Total Variation Meets Differential Privacy

Elena Ghazi*, Ibrahim Issa*†
*Electrical and Computer Engineering Department, American University of Beirut, Beirut, Lebanon
†Center for Advanced Mathematical Sciences, American University of Beirut, Beirut, Lebanon
ebg03@mail.aub.edu, ibrahim.issa@aub.edu.lb

***Abstract*—The framework of approximate differential privacy is considered, and augmented by leveraging the notion of "the total variation of a (privacy-preserving) mechanism" (denoted by η-TV). With this refinement, an exact composition result is derived, and shown to be significantly tighter than the optimal bounds for differential privacy (which do not consider the total variation). Furthermore, it is shown that (ε, δ)-DP with η-TV is closed under subsampling. The induced total variation of commonly used mechanisms are computed. Moreover, the notion of total variation of a mechanism is studied in the local privacy setting and privacy-utility tradeoffs are investigated. In particular, total variation distance and KL divergence are considered as utility functions and studied through the lens of contraction coefficients. Finally, the results are compared and connected to the locally differentially private setting.**

I. Introduction

Given a database D that contains private records (e.g., medical records), one would like to design a mechanism M to answer queries issued by a curious (but not necessarily malicious) analyst (e.g., medical researchers). Such mechanism should provide useful answers while preserving the privacy of the individuals whose records are in D. In this context, differential privacy (DP) [1] was developed, and subsequently widely adopted, to quantify the privacy guarantees of a given mechanism M. Essentially, given any database D, any possible record, and the output $M(D)$ of a differentially-private M, the analyst cannot tell (except with insignificant probability) whether or not the record appears in D.

However, the analyst may issue several queries, thus degrading the privacy guarantees. Bounds on the resulting guarantees are called *composition theorems*. Notably, pure differential privacy yields pessimistic composition theorems. As such, researchers have developed relaxations and variations of differential privacy to yield better composition behaviour, e.g., approximate DP [2], Rényi-DP [3], concentrated-DP [4], Gaussian DP [5], etc.

Herein, we focus on approximate DP and Gaussian DP as they have the clearest operational interpretation. In particular, Wasserman and Zhou [6] and Kairouz *et al.* [7] provided an equivalent characterization of (ε, δ)-DP in terms of the achievable region of errors (type I and type II) of a binary hypothesis testing experiment. Utilizing this view, Kairouz *et al.* [7] proved an exact composition result for (ε, δ)-DP in terms of a family of $\{(\varepsilon_j, \delta_j)\}_j$ parameters. As observed by Dong *et al.* [5], the binary hypothesis view also elucidates why composition results for (ε, δ)-DP are pessimistic: the framework is not rich enough to capture the induced privacy region of the composed mechanisms (with one (ε, δ) pair). Instead, they parameterized the privacy guarantee by (the lower convex envelope of) the privacy region itself, and denoted it by f-DP [5]. Hence, they showed that f-DP is closed under composition, and proved an interesting limiting behavior wherein, under certain assumptions, the limit of the composition of n f-DP mechanisms converges to the guarantees of the Gaussian mechanism.

In this work, we introduce a simple refinement to (approximate) differential privacy that yields better composition results. Namely, we leverage the total variation (TV) of a mechanism, denoted by η-TV, so that we keep track of both the (ε, δ)-parameters for DP and the η-parameter for TV. This retains an equivalent characterization in terms of binary hypothesis testing. In addition to enabling tighter analyses, explicitly incorporating total variation offers important advantages in the context of learning algorithms. In particular, η-TV has emerged as the target measure to control *membership inference attacks* (MIA) — which pose an orthogonal concern compared to differential privacy. Moreover, η-TV can be used to bound the generalization error of the algorithms.

Herein, our contributions consist of the following:

- We prove an *exact* composition bound for (ε, δ)-DP coupled with η-TV (Section III).
- We show the provided bounds can be significantly tighter than bounds where total variation is not taken into consideration (using examples, as well as asymptotic analysis [5]) (Sections III and V).
- We show that (ε, δ)-DP with η-TV is closed under subsampling (where the mechanism computes the query answer on a random subset of the database) (Section IV).
- We compute the total variation of commonly used mechanisms and demonstrate an interesting connection with the staircase mechanism [8] (Section VI).
- We analyze the differentially private stochastic gradient descent algorithm (Section VI-D).

We also study the privacy setting in which data remains private even from the statistician. Given distributions P_0 and P_1 and an ε-locally differentially private (ε-LDP) privatization mechanism Q, let M_0 and M_1 be the induced output marginals. Herein, the goal is to design

This work was presented in part at the IEEE International Symposium on Information Theory, June 2023 (Taipei, Taiwan). Part of this work was done while Ibrahim Issa was a visiting professor at EPFL (Lausanne, Switzerland). Elena Ghazi was at the American University of Beirut (Beirut, Lebanon). This work was supported by the University Research Board at the American University of Beirut.

a mechanism that satisfies a certain privacy constraint (to satisfy the data owners), while maximizing a certain statistical utility function (to satisfy the statistician). Duchi *et al.* [9] proved bounds on the symmetrized KL divergence between M_0 and M_1 in the ε-LDP setting. Kairouz *et al.* [10] provided an ε-LDP binary mechanism that maximizes the total variation between M_0 and M_1 for all P_0 and P_1, and approximates the mechanism that maximizes the KL divergence between the marginals. We consider the notion of total variation of a mechanism in the local privacy setting and study privacy-utility tradeoffs (Section VII). In particular, our contributions consist of the following.

- We provide a mechanism that maximizes the total variation between the marginals in the case of ε-LDP with η-TV.
- We generalize bounds on the contraction coefficient of KL divergence for a privacy-preserving mechanism [9]–[11] by accounting simultaneously for ε-LDP and η-TV constraints.
- We generalize bounds on χ^2 of the marginal distributions in terms of the total variation of the input distributions [9], [11].
- Finally, given an ε-LDP mechanism, we show how to construct an ε-LDP with η-TV mechanism, and use this result to connect the corresponding privacy-utility tradeoffs in the two settings (i.e., enforcing ε-LDP constraint versus enforcing ε-LDP and an η-TV constraint).

Finally, it is worth mentioning that it is typically simple to compute bounds on total variation using, for instance, KL divergence or Chernoff information. As such, the computational overhead of keeping track of total variation is relatively low. Furthermore, it may be estimated from data in a "black-box" manner [12].

Prior Work: Total variation is arguably a "natural" measure to consider in privacy analysis and has indeed been used in the literature as a privacy metric: Barber and Duchi [13] compared various definitions of privacy for several estimation problems, including *α-total variation privacy*. Geng *et al.* [14] derived the optimal $(0, \delta)$-differentially private noise-adding mechanism for single real-valued query function under a cost-minimization framework. Jia *et al.* [15] used the notion of *total variation distance privacy* to accurately estimate privacy risk, despite it being a weaker privacy definition than differential privacy. The total variation also appeared as a bound on generalization metrics in the context of machine learning, and consequently as a bound on vulnerability to Membership Inference Attacks (MIAs). Dwork *et al.* [16] examined the total variation as a notion of fairness. Bassily *et al.* [17] defined *TV-stability* as a notion of algorithmic stability used to bound generalization error, which later appeared in Raginsky *et al.*'s analysis of bias in learning algorithms [18]. Kulynych *et al.* [19], [20] proved through the notion of *distributional generalization* that the total variation bounds vulnerability to MIAs, as well as disparate vulnerability against MIAs (unequal success rate of MIAs against different population subgroups). Chatzikokolakis *et al.* [12] also studied *Bayes security*, a security metric inspired by the cryptographic advantage, equal to the complement of the total variation. Additionally (as will be seen in this paper), it naturally appears in many existing analysis. Note that our work differs from existing frameworks that leverage the notion of total variation in a privacy setting (e.g., [13]–[15]) in that we suggest keeping track of the total variation of a mechanism *in addition to* differential privacy parameters, rather than as a standalone metric. This framework continues to leverage the strong guarantees of differential privacy, while additionally accounting for the mechanism-specific property of total variation.

Novel Contributions. In addition to our ISIT findings, we consider applications of our results in the context of membership inference attacks, as well as differentially private learning algorithms (namely, noisy stochastic gradient descent). Furthermore, we study the notion of total variation of a mechanism in the local privacy setting. We obtain generalized privacy-utility bounds, showing that one can obtain more precise guarantees by accounting for both ε and η parameters.

II. Preliminaries and Definitions

Fix an alphabet $\mathcal{X}$ and an integer $m \in \mathbb{N}$. A database D is an element in $\mathcal{X}^m$. Two databases that differ in one entry are called neighboring databases. A (query-answering) mechanism M is a randomized map from $\mathcal{X}^m$ to an output space, which we denote by $\mathcal{Y}$.

Definition 1 (Differential Privacy (DP) [1]): Given $\varepsilon \geq 0$ and $\delta \in [0, 1]$, a mechanism M is (ε, δ)-differentially private if, for all neighboring databases D_0 and D_1 and all $S \subseteq \mathcal{Y}$,

$$\mathbf{Pr}(M(D_0) \in S) \leq e^{\varepsilon}\mathbf{Pr}(M(D_1) \in S) + \delta.$$

Wasserman and Zhou [6] and Kairouz *et al.* [7] provide an equivalent characterization of $(\varepsilon, 0)$- and (ε, δ)-DP, respectively, in terms of binary hypothesis testing. In particular, consider a mechanism M, two neighboring databases D_0 and D_1, and let P_0 and P_1 be the corresponding distributions over $\mathcal{Y}$ of $M(D_0)$ and $M(D_1)$. Given a random output $Y \in \mathcal{Y}$, an adversary aims to distinguish between

$$H_0\colon Y \sim P_0, \quad \text{and} \quad H_1\colon Y \sim P_1.$$

Let h be any (possibly randomized) decision function, $h : \mathcal{Y} \to \{0, 1\}$, and denote by $\beta_{\mathrm{I}}(P_0, P_1, h) = P_0(h(Y) = 1)$ and $\beta_{\mathrm{II}}(P_0, P_1, h) = P_1(h(Y) = 0)$ the type I and type II errors, respectively. Among such decision functions, the ROC curve (also called the tradeoff function [5]) describes the best type II error that can be achieved for a given level of type I error:

Definition 2 (ROC): For a binary hypothesis testing experiment with distributions P_0 and P_1, the ROC curve, $f : [0, 1] \to [0, 1]$, is defined as

$$f(P_0, P_1)(t) = \inf_h \{\beta_{\mathrm{II}}(P_0, P_1, h) : \beta_{\mathrm{I}}(P_0, P_1, h) \leq t\}.$$

DP can then be described in terms of the ROC curves:

Theorem 1 ([7, Theorem 2.1]): Given $\varepsilon \geq 0$ and $\delta \in [0, 1]$, a mechanism M is (ε, δ)-DP if and only if for

all neighboring databases D_0 and D_1 and all $t \in [0,1]$,

$$t + e^{\varepsilon} f(P_0, P_1)(t) \geq 1 - \delta,$$
$$e^{\varepsilon} t + f(P_0, P_1)(t) \geq 1 - \delta,$$

where $M(D_0) \sim P_0$ and $M(D_1) \sim P_1$.
The region $(\beta_{\mathrm{I}}, \beta_{\mathrm{II}})$ corresponding to the above constraints will also be referred to as the "privacy region" of the mechanism. The ROC curve corresponding to (ε, δ)-DP is shown in Figure 1 (region in gray). We

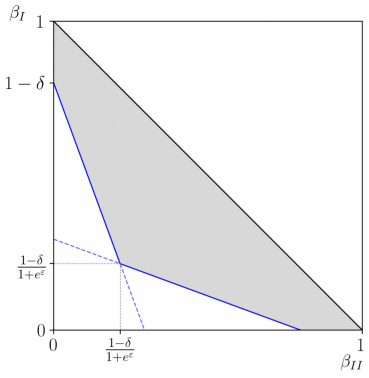


Fig. 1. ROC corresponding to (ε, δ)-DP.

augment the (ε, δ)-DP framework using the notion of "total variation of the mechanism" (which first appeared in [13]). First, recall

Definition 3 (Total Variation): Given two distributions P and Q over a common alphabet $\mathcal{Y}$, the total variation is defined as follows:

$$d_{TV}(P, Q) = \sup_{A \subseteq \mathcal{Y}} (P(A) - Q(A)).$$

Definition 4 (Total Variation of a Mechanism): Given $\eta \in [0,1]$, a mechanism M has total variation less than η (or $d_{TV}(M) \leq \eta$ or η-TV, for short) if, for all neighboring databases D_0 and D_1,

$$d_{TV}(P_0, P_1) \leq \eta, \tag{1}$$

where $M(D_0) \sim P_0$ and $M(D_1) \sim P_1$.
Total variation is closely related to hypothesis testing: given two distributions P_0 and P_1, then $d_{TV}(P_0, P_1) \leq \eta$ if and only if[1] $f(P_0, P_1)(t) \geq 1 - \eta - t$ for all $t \in [0,1]$. As such, M is η-TV can be rewritten as $(0, \eta)$-DP.

Corollary 1: Given $\varepsilon \geq 0$, $\delta \in [0,1]$, and $\eta \in [0, \delta + (1-\delta)\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}]$, a mechanism M is (ε, δ)-DP and η-TV if and only if, for all neighboring databases D_0 and D_1 and all $t \in [0,1]$, As such, M is η-TV can be rewritten as $(0, \eta)$-DP.

$$t + e^{\varepsilon} f(P_0, P_1)(t) \geq 1 - \delta, \tag{2}$$
$$e^{\varepsilon} t + f(P_0, P_1)(t) \geq 1 - \delta, \tag{3}$$
$$t + f(P_0, P_1)(t) \geq 1 - \eta, \tag{4}$$

[1]This can be seen by noting the following: consider the line with slope -1 tangent to the curve f, then its y-intercept is given by

$$\begin{aligned} -f^{\star}(-1) &= -\sup_{s \in [0,1]} \{-s - f(s)\} = \inf_{s \in [0,1]} \{s + f(s)\} \\ &= \inf_{A \subseteq \mathcal{Y}} P_0(A) + P_1(A^c) = 1 + \inf_{A \subseteq \mathcal{Y}} P_0(A) - P_1(A) \\ &= 1 - d_{TV}(P_0, P_1). \end{aligned}$$

where $M(D_0) \sim P_0$ and $M(D_1) \sim P_1$.

Remark 1: The maximum possible value for η can be seen from the region in Figure 1, where $\beta_{\mathrm{I}} + \beta_{\mathrm{II}} \geq \frac{2(1-\delta)}{1+e^{\varepsilon}} = 1 - \left(\delta + (1-\delta)\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\right)$.
The corresponding ROC curve is shown in Figure 2, with the privacy region shaded in gray. Compared with Figure 1, the effect of introducing the η-TV constraint corresponds to "slicing" the original region ("moving it away" from the origin). Jia *et al.* [15, Theorem 4] concurrently studied η-TV privacy in the hypothesis testing framework. However, their formulation does not simultaneously incorporate both ε-DP and η-TV (as opposed to this work). Nevertheless, enforcing both ε-DP and η-TV constraints can be captured within the f-DP framework [5], by choosing f that satisfies the inequalities of Corollary 1 with equality.

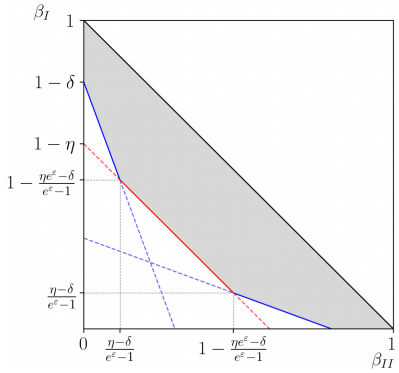


Fig. 2. ROC corresponding to (ε, δ)-DP η-TV.

III. Main Result: Adaptive Composition

Our main result characterizes exactly the k-fold adaptive composition of (ε, δ)-DP and η-TV mechanisms. More precisely, the k-fold binary hypothesis experiment is defined as follows:

- An adversary $\mathcal{A}$ chooses two neighboring databases D_0 and D_1.
- A parameter $b \in \{0,1\}$ is fixed (unknown to $\mathcal{A}$).
- $\mathcal{A}$ issues k queries $q_1, q_2, \ldots, q_k$, and receives $M_1(D_b), M_2(D_b), \ldots, M_k(D_b)$. The choice of each query (or equivalently, the mechanism M_i) may depend on the outputs of previous queries. In particular, letting $\mathcal{Y}_i$ be the output alphabet of M_i, then M_i is a mechanism $M_i : \mathcal{D} \times \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_{i-1} \to \mathcal{Y}_i$, satisfying: for all $y^{i-1} \in Y_1 \times Y_2 \ldots Y_{i-1}$, $M_i(.|y^{i-1})$ is an (ε, δ)-DP and η-TV mechanism.
- $\mathcal{A}$ guesses $\hat{b} \in \{0,1\}$.

Given that each M_i is an (ε, δ)-DP and η-TV mechanism, what is the privacy region induced by the above experiment? Allowing the choice of each query to depend on the outcomes of previous queries is referred to as *adaptive composition*.

Theorem 2: For any $\varepsilon \geq 0$, $\delta \in [0,1]$, and $\eta \in [\delta, \delta + \frac{(e^{\varepsilon}-1)(1-\delta)}{1+e^{\varepsilon}}]$, the class of (ε, δ)-differentially private and η-total variation mechanisms satisfies

$$(j\varepsilon, 1 - (1-\delta)^k (1-\delta_j))\text{-differential privacy with}$$
$$d_{TV} = 1 - (1-\delta)^k (1-\delta_0)$$

under k-fold adaptive composition, for all $j \in \{0, 1, ..., k\}$, where

$$\delta_j = \sum_{a=0}^{k-j-1} \binom{k}{a} \sum_{\ell=0}^{\lceil \frac{k-j-a}{2} \rceil - 1} \binom{k-a}{\ell} \left(\frac{1-\alpha}{1+e^{\varepsilon}}\right)^{k-a} \cdot \alpha^{a} \left(e^{(k-\ell-a)\varepsilon} - e^{(\ell+j)\varepsilon}\right) \quad (5)$$

and $\alpha = 1 - \dfrac{(\eta-\delta)(1+e^{\varepsilon})}{(1-\delta)(e^{\varepsilon}-1)}$.

The proof follows the machinery developed by Kairouz *et al.* [7]. In particular, we introduce a "dominating" mechanism, i.e., a mechanism which exactly achieves the region described by the equations of Corollary 1. The achievability is then shown by analyzing the (non-adaptive) composition of the dominating mechanism (which admits a simple form). The key component of the converse, similarly to [7], is a result by Blackwell [21, Corollary of Theorem 10] which states the following: for two binary hypothesis testing experiments A and B, if $f_B(t) \geq f_A(t)$ for all $t \in [0, 1]$, then B can be simulated from A. This is why the dominating mechanism yields the worst-case degradation (other mechanisms can be simulated from it).

A. Comparison

The composition bound proved by Kairouz *et al.* [7] states that for any ε and $\delta \in [0, 1]$, the class of (ε, δ)-differentially private mechanisms satisfies

$$(j\varepsilon, 1 - (1-\delta)^k(1-\delta_j))\text{-differential privacy}$$

under k-fold adaptive composition, for all $j \in \{k - 2i : i = 0, 1, \ldots, \lfloor k/2 \rfloor\}$, where

$$\delta_j = \frac{\sum_{\ell=0}^{\frac{k-j}{2}-1} \binom{k}{\ell} \left(e^{(k-\ell)\varepsilon} - e^{(\ell+j)\varepsilon}\right)}{(1+e^{\varepsilon})^k}. \quad (6)$$

Although the above bound is tight, commonly used mechanisms, like the Laplace mechanism, do not achieve the entire region described in Theorem 1. Taking into consideration the mechanism's total variation leads to a better bound on its privacy region, as shown in Figure 3 (details of the computations deferred to Section VI-A).

We illustrate the improved composition bound in Figure 4, with $\varepsilon = 1$, $\delta = 0$, and $\eta = 0.7\frac{e^{\varepsilon}-1}{1+e^{\varepsilon}}$ (i.e., $\alpha = 0.3$). Note that the refined composition bound involves lines with slopes of the form $-e^{\pm j\varepsilon}$ for $j \in \{0, 1, ..., k\}$, while the bound introduced in [7] only involves values of j such that $0 \leq j \leq k$ and j and k have the same parity. Even if one were to only observe the lines that correspond to values of j that have the same parity as k (blue vs green in Figure (4b)), the refined bound still improves on the previous bound.

Remark 2: Jia *et al.* [15, Theorem 12] derived the k-fold composition for any η with $\varepsilon = \log \frac{1+\eta}{1-\eta}$. However, this reduces to the standard ε-DP setting ($\eta = \frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}$), which has already been derived by Kairouz *et al.* [7]. The same applies to [15, Theorem 18]. Hence, contrary to our result, the η parameter is not used to further bound the privacy region and obtain tighter composition results.

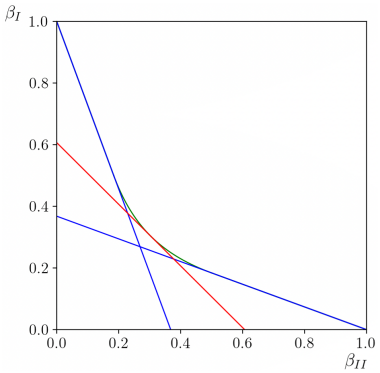
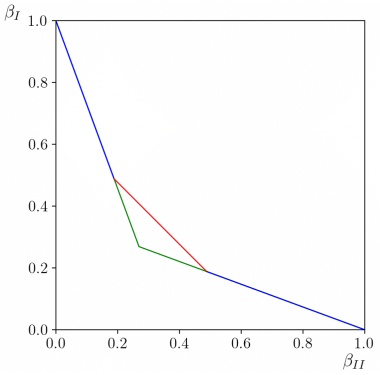
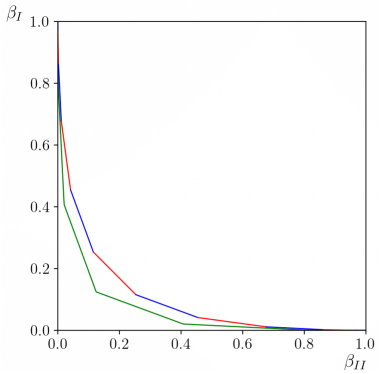


Fig. 3. The Laplace mechanism's exact tradeoff function is shown in green (for $\varepsilon = 1$). The blue lines correspond to the differential privacy constraints. The red line corresponds to $\beta_{\mathrm{I}} = 1 - \eta_{\mathrm{Lap}} - \beta_{\mathrm{II}}$ where $\eta_{\mathrm{Lap}} = 1 - e^{-\varepsilon/2}$ is the total variation of the Laplace mechanism.



(a) $k = 1$.



(b) $k = 5$.

Fig. 4. The induced tradeoff functions of the dominating mechanisms for $\varepsilon = 1$, $\delta = 0$, and $\alpha = 0.3$. The green line shows the tradeoff function of ε-DP (given by equation (6)). The blue and red lines plot the tradeoff given by Theorem 2, where they correspond, respectively, to odd (same parity as $k = 5$) and even choices of j in equation (5).

B. Discussion

1) Membership Inference Attacks: Prior works have utilized the notion of total variation to capture an adversary's advantage and to bound vulnerability to Membership Inference Attacks (MIAs), in which the attacker tries to determine whether a given record was part of the model's training data. Chatzikokolakis *et al.* [12] studied the notion of Bayes security (denoted β^*), a metric inspired by cryptographic advantage, and which is proven to be the complement of the total variation of the two maximally distant rows of the channel. The authors also proved a parallel composition result, which states

that the composition of two channels that are β_1^*-secure and β_2^*-secure yields a $\beta_1^*\beta_2^*$-secure channel. We can recover this result for the special case $\beta_1^* = \beta_2^*$ by considering the total variation of the composed dominating η-TV mechanism with $\eta = 1 - \beta^*$ (or, equivalently, the dominating $(0, \eta)$-DP mechanism), which is $1-(1-\eta)^k$ (this can be seen by noting that the total variation under composition is $d_{TV} = 1 - (1 - \delta)^k(1 - \delta_0)$, and that $\delta_0 = 0$ in equation (6) for $\varepsilon = 0$ and $\delta = \eta$). Kulynych *et al.* [19] studied the phenomenon of disparate vulnerability to MIAs, i.e., the unequal success rate of MIAs against different population subgroups (or the difference between subgroup vulnerability to MIAs). The authors showed that the worst-case vulnerability of a subgroup to MIAs is bounded by the total variation of the mechanism, and consequently, so is the disparity.

Considering the importance of MIA attacks and subsequently the importance of the total variation of the mechanism, it is crucial to keep track of the evolution of the η parameter under composition. Indeed, our result enables accurate tracking of η by coupling the η-TV and ε-DP constraints. By contrast, studying the worst-case evolution of η without taking ε into consideration would yield much looser bounds: η would evolve as $1-(1-\eta)^k$. Moreover, from a mechanism design point of view, our framework enables greater flexibility. In particular, when comparing two mechanisms, a designer may want to consider both their (ε, δ)-parameters as well as their η-parameters to have a more complete comparison.

2) Differentially Private Stochastic Gradient Descent: The more refined analysis of the composition of privacy-preserving mechanisms is crucial in studying algorithms like DP-SGD, which involves subsampling followed by a large number of compositions (so that standard composition results yield loose bounds). We derive a subsampling result in Section IV, and couple it with our composition result to analyze DP-SGD in Section VI-D.

C. Proof

We propose an (ε, δ)-differentially private mechanism with total variation η that achieves the entire privacy region corresponding to (ε, δ)-DP and η-TV (described in Corollary 1, shown in Figure 2), and that therefore *dominates* all (ε, δ)-differentially private mechanisms with total variation η. We then derive its privacy region under non-adaptive k-fold composition, which consists of the intersection of several $(\varepsilon_j, \delta_j)$-DP privacy regions. Finally, we show that this derived privacy region is the largest achievable privacy region under k-fold (adaptive) composition of (ε, δ)-DP and η-TV mechanisms, hence proving the tightness of our result.

1) Dominating Mechanism: Kairouz [7] *et al.* introduced a mechanism, which we will denote by $M_{\varepsilon,\delta} : \mathcal{X}^m \to \{0, 1\}$, that dominates all (ε, δ)-differentially private mechanisms in the following sense: for any (ε, δ)-DP mechanism $M : \mathcal{X}^m \to \mathcal{Y}$, for any pair of databases D_0 and D_1, there exists a (possibly randomized) mapping $T : \{0, 1\} \to \mathcal{Y}$ such that, $M(D)$ and $T(M_{\varepsilon,\delta}(D))$ have the same distribution, where $D \in \{D_0, D_1\}$. We adopt a similar approach, and provide a mechanism that dominates all (ε, δ)-differentially private mechanisms with total variation η. Due to lack of space, we provide herein the proof of the case $\delta = 0$ only and include the details on how we generalize the result to approximate differential privacy in Appendix A.

The following mechanism (as in [7]) does not depend on the database entries or the query; it only depends on the hypothesis, and will be denoted by $M_{\varepsilon,0,\eta}$. It outputs $X_0 \sim P_0$ in the case of the null hypothesis, and $X_1 \sim P_1$ in the case of the alternative hypothesis, where P_0 and P_1 are defined as follows, for $\varepsilon \geq 0$ and $\alpha \in [0, 1]$:

$$P_0 = \begin{cases} \frac{(1-\alpha)e^\varepsilon}{(1+e^\varepsilon)}, & x = 0, \\ \alpha, & x = 1, \\ \frac{(1-\alpha)}{(1+e^\varepsilon)}, & x = 2, \end{cases} \quad P_1 = \begin{cases} \frac{(1-\alpha)}{(1+e^\varepsilon)}, & x = 0, \\ \alpha, & x = 1, \\ \frac{(1-\alpha)e^\varepsilon}{(1+e^\varepsilon)}, & x = 2. \end{cases} \tag{7}$$

The above mechanism's total variation is $d_{TV}(P_0, P_1) = \frac{(e^\varepsilon-1)(1-\alpha)}{(1+e^\varepsilon)}$. So for a target $\eta \in [0, \frac{e^\varepsilon-1}{e^\varepsilon+1}]$, we set $\alpha = 1 - \eta\frac{e^\varepsilon+1}{e^\varepsilon-1}$. For $\eta = \frac{e^\varepsilon-1}{e^\varepsilon+1}$, we recover $M_{\varepsilon,0}$. It is easy to verify that $f(P_0, P_1)$ exactly achieves the region described in Corollary 1.

2) Composed Dominating Mechanism: We show that the region described in Theorem 2 is achievable by analyzing the non-adaptive composition of $M_{\varepsilon,0,\eta}$, i.e., in the k-fold binary hypothesis experiment, we set $M_i = M_{\varepsilon,0,\delta}$ for all $i = 1, 2, \ldots, k$.

Since the privacy region of the composed mechanism is a convex set, we can describe its boundaries with a set of lines tangent to its envelope. For a slope $-\gamma$, we want to find the largest shift $1 - \delta'$ such that the line $\beta_{\mathrm{I}} = -\gamma\beta_{\mathrm{II}} + 1 - \delta'$ is below the tradeoff curve.

For some acceptance set A, consider the point $(\beta_{\mathrm{II}}(A), \beta_{\mathrm{I}}(A))$ included in the composed mechanism's privacy region. We want $-\gamma\beta_{\mathrm{II}}(A) + 1 - \delta' \leq \beta_{\mathrm{I}}(A)$:

$$\begin{aligned} \delta' &\geq 1 - \beta_{\mathrm{I}}(A) - \gamma\beta_{\mathrm{II}}(A) && (8) \\ &= P_0(A) - \gamma P_1(A). && (9) \end{aligned}$$

We choose the smallest δ' such that the inequality (8) is satisfied for all A, i.e.,

$$\delta' = \max_{A \in \mathcal{X}} \{P_0(A) - \gamma P_1(A)\}.$$

We represent the output of a k-fold composition experiment as a sequence $s^k \in \{0, 1, 2\}^k$, and use the following notations:

$$a = |\{i : s_i^k = 1\}|, \text{ and } \ell = |\{i : s_i^k = 2\}|. \tag{10}$$

Let $\tilde{P}_0$ and $\tilde{P}_1$ be the probabilities of obtaining a sequence s^k under k-fold composition when the true database is D_0 and D_1, respectively. Since the outputs of the mechanisms are i.i.d.,

$$\begin{aligned} \tilde{P}_0(s^k) &= P_0(X = 0)^{k-\ell-a} P_0(X = 1)^a P_0(X = 2)^\ell \\ &= \left(\frac{1-\alpha}{1+e^\varepsilon}\right)^{k-a} e^{(k-\ell-a)\varepsilon}\alpha^a, \end{aligned}$$

and

$$\begin{aligned} \tilde{P}_1(s^k) &= P_1(X = 0)^{k-\ell-a} P_1(X = 1)^a P_1(X = 2)^\ell \\ &= \left(\frac{1-\alpha}{1+e^\varepsilon}\right)^{k-a} e^{\ell\varepsilon}\alpha^a. \end{aligned}$$

By the Neyman-Pearson lemma [22, Theorem 11.7.1], points on the ROC curve are achieved by tests of the form $\{s^k : \frac{\tilde{P}_0(s^k)}{\tilde{P}_1(s^k)} > T\}$, for some threshold T. Letting

$$E \equiv \{0 \le \tilde{\varepsilon} \le \infty : \tilde{P}_0(s_k) = e^{\tilde{\varepsilon}} \tilde{P}_1(s_k) \text{ for some } x \in \mathcal{X}\},$$

any $\tilde{\varepsilon} \notin E$ does not contribute to the curve.

Remark 3: It is sufficient to consider $\tilde{\varepsilon} \ge 0$ as the tradeoff function is symmetric due to the symmetry of P_0 and P_1. The symmetry is then accounted for by the symmetry of (ε, δ)-DP.
Noting that

$$\frac{\tilde{P}_0(s_k)}{\tilde{P}_1(s_k)} = \frac{e^{(k-\ell-a)\varepsilon}}{e^{\ell\varepsilon}}, \text{ it follows } \tilde{\varepsilon} = (k-2\ell-a)\varepsilon.$$

Therefore, the distinct values that $\tilde{\varepsilon}$ can take are of the form $\varepsilon_j = j\varepsilon$ where $j \in \{0, 1, ..., k\}$. To every ε_j corresponds a δ_j of the form

$$\delta_j = \max_{A \in \mathcal{X}^k} \{\tilde{P}_0(A) - e^{\varepsilon_j} \tilde{P}_1(A)\},$$

where A is the acceptance region. For a fixed j, we consider the sequences s_k such that $\tilde{P}_0(s_k) - e^{j\varepsilon} \tilde{P}_1(s_k) > 0$, i.e.

$$\left(\frac{1-\alpha}{1+e^{\varepsilon}}\right)^{k-a} \alpha^a \left(e^{(k-\ell-a)\varepsilon} - e^{(\ell+j)\varepsilon}\right) > 0.$$

There are $\binom{k}{a}\binom{k-a}{\ell}$ sequences of the form described in (10), but we are only interested in the sequences that satisfy $e^{(k-\ell-a)\varepsilon} - e^{(\ell+j)\varepsilon} > 0$, or $k - j > 2\ell + a$, hence the summation bounds in equation (5).

3) Converse: Since the tradeoff function of $M_{\varepsilon,0,\eta}$, $f_{\varepsilon,0,\eta}$, matches with equality the constraints given in Corollary 1, the tradeoff function of any ε-DP and η-TV mechanism M must satisfy $f_M(t) \ge f_{\varepsilon,0,\eta}(t)$ for all $t \in [0, 1]$. Hence by Blackwell's result [21, Corollary of Theorem 10], M can be simulated from $M_{\varepsilon,0,\delta}$. Given this fact, one can follow exactly the converse proof in [7] to show that the output of the k-fold composition of any ε-DP and η-TV mechanisms can also be simulated using the (non-adaptive) k-fold composition of $M_{\varepsilon,0,\eta}$, hence the corresponding tradeoff function will be greater than the composition of $M_{\varepsilon,0,\eta}$. The intuition is the following: Let M_{D_1} and M_{D_2} be two instances of the dominating (ε, δ)-DP with η-TV mechanism. The adversary chooses a mechanism M_1, which is (ε, δ)-DP with η-TV, observes its output, and based on it, chooses M_2, which is also (ε, δ)-DP with η-TV. (M_1, M_2) can be simulated from (M_{D_1}, M_{D_2}), since M_1 can be simulated from M_{D_1}, and M_2 can be simulated from M_{D_2}.

IV. Subsampling

Subsampling is a simple method to "amplify" privacy guarantees: before answering a query on a given database of size n, first choose uniformly at random a subset of size m, $1 \le m \le n$. This procedure will be denoted by Sample_m. Then, compute the query answer on the sampled database.

Proposition 1: Given $\varepsilon > 0$, $\delta \in [0, 1]$, and $\eta \in [0, 1]$, if $M : \mathcal{X}^m \to \mathcal{Y}$ is (ε,δ)-DP and η-TV, then the subsampled mechanism $M \circ \mathrm{Sample}_m : \mathcal{X}^n \to \mathcal{Y}$ is $p\eta$-TV and $(\log(1 + p(e^{\varepsilon} - 1)), p\delta)$-DP on $\mathcal{X}^n$, where $m \le n$ and $p = \frac{m}{n}$. Moreover, this result is tight.
The proof (deferred to Appendix C) is based on the privacy amplification by subsampling result for (ε,δ)-DP appearing in [23]–[26] and stated in [27, Theorem 29]. As such, (ε, δ)-DP with η-TV is closed under subsampling. Moreover, as can be seen from the proof, it holds in general that if M is η-TV, then $M \circ \mathrm{Sample}_m$ is $p\eta$-TV. This result was concurrently derived by Jia *et al.* [15], but we include a simple proof herein and further show tightness.

V. Asymptotic Behaviour

Inspired by the hypothesis testing view of differential privacy [6], [7], Dong *et al.* [5] introduced a new privacy measure, called f-differential privacy, based directly on tradeoff functions (i.e., ROC curves):

Definition 5 (f-DP [5]): Let $f : [0, 1] \to [0, 1]$ be a convex, continuous, non-increasing function, satisfying $f(x) \le 1 - x$. A mechanism M is f-DP if for all neighboring databases D_0 and D_1 and $t \in [0, 1]$,

$$f(P_0, P_1)(t) \ge f(t), \tag{11}$$

where $M(D_0) \sim P_1$ and $M(D_1) \sim P_1$.

As such, (ε, δ)-DP, as well as (ε, δ)-DP with η-TV, are special cases of f-DP, with f given by the corresponding curves shown in Figure 2.

Notably, the composition of f-DP mechanisms is also f-DP (with a potentially different f), i.e., f-DP is closed under composition. Moreover, the authors [5] show that composition for f-DP follows a central limit theorem-like behavior in the following sense. Given $\mu \in \mathbb{R}$, let

$$G_\mu(\alpha) = f(\mathcal{N}(0, 1), \mathcal{N}(\mu, 1))(\alpha), \tag{12}$$

for $\alpha \in [0, 1]$. Then, the composition of n f-DP mechanisms can be approximated by G_μ (with an appropriate μ) [5, Theorem 3.4]. Indeed, under suitable assumptions, the limit of the composition is equal to G_μ [5, Theorem 3.5]. For instance, letting $\varepsilon_n = \frac{\mu}{\sqrt{n}}$, then the composition of n ε_n-DP mechanisms converges to G_μ [5, Theorem 3.6]. Here, we provide an analogous result for ε-DP with η-TV (proof deferred to Appendix B):

Theorem 3: Consider a triangular array $\{(\varepsilon_{ni}, \eta_{ni}) : 1 \le i \le n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} \sum_{i=1}^{n} \varepsilon_{ni}\eta_{ni} = \frac{\mu^2}{2}$ for some $\mu \ge 0$, and $\lim_{n\to\infty} \max_{1\le i\le n} \varepsilon_{ni} = 0$. Let $M_{ni} : \mathcal{X}^m \to \mathcal{Y}_{ni}$ be an ε_{ni}-DP and η_{ni}-TV mechanism, and let $M_n(D) = (M_{n1}(D), \ldots, M_{nn}(D))$, $D \in \mathcal{X}^m$. Then, there exists a sequence of functions $\{f_n\}$ such that $M_n(D)$ is f_n-DP and

$$\lim_{n\to\infty} f_n = G_\mu, \tag{13}$$

uniformly over $[0, 1]$, where G_μ is defined in Eq. (12).

For pure differential privacy, $\eta = \frac{e^{\varepsilon}-1}{e^{\varepsilon}+1} \approx \frac{\varepsilon}{2}$ for small ε, so that the above result matches Theorem 3.6 of [5]. However, if for small ε, $\eta \approx \frac{\varepsilon}{c_0}$ for some $c_0 > 2$, then: if Theorem 3.6 of [5] yields a parameter μ, Theorem 3 will yield $\mu\sqrt{\frac{2}{c_0}}$. It is worth noting that one can always

reduce the total variation of a mechanism by a factor $(1-\delta')$ by passing its output through an erasure channel with parameter δ'. Moreover, let $c_1 \in (0, 1/2)$, $\varepsilon_{ni} = \frac{\mu}{n^{c_1}}$, and $\eta_{ni} = \frac{\mu}{2n^{1-c_1}}$, for all $1 \leq i \leq n$. Then, the conditions of the theorem are satisfied, and the limit of the composition is G_μ. As opposed to decaying as $1/\sqrt{n}$, ε_n could decay with an arbitrarily small power of n, as long as this is "compensated" for with a faster decay for the total variation. Finally, it is worth noting that, as compared to the f-DP framework, (ε, δ)-DP with η-TV is significantly simpler as we keep track of only three parameters, while still yielding considerable advantages over (ε, δ)-DP. Moreover, it is easier to work with from a design perspective (choosing ε and η parameters versus choosing a function).

Remark 4: Although Theorem 3 was stated for non-adaptive composition, one could allow adaptive choices of M_{ni} and retain the same result (this is due to a tightness result by Dong *et al.* [5, Theorem 3.2]).

VI. Applications

In this section, we compute the total variation for the Laplace, Gaussian, and staircase mechanisms, to demonstrate improvement and/or simplicity of analysis for commonly used mechanisms. Using well-known bounds on the total variation (e.g., Pinsker's inequality), one could also derive simple closed-form bounds on the total variation of the composition. While one could efficiently approximate the composed tradeoff functions of mechanisms using Fourier-based methods [28]–[30], doing so requires the CDF of the "privacy loss" random variable and may not lead to a closed form. We also discuss an additional use-case for keeping track of the total variation of a mechanism in the context of Membership Inference Attacks.

A. Laplace Mechanism

The probability density function of a Laplace distribution centered at 0 and with scale b is

$$Lap(x|b) = \frac{1}{2b}\exp(-\frac{|x|}{b}), \tag{14}$$

and will be denoted by $Lap(b)$. For a (query) function $q : \mathcal{X}^m \to \mathbb{R}$, the output of the Laplace mechanism, denoted M_{Lap}, is $M_{\text{Lap}}(X) = q(X) + Y$ where $Y \sim Lap(\Delta/\varepsilon)$, and the query sensitivity Δ is defined as

$$\Delta = \max_{D \text{ and } D' \text{ neighbours}} |q(D) - q(D')|. \tag{15}$$

M_{Lap} is ε-DP and $d_{TV}(M_{\text{Lap}}) = 1 - e^{-\frac{\varepsilon}{2}}$ (derivation in Appendix D). This result was concurrently derived by Jia *et al.* [15] and Chatzikokolakis *et al.* [12].

B. Gaussian Mechanism

The total variation of a Gaussian mechanism operating on a statistic θ with variance $\sigma^2 = \frac{sens(\theta)^2}{\mu^2}$ for $\mu \geq 0$ is

$$d_{TV} = 2\Phi\Big(\frac{\mu}{2}\Big) - 1, \tag{16}$$

where Φ is the Standard Normal CDF. This result was proven by Jia *et al.* [15] and Chatzikokolakis *et al.* [12], and appeared implicitly in [31]. We provide an alternative derivation in Appendix E.

C. Staircase Mechanism

Geng *et al.* [32] studied privacy-utility tradeoffs using differential privacy. They demonstrated that, for a large family of utility functions, the optimal noise-adding mechanism is the "staircase mechanism". That is, the noise is drawn from a staircase-shaped probability distribution with probability density function

$$f_\gamma(x) = \begin{cases} a(\gamma), & x \in [0, \gamma\Delta), \\ e^{-\varepsilon}a(\gamma), & x \in [\gamma\Delta, \Delta), \\ e^{-k\varepsilon}f_\gamma(x - k\Delta), & [k\Delta, (k+1)\Delta),\ k \in \mathbb{N}, \\ f_\gamma(-x), & x < 0. \end{cases}$$

where $a(\gamma) = \dfrac{1 - e^{-\varepsilon}}{2\Delta(\gamma + e^{-\varepsilon}(1-\gamma))}$, $\gamma \in \mathbb{R}^+$. The total variation of the staircase mechanism is given by (derivation in Appendix F1):

$$d_{TV} = \begin{cases} \dfrac{(1-e^{-\varepsilon})(2\gamma(1-e^{-\varepsilon}) + e^{-\varepsilon})}{2(\gamma + e^{-\varepsilon}(1-\gamma))}, & \gamma \in [0, \frac{1}{2}), \\ \dfrac{1 - e^{-\varepsilon}}{2(\gamma + e^{-\varepsilon}(1-\gamma))}, & \gamma \in [\frac{1}{2}, \infty). \end{cases}$$

Remarkably, the staircase mechanism with total variation η achieves exactly the privacy region corresponding to $(\varepsilon, 0)$-DP and η-TV (proof in Appendix F2).

For $\gamma = \frac{1}{2}$, $d_{TV} = \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$, which corresponds to the total variation of the $(\varepsilon, 0)$-differentially private dominating mechanism introduced by Kairouz *et al.* [7], or to that of the mechanism that we introduced in Section III-C for $\alpha = 0$. For $\gamma = 0.0139$, $\alpha \approx 0.3$, the corresponding tradeoff function consists of the blue and red lines in Figure 4a, and therefore the composition exactly corresponds to the blue and red lines in Figure 4b.

D. Differentially Private Stochastic Gradient Descent

We consider the noisy SGD algorithm introduced in [33]. Given a dataset of size n, at every step, noisy SGD generates a uniformly random subsample of size m, computes the corresponding gradient, clips it (in ℓ_2-norm), then adds i.i.d Gaussian noise to each component of the gradient. Dong *et al.* [5] consider and analyze the same setup for comparison purposes, which we adopt here.

Given specific parameters of the noisy SGD algorithm (batch size $m = 256$, learning rate $= 0.25$, and clipping threshold $= 1.5$), Dong *et al.* [5] show that every step satisfies $\frac{1}{1.3}$-Gaussian differential privacy (GDP) [5, Definition 2.6]. For every choice of ε, one obtains a corresponding δ using [5, Corollary 2.13]. In particular, if a mechanism is μ-GDP then it is$(\varepsilon, \delta(\varepsilon))$-DP for

$$\delta(\varepsilon) = \Phi\Big(-\frac{\varepsilon}{\mu} + \frac{\mu}{2}\Big) - e^\varepsilon\Phi\Big(-\frac{-\varepsilon}{\mu} - \frac{\mu}{2}\Big), \tag{17}$$

where Φ is the CDF of the standard normal distribution [5, Corollary 2.13]. As such,

$$\eta = \delta(0) = \Phi\Big(\frac{\mu}{2}\Big) - e^\varepsilon\Phi\Big(-\frac{\mu}{2}\Big). \tag{18}$$

For every choice of (ε, δ), we may then use Theorem 2 to derive an outer bound on the composition of noisy

SGD iterations. Moreover, we consider the intersection of all such regions (for all chosen pairs (ε, δ)). The result is illustrated in Figure 5.

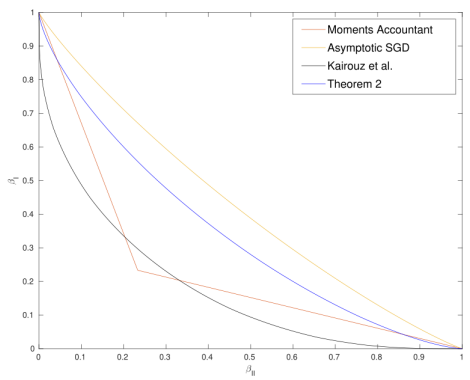


Fig. 5. Comparison of the resulting composition bounds: dataset size $n = 60,000$, batch size $m = 256$, learning rate $= 0.25$, clipping threshold $= 1.5$. Noisy-SGD algorithm running for 15 epochs. Moments accountant refers to the method developed by Abadi *et al.* [33] which yields $(1.19, 10^{-5})$-DP; Asymptotic SGD corresponds to the asymptotic result of [5, Corollary 5.4] (both of these curves were retrieved from [5]); Theorem 2 is applied for all $\varepsilon \in [0.5, 3.4]$ in steps of 0.1 while using η of the mechanism; Kairouz *et al.* [7] is applied for the same values of ε.

As can be seen from the figure, our method yields a significant improvement over both Abadi *et al.*'s as well as Kairouz *et al.*'s result [7], [33]. Although our region is looser than the region of Dong *et al.* [5], their result only holds asymptotically for a specific relationship between the sampling rate and the number of epochs, whereas we obtain a finite-n analysis. It is worth noting that our computations involve an approximation of Theorem 2 to avoid numerical issues. In particular, we used the method of types [22], and approximate the probability of a type using KL divergence [34, Exercise I.2]

VII. Total Variation with Local Differential Privacy

The local privacy setting refers to the case in which there is no "trusted" data curator. Hence, every data owner shares with the curator *noisy* versions of their data. In this setting, local differential privacy has emerged as the main framework to guarantee that the noisy versions are indeed private [9]. Herein, we have "opposing" goals of the data curator and the data owners. For instance, suppose the data corresponds to "web browsing history", then every data owner would like to keep their browsing history private, while the (untrusted) data curator (who is typically offering a service to the data owners) would like to perform statistical analyses on the data. As such, privacy-utility tradeoffs correspond to finding a mechanism that satisfies a given privacy constraint, yet maximizes a certain statistical utility. One can pose a simplified abstract problem as follows: given two distributions P_0 and P_1, and a privatization mechanism Q, let M_0 and M_1 be the induced distributions. Suppose the data curator measures utility through a function $U(M_0||M_1)$. Then we need

$$\max_{Q: Q \text{ satisfies privacy constraint}} U(M_0||M_1). \tag{19}$$

A common choice for U is an f-divergence (e.g., KL, χ^2, etc.). In practice, ε-LDP analysis often lead to pessimistic results [9], [12]. Moreover, the analysis of commonly-used mechanisms (such as M-ary randomized response) may be loose in the LDP framework.

Herein, we study the notion of total variation of a mechanism in the local privacy setting, while maintaining the ε-LDP constraint. This offers more precise results for analyzing mechanisms, and enables greater flexibility in mechanism design (for instance, one may increase ε and decrease η to obtain an acceptable privacy region and improve utility). We mainly focus on f-divergences as utility functions, and study them through the lens of contraction coefficients as done in [11].

A. Preliminaries

Given a convex function $f : [0, \infty) \to \mathbb{R}$ satisfying $f(1) = 0$, the f-divergence between two distributions P_0 and P_1 (over a common alphabet $\mathcal{X}$) is defined by

$$D_f(P_0||P_1) = \mathbf{E}_Q\left[\frac{dP_0}{dP_1}\right], \tag{20}$$

where $\frac{dP_0}{dP_1}$ is the Radon-Nikodym derivative. The family of f-divergences includes commonly used divergences such as: total variation distance (by choosing $f(t) = \max\{0, t-1\}$), KL divergence ($f(t) = t \log t$), and χ^2-divergence ($f(t) = (t-1)^2$). It is well known that f-divergences satisfy the data processing inequality. That is, given P_0, P_1 and a channel $Q_{Y|X}$, let M_0 and M_1 be the induced marginal distributions over $\mathcal{Y}$ respectively, i.e., for all $S \subseteq \mathcal{Y}$,

$$M_\nu(S) \equiv \sum_{x \in \mathcal{X}} Q(S|x) P_\nu(x), \tag{21}$$

where $\nu \in \{0, 1\}$. Then, the data processing inequality states that

$$D_f(M_0||M_1) \leq D_f(P_0||P_1). \tag{22}$$

Contraction coefficients are concerned with strength the above inequality to $D_f(M_0||M_1) \leq \eta_f(Q) D_f(P_0||P_1)$ with $\eta_f(Q) < 1$. To that end, the contraction coefficient for f-divergence of a given channel Q is defined as

$$\eta_f(Q) = \sup_{P_0, P_1 : D_f(P_0||P_1) \neq 0} \frac{D_f(M_0||M_1)}{D_f(P_0||P_1)}. \tag{23}$$

B. Definitions and Preliminary Results

Definition 6 (Local Differential Privacy): For $\varepsilon \geq 0$, a mechanism Q is ε-locally differentially private if

$$\sup_{S \subseteq \mathcal{Y}, x, x' \in \mathcal{X}} \frac{Q(S|x)}{Q(S|x')} \leq e^\varepsilon. \tag{24}$$

The total variation of a mechanism in given by:

Definition 7: Given $\eta \in [0, 1]$, a mechanism $Q_{Y|X}$ is said to be η-TV locally (or $d_{TV}(Q) \leq \eta$ for short) if

$$\max_{x, x'} d_{TV}(Q(.|X = x), Q(.|X = x')) = \eta. \tag{25}$$

In other words, the Dobrushin (contraction) coefficient [35] of Q, typically denoted by $\eta_{TV}(Q)$, is given by η, i.e., $\eta_{TV}(Q) = \eta$. The connection to contraction

coefficients is reminiscent of the recent results on the connection between local differential privacy and the contract coefficients for E_γ divergences [36], [37].

Given $\varepsilon > 0$ and an ε-LDP channel Q, it is known that [7, Remark A.1]

$$d_{TV}(Q) \leq \frac{e^\varepsilon + 1}{e^\varepsilon - 1}. \tag{26}$$

Now given $\varepsilon > 0$ and $0 \leq \eta \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$, define

$$\mathcal{Q}_{\varepsilon,\eta} = \{Q : Q \text{ satisfies } \varepsilon\text{-LDP and } \eta\text{-TV}\}. \tag{27}$$

Given a collection of distributions $P_0, P_1, \ldots, P_{M-1}$, we slightly abuse notation and use the tuple $(P_0, P_1, \ldots, P_{M-1})$ to denote the M-ary channel $Q : \{0, 1, \ldots, M-1\} \to \mathcal{Y}$ defined by $Q(y|i) = P_i(y)$, for all $i \in \{0, 1, \ldots, M-1\}$ and $y \in \mathcal{Y}$.

The distributions defined in equation (7) will play a special role, and will be denoted by $P_0^\star$ and $P_1^\star$. That is,

$$\begin{bmatrix} P_0^\star \\ P_1^\star \end{bmatrix} = \begin{bmatrix} \eta\frac{e^\varepsilon}{e^\varepsilon-1} & \eta\frac{1}{e^\varepsilon-1} & 1-\eta\frac{e^\varepsilon+1}{e^\varepsilon-1} \\ \eta\frac{1}{e^\varepsilon-1} & \eta\frac{e^\varepsilon}{e^\varepsilon-1} & 1-\eta\frac{e^\varepsilon+1}{e^\varepsilon-1} \end{bmatrix}, \tag{28}$$

where for notational convenience, we suppressed the dependence of $P_0^\star$ and $P_1^\star$ on ε and η. The corresponding binary-input channel will be denoted by $Q^\star_{\varepsilon,\eta} : \{0,1\} \to \{0,1,2\}$, i.e., for all $y \in \{0,1,2\}$,

$$Q^\star_{\varepsilon,\eta}(y|0) = P_0(y) \text{ and } Q^\star_{\varepsilon,\eta}(y|1) = P_1(y). \tag{29}$$

$Q^\star_{\varepsilon,\eta}$ will be referred to as the dominating mechanism.

Considering the equivalence with the Dobrushin contraction coefficient, we get the following result:

Proposition 2: Given distributions P_0 and P_1 over a common alphabet $\mathcal{Y}$, $\varepsilon \geq 0$, and $0 \leq \eta \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$,

$$\sup_{Q \in \mathcal{Q}_{\varepsilon,\eta}} d_{TV}(M_0, M_1) = \eta d_{TV}(P_0, P_1). \tag{30}$$

Moreover, the supremum is achieved by the *binary mechanism with erasure* $Q^{(be)} : \mathcal{Y} \to \{0,1,2\}$ defined by

$$Q^{(be)}(.|y) = Q^\star_{\varepsilon,\eta}(.|\mathbf{1}\{P_1(y) > P_0(y)\}), \tag{31}$$

where $\mathbf{1}\{\}$ is the indicator function and $Q^\star_{\varepsilon,\eta}$ is the dominating mechanism defined in equation (29).

This recovers and generalizes a result by Kairouz *et al.* [10, Theorem 6, Corollary 11], which performs the optimization over all ε-LDP channels, i.e., it corresponds to $\eta = \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$ (proof deferred to Appendix G).

The following result will be very useful in the sequel.

Theorem 4: Given a convex function $f : [0, \infty) \to \mathbb{R}_+$ satisfying $f(1) = 0$, $\varepsilon \geq 0$, and $0 \leq \eta \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$,

$$\begin{aligned} \sup_{(P_0,P_1)\in\mathcal{Q}_{\varepsilon,\eta}} D_f(P_0||P_1) &= D_f(P_0^\star||P_1^\star) \\ &= \frac{\eta(f(e^\varepsilon) + e^\varepsilon f(e^{-\varepsilon}))}{e^\varepsilon - 1}. \end{aligned} \tag{32}$$

Denoting by $\mathcal{Q}_\varepsilon$ the set of all ε-LDP mechanisms, one obtains as an immediate corollary that

$$\sup_{(P_0,P_1)\in\mathcal{Q}_\varepsilon} D_f(P_0||P_1) = \frac{1}{e^\varepsilon + 1}\big(f(e^\varepsilon) + e^\varepsilon f\big(e^{-\varepsilon}\big)\big).$$

Proof: Consider two distributions P_0 and P_1 over a common alphabet $\mathcal{Y}$ such that $(P_0, P_1) \in \mathcal{Q}_{\varepsilon,\eta}$. Let $f(P_0, P_1)$ be the corresponding ROC curve (cf. Definition 2), Then, by Corollary 1, $f(P_0, P_1) \geq f(P_0^\star, P_1^\star)$ since $(P_0^\star, P_1^\star)$ achieves equality in Corollary 1. Hence, it follows from Blackwell's theorem [21] that there exists a channel $Q : \{0,1,2\} \to \mathcal{Y}$ such that $P_i = Q \circ P_i^\star$, $i \in \{0,1\}$. Thus it follows from the data processing inequality for f-divergences that

$$D_f(P_0||P_1) = D_f(Q \circ P_0^\star||Q \circ P_1^\star) \leq D_f(P_0^\star||P_1^\star).$$

C. Contraction Coefficient for KL Divergence

Given two distributions P_0 and P_1 over the same alphabet $\mathcal{Y}$, the KL divergence is given $D(P_0||P_1) = \mathbf{E}\Big[\log \frac{dP_0}{dP_1}\Big]$. We provide an exact characterization of the maximum contraction coefficient η_{KL} for channels $Q \in \mathcal{Q}_{\varepsilon,\eta}$.

Theorem 5: Given $\varepsilon \geq 0$ and $0 \leq \eta \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$,

$$\sup_{Q \in \mathcal{Q}_{\varepsilon,\eta}} \eta_{KL}(Q) = \eta_{KL}\big(Q^\star_{\varepsilon,\eta}\big) = \eta\frac{e^\varepsilon - 1}{e^\varepsilon + 1}. \tag{33}$$

Maximizing $\eta_{KL}(Q)$ for ε-LDP channels corresponds to setting η to its maximum value, i.e., $\eta = \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$ and thus

$$\sup_{Q \in \mathcal{Q}_\varepsilon} \eta_{KL}(Q) = \left(\frac{e^\varepsilon - 1}{e^\varepsilon + 1}\right)^2, \tag{34}$$

which recovers Theorem 1 of [11]. Moreover, consider the M-ary randomized response mechanism defined by $Q_R : \{1, 2, \ldots, M\} \to \{1, 2, \ldots, M\}$ satisfying

$$Q_R(j|i) = \begin{cases} \frac{e^\varepsilon}{e^\varepsilon + M - 1}, & i = j \\ \frac{1}{e^\varepsilon + M - 1}, & i \neq j. \end{cases} \tag{35}$$

Then, by Theorem 5 and [38, Theorem 1], we get

$$\eta_{\chi^2}(Q_R) = \eta_{KL}(Q_R) \leq \frac{(e^\varepsilon - 1)^2}{(e^\varepsilon + M - 1)(e^\varepsilon + 1)}. \tag{36}$$

By Proposition 1 of [11], the upper bound is tight, thus highlighting the refined analysis gained by incorporating TV in our analysis. Indeed, the proof of Theorem 5 generalizes (and in fact simplifies) the proof of Asoodeh and Zhang [11, Theorem 1] by utilizing Lemma 4.

Moreover, one can make use of Theorem 5 for a more refined comparison of two privatization mechanisms. For instance, suppose we construct two feasible (according to certain privacy constraints) privatization mechanisms Q_1 and Q_2. If Q_1 is stochastically degraded from Q_2 (i.e., Q_1 can be simulated from Q_2), then for any utility function that satisfies the data processing inequality (e.g., an f-divergence), Q_2 is a better choice (and vice versa if Q_2 is degraded from Q_1). If there is no degradedness relationship between the two, then how do we choose? In this case, one may use $\eta_Q \frac{e^\varepsilon Q - 1}{e^\varepsilon Q + 1}$ as a simple (and generally easily to compute) proxy for the utility function.

Proof of Theorem 5: By Theorem 1 of [39], it suffices to consider binary-input channels $Q = (P_0, P_1)$. Now by (the proof of) Theorem 21 in [40], we have

$$\eta_{KL}((P_0, P_1)) = \sup_{\beta \in (0,1)} LC_\beta(P_0||P_1), \tag{37}$$

where $LC_\beta(P_0||P_1)$ is an f-divergence, known as the Le Cam divergence, and is given by [41]–[43]

$$LC_\beta(P_0||P_1) = \beta(1-\beta)\mathbf{E}_Q\left[\frac{(\frac{dP}{dQ}-1)^2}{\beta\frac{dP}{dQ}+1-\beta}\right]. \quad (38)$$

Therefore,

$$\begin{aligned}\sup_{Q\in\mathcal{Q}_{\varepsilon,\delta}} \eta_{KL}(Q) &= \sup_{(P_0,P_1)\in\mathcal{Q}_{\varepsilon,\delta}} \eta_{KL}((P_0,P_1))\\ &= \sup_{(P_0,P_1)\in\mathcal{Q}_{\varepsilon,\delta}} \sup_{\beta\in(0,1)} LC_\beta(P_0||P_1)\\ &= \sup_{\beta\in(0,1)} \sup_{(P_0,P_1)\in\mathcal{Q}_{\varepsilon,\delta}} LC_\beta(P_0||P_1)\\ &\overset{(a)}{=} \sup_{\beta\in(0,1)} LC_\beta(P_0^\star||P_1^\star)\\ &= \eta(e^\varepsilon-1)\sup_{\beta\in(0,1)}\beta(1-\beta)\cdot\\ &\quad\left(\frac{1}{\beta e^\varepsilon+1-\beta}+\frac{1}{(1-\beta)e^\varepsilon+\beta}\right)\\ &\overset{(b)}{=} \eta\frac{e^\varepsilon-1}{e^\varepsilon+1}, \end{aligned} \quad (39)$$

where (a) follows from Lemma 4, and (b) follows from elementary algebraic manipulations (the supremum is achieved for $\beta = 1/2$). ∎

D. Contraction in terms of TV

A well-known feature of ε-locally DP mechanisms is that, even if the input KL divergence $D(P_0||P_1)$ is infinite, the induced output KL divergence $D(M_0||M_1)$ is finite. This is captured by bounding $D(M_0||M_1)$ in terms of $d_{TV}(P_0||P_1)$ (which is finite by definition). In particular, we have $D(M_0||M_1) \le \log(1+\chi^2(M_0||M_1))$ [44] and

Theorem 6: Given $\varepsilon \ge 0$, $0 \le \eta \le \frac{e^\varepsilon+1}{e^\varepsilon-1}$, and $Q : \mathcal{X} \to \mathcal{Y}$ satisfying $Q \in \mathcal{Q}_{\varepsilon,\eta}$, then for any distributions P_0 and P_1 on $\mathcal{X}$, the induced marginals M_0 and M_1 satisfy

$$\chi^2(M_0||M_1) \le 4\eta(e^\varepsilon-1)(e^{-\varepsilon}+1)d_{TV}^2(P_0,P_1). \quad (40)$$

Considering the maximum value of $\eta = \frac{e^\varepsilon+1}{e^\varepsilon-1}$, this recovers Theorem 2 of [11]. Indeed, once again the proof refines and simplifies the proof of [11].

Proof: As done by Asoodeh and Zhang [11], we first utilize [45, Proposition 8]:

$$\chi^2(M_0||M_1) \le 4d_{TV}^2(P_0,P_1)\max_{x,x'}\chi^2(Q(.|x)||Q(.|x')).$$

Now note that it follows straightforwardly from Definitions 6 and 7 that $Q \in \mathcal{Q}_{\varepsilon,\eta} \Rightarrow (Q(.|x)||Q(.|x')) \in \mathcal{Q}_{\varepsilon,\eta}$ for all $(x,x') \in \mathcal{X}^2$. Hence, it follows from Lemma 4 that

$$\begin{aligned}\chi^2(M_0||M_1) &\le 4d_{TV}^2(P_0,P_1)\max_{(x,x')\in\mathcal{X}^2}\chi^2(P_0^\star||P_1^\star)\\ &= 4\eta(e^\varepsilon-1)(e^{-\varepsilon}+1)d_{TV}^2(P_0,P_1),\end{aligned}$$

where the equality is obtained by plugging $f(t) = (t-1)^2$ in Lemma 4 (and some basic algebraic manipulations). ∎

1) Performance of the Binary Mechanism with Erasure: Fix P_0 and P_1. Let $Q^\star \in \mathcal{Q}_{\varepsilon,\delta}$ be the mechanism maximizing $D(M_0||M_1)$ and denote the induced marginals by $M_0^\star$ and $M_1^\star$. Now consider the binary mechanism with erasure $Q^{(be)}$ (cf. Proposition 2) and let $M_0^{(be)}$ and $M_1^{(be)}$ be the induced marginals. Then, by Theorem 6,

$$\begin{aligned}&4\eta(e^\varepsilon-1)(e^{-\varepsilon}+1)d_{TV}^2(P_0,P_1) \ge \chi^2(M_0^\star||M_1^\star)\\ &\ge D(M_0^\star||M_1^\star) \ge D\Big(M_0^{(be)}||M_1^{(be)}\Big)\\ &\overset{(a)}{\ge} 2d_{TV}^2\Big(M_0^{(be)}||M_1^{(be)}\Big) \overset{(b)}{=} 2\eta^2 d_{TV}^2(P_0||P_1),\end{aligned}$$

where (a) follows from Pinsker's inequality and (b) follows from Proposition 2. As such,

$$\frac{D\Big(M_0^{(be)}||M_1^{(be)}\Big)}{D(M_0^\star||M_1^\star)} \ge \frac{\eta}{2(e^\varepsilon-1)(e^{-\varepsilon}+1)}. \quad (41)$$

Consider the high-privacy regime where $\varepsilon \ll 1$, and suppose $\eta \ge c\varepsilon$ for some $c \in (0,1/2)$. Then the right-hand side of (41) goes to $\frac{c}{4}$ as ε goes to 0. That is, for the high-privacy regime, the binary mechanism with erasure is order-optimal. It is worth noting that the binary mechanism in the local DP setting (where no TV constraint is imposed) is optimal for all ε's below some threshold $\varepsilon^\star$ [10, Theorem 5].

E. Conversion from Local Differential Privacy

Consider the results in Proposition 2, and Theorems 5 and 6. By comparison with the pure local DP counterparts, we notice that in all three cases, the relationship is a multiplicative factor equal to $\eta\frac{e^\varepsilon+1}{e^\varepsilon-1}$. We show that an inequality of this form hold more generally.

In particular, fix two distributions P_0 and P_1 on $\mathcal{X}$ and a convex function f satisfying $f(1) = 0$. Let,

$$\mathrm{OPT}_{\varepsilon,\eta} = \max_{Q\in\mathcal{Q}_{\varepsilon,\eta}} D_f(M_0||M_1),$$

and

$$\mathrm{OPT}_{\varepsilon} = \max_{Q\in\mathcal{Q}_{\varepsilon}} D_f(M_0||M_1).$$

Clearly, $\mathrm{OPT}_\varepsilon \ge \mathrm{OPT}_{\varepsilon,\eta}$. Our next result proves a reverse inequality:

Theorem 7: For any f-divergence (where f is convex and satisfies $f(1) = 0$), $\varepsilon \ge 0$, and $\eta \in [0, \frac{e^\varepsilon-1}{e^\varepsilon+1}]$,

$$\mathrm{OPT}_{\varepsilon,\eta} \ge \eta\frac{e^\varepsilon+1}{e^\varepsilon-1}\mathrm{OPT}_\varepsilon. \quad (42)$$

The proof relies on the following observation: given a channel $Q \in \mathcal{Q}_\varepsilon$, then composing Q with an erasure channel W with parameter α yields a channel $W \circ Q \in \mathcal{Q}_{\varepsilon,(1-\alpha)\frac{e^\varepsilon-1}{e^\varepsilon+1}}$. The details are deferred to Appendix H.

REFERENCES

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography: 3rd Theory of Cryptography Conference, New York, NY, USA, Mar. 4-7, 2006.* Springer, 2006, pp. 265–284.

[2] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2006, pp. 486–503.

[3] I. Mironov, "Rényi differential privacy," in *IEEE 30th Computer Security Foundations Symposium*. IEEE, 2017, pp. 263–275.

[4] C. Dwork and G. N. Rothblum, "Concentrated differential privacy," *arXiv preprint arXiv:1603.01887*, 2016.

[5] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," 2019. [Online]. Available: https://arxiv.org/abs/1905.02383

[6] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 375–389, 2010.

[7] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *International conference on machine learning*. PMLR, 2015, pp. 1376–1385.

[8] Q. Geng, P. Kairouz, S. Oh, and P. Viswanath, "The staircase mechanism in differential privacy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1176–1184, 2015.

[9] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy, data processing inequalities, and statistical minimax rates," 2014.

[10] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," vol. 17, no. 17, 2016, pp. 1–51.

[11] S. Asoodeh and H. Zhang, "Contraction of locally differentially private mechanisms," *arXiv preprint arXiv:2210.13386*, 2022.

[12] K. Chatzikokolakis, G. Cherubin, C. Palamidessi, and C. Troncoso, "Bayes security: A not so average metric," in *IEEE 36th Computer Security Foundations Symposium*, 2023, pp. 388–406.

[13] R. F. Barber and J. C. Duchi, "Privacy and statistical risk: Formalisms and minimax bounds," 2014.

[14] Q. Geng, W. Ding, R. Guo, and S. Kumar, "Optimal noise-adding mechanism in additive differential privacy," in *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 11–20.

[15] J. Jia, C. Tan, Z. Liu, X. Li, Z. Liu, S. Lv, and C. Dong, "Total variation distance privacy: Accurately measuring inference attacks and improving utility," *Information Sciences*, vol. 626, pp. 537–558, 2023.

[16] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. New York, NY, USA: Association for Computing Machinery, 2012, p. 214–226.

[17] R. Bassily, K. Nissim, A. Smith, T. Steinke, U. Stemmer, and J. Ullman, "Algorithmic stability for adaptive data analysis," in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery, 2016, p. 1046–1059.

[18] M. Raginsky, A. Rakhlin, M. Tsao, Y. Wu, and A. Xu, "Information-theoretic analysis of stability and bias of learning algorithms," in *IEEE Information Theory Workshop*, 2016.

[19] M. Yaghini, B. Kulynych, and C. Troncoso, "Disparate vulnerability: on the unfairness of privacy attacks against machine learning," *CoRR*, vol. abs/1906.00389, 2019.

[20] B. Kulynych, Y.-Y. Yang, Y. Yu, J. Błasiok, and P. Nakkiran, "What you see is what you get: Principled deep learning via distributional generalization," vol. 35, 2022, pp. 2168–2183.

[21] D. Blackwell, "Equivalent comparisons of experiments," *The annals of mathematical statistics*, pp. 265–272, 1953.

[22] T. M. Cover and J. A. Thomas, "Elements of information theory 2nd edition," *Willey-Interscience: NJ*, 2006.

[23] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.

[24] A. Smith, "Differential privacy and the secrecy of the sample," Sep 2009. [Online]. Available: https://adamdsmith.wordpress.com/2009/09/02/sample-secrecy/

[25] K. Chaudhuri and N. Mishra, "When random sampling preserves privacy," in *Annual International Cryptology Conference*. Springer, 2006, pp. 198–213.

[26] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," *Advances in neural information processing systems*, vol. 31, 2018.

[27] T. Steinke, "Composition of differential privacy & privacy amplification by subsampling," *arXiv preprint arXiv:2210.00597*, 2022.

[28] A. Koskela, J. Jälkö, L. Prediger, and A. Honkela, "Tight differential privacy for discrete-valued mechanisms and for the subsampled gaussian mechanism using FFT," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 3358–3366.

[29] S. Gopi, Y. T. Lee, and L. Wutschitz, "Numerical composition of differential privacy," vol. 34, 2021, pp. 11 631–11 642.

[30] Y. Zhu, J. Dong, and Y.-X. Wang, "Optimal accounting of differential privacy via characteristic function," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 4782–4817.

[31] B. Balle and Y.-X. Wang, "Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising," in *International Conference on Machine Learning*. PMLR, 2018, pp. 394–403.

[32] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 2371–2375.

[33] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[34] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning (Draft)*. Cambridge University Press, 2023.

[35] R. L. Dobrushin, "Central limit theorem for nonstationary markov chains. i," *Theory of Probability & Its Applications*, vol. 1, no. 1, pp. 65–80, 1956.

[36] S. Asoodeh, M. Aliakbarpour, and F. P. Calmon, "Local differential privacy is equivalent to contraction of an f-divergence," in *IEEE International Symposium on Information Theory*. IEEE, 2021, pp. 545–550.

[37] B. Zamanlooy and S. Asoodeh, "Strong data processing inequalities for locally differentially private mechanisms," in *IEEE International Symposium on Information Theory*. IEEE, 2023, pp. 1794–1799.

[38] M.-D. Choi, M. B. Ruskai, and E. Seneta, "Equivalence of certain entropy contraction coefficients," *Linear algebra and its applications*, vol. 208, pp. 29–36, 1994.

[39] O. Ordentlich and Y. Polyanskiy, "Strong data processing constant is achieved by binary inputs," *IEEE Transactions on Information Theory*, vol. 68, no. 3, pp. 1480–1481, 2021.

[40] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and bayesian networks," in *Convexity and Concentration*. Springer, 2017, pp. 211–249.

[41] L. Le Cam, *Asymptotic methods in statistical decision theory*. Springer Science & Business Media, 2012.

[42] L. Györfi and I. Vajda, "A class of modified pearson and neyman statistics," *Statistics & Risk Modeling*, vol. 19, no. 3, pp. 239–252, 2001.

[43] M. Raginsky, "Strong data processing inequalities and Φ-sobolev inequalities for discrete channels," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3355–3389, 2016.

[44] I. Sason and S. Verdú, "f-divergence inequalities," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 5973–6006, 2016.

[45] J. C. Duchi and F. Ruan, "The right complexity measure in locally private estimation: It is not the fisher information," *arXiv preprint arXiv:1806.05756*, 2018.

APPENDIX

A. Proof for Approximate Differential Privacy

Generalizing the result to approximate differential privacy requires the following (dominating) mechanism, which outputs an integer sampled from P_0 when the true database is D_0 (null hypothesis) and from P_1 when the true database is D_1 (alternative hypothesis). For $\varepsilon \geq 0$, $\delta > 0$, and $\alpha \in [0, 1]$, define:

$$P_0 = \begin{cases} \delta & \text{for } x = 0, \\ \frac{(1-\delta)(1-\alpha)e^{\varepsilon}}{(1+e^{\varepsilon})} & \text{for } x = 1, \\ \alpha(1-\delta) & \text{for } x = 2, \\ \frac{(1-\delta)(1-\alpha)}{(1+e^{\varepsilon})} & \text{for } x = 3, \\ 0 & \text{for } x = 4, \end{cases} \tag{43}$$

and

$$P_1 = \begin{cases} 0 & \text{for } x = 0, \\ \frac{(1-\delta)(1-\alpha)}{(1+e^{\varepsilon})} & \text{for } x = 1, \\ \alpha(1-\delta) & \text{for } x = 2, \\ \frac{(1-\delta)(1-\alpha)e^{\varepsilon}}{(1+e^{\varepsilon})} & \text{for } x = 3, \\ \delta & \text{for } x = 4. \end{cases} \tag{44}$$

The total variation of the above mechanism is $d_{TV} = \delta + \frac{(1-\delta)(1-\alpha)(e^{\varepsilon}-1)}{1+e^{\varepsilon}}$. Setting $\alpha = 0$ yields the same mechanism proposed in [7].

The proof is analogous to case of $\delta = 0$ in Section III-C2. To see how the expression $\max_{S \in \mathcal{X}^k}\{\tilde{P}_0(S) - e^{\varepsilon_j}\tilde{P}_1(S)\} = 1 - (1-\delta)^k(1-\delta_j)$ in Theorem 2 was derived, we split it into two parts $(p_1 + p_2)$, such that $p_1 = 1 - (1-\delta)^k$ and $p_2 = (1-\delta)^k\delta_j$.
We are interested in the sequences s_k that achieve the inequality $\tilde{P}_0(s_k) - e^{\tilde{\varepsilon}}\tilde{P}_1(s_k) > 0$:

- p_1: This inequality is satisfied for all output sequences in which $x = 0$ appears at least once, because the probability of such a sequence under P_1 is always 0. The probability of obtaining a sequence with no "0"s under P_0 is $(1-\delta)^k$. Taking the complement to obtain the probability that a sequence contains at least one "0" under P_0, we get $p_1 = 1 - (1-\delta)^k$.
- p_2: We now consider the sequences that do not contain a "0" and that lead to a positive $\tilde{P}_0(s_k) - e^{\tilde{\varepsilon}}\tilde{P}_1(s_k)$. We apply the same logic that we followed in the case of $\delta = 0$. Factoring out $(1-\delta)^k$ from the result, we obtain $p_2 = (1-\delta)^k\delta_j$.

B. Proof of Theorem 3

The theorem is an application of Dong *et al.*'s result, in particular Theorem 3.5 in "Gaussian Differential Privacy" [5], specialized for ε-DP and η-TV. Note that, by definition, M_{ni} is $f_{\varepsilon_{ni},0,\eta_{ni}}$-DP. Now, following the definitions in [5],

$$\begin{aligned} kl(f_{\varepsilon,0,\eta}) &:= -\int_0^1 \log\left|f'_{\varepsilon,0,\eta}(t)\right| dt \\ &= -\left(\int_0^{\frac{\eta}{e^{\varepsilon}-1}} \varepsilon dt + \int_{1-\eta\frac{e^{\varepsilon}}{e^{\varepsilon}+1}}^1 (-\varepsilon) dt\right) \\ &= \varepsilon\eta. \end{aligned}$$

Similarly,

$$\kappa_2(f_{\varepsilon,0,\eta}) := \int_0^1 \log^2\left|f'_{\varepsilon,0,\eta}(t)\right| dt = \eta\varepsilon^2\frac{e^{\varepsilon}+1}{e^{\varepsilon}-1},$$

and

$$\kappa_3(f_{\varepsilon,0,\eta}) := \int_0^1 \left|\log\left|f'_{\varepsilon,0,\eta}(t)\right|\right|^3 dt = \eta\varepsilon^3\frac{e^{\varepsilon}+1}{e^{\varepsilon}-1}.$$

Now,

$$\sum_{i=1}^n kl(f_{\varepsilon_{ni},0,\eta_{ni}}) = \sum_{i=1}^n \varepsilon_{ni}\eta_{ni} \to \frac{\mu^2}{2}, \tag{45}$$

where the limit holds by assumption of the theorem. Similarly,

$$\max_{1\leq i\leq n} kl(f_{\varepsilon_{ni},0,\eta_{ni}}) = \max_i \eta_{ni}\varepsilon_{ni} \leq \max_i \varepsilon_{ni} \to 0, \tag{46}$$

where the limit holds by assumption. Since $kl(f) \geq 0$, then $\max_{1\leq i\leq n} kl(f_{\varepsilon_{ni},0,\eta_{ni}}) \to 0$. Moreover, note that

$$\begin{aligned} &\sum_{i=1}^n \left|\frac{1}{2}\kappa_2(f_{\varepsilon_{ni},0,\eta_{ni}}) - kl(f_{\varepsilon_{ni},0,\eta_{ni}})\right| \\ &= \sum_{i=1}^n \varepsilon_{ni}\eta_{ni}\left|\frac{\varepsilon_{ni}(e^{\varepsilon}_{ni}+1)}{2(e^{\varepsilon}_{ni}-1)} - 1\right| \\ &\leq \sum_{i=1}^n \varepsilon_{ni}\eta_{ni}\varepsilon^2_{ni} \\ &\leq \max_{1\leq i\leq n} \varepsilon^2_{ni} \sum_{i=1}^n \varepsilon_{ni}\eta_{ni} \to 0, \end{aligned}$$

where the first inequality follows from the Taylor expansion of $t(e^t+1)/(e^t-1)$. Hence,

$$\sum_{i=1}^n \kappa_2(f_{\varepsilon_{ni},0,\eta_{ni}}) \to \mu^2. \tag{47}$$

Similarly, one can show that $\sum_{i=1}^n \kappa_3(f_{\varepsilon_{ni},0,\eta_{ni}}) \to 0$. Coupled with equations (45), (46), and (47), the assumptions of Theorem 3.5 of [5] are satisfied, hence the limit of the composition converges uniformly to G_μ.

C. Proof of Proposition 1

Since M is (ε, δ)-DP then $M \circ \text{Sample}_m$ is $(\log(1 + p(e^{\varepsilon} - 1)), p\delta)$-DP by Theorem 29 of [27]. Similarly, M being η-TV can be rewritten as $(0, \eta)$-DP (see, for instance, equations of Corollary 1). Hence, $M \circ \text{Sample}_m$ is (0,$p\eta$)-DP, i.e., it is $p\eta$-TV, as desired. Although for (ε, δ)-pair, the result is known to be tight, it is not clear a priori whether the result is tight *simultaneously* for both constraints. We show that is tight by exhibiting a mechanism that achieves the region. Consider two neighboring databases D_0 and D_1 that differ on one entry. Say there exists an element $a \in D_0$ but $a \notin D_1$. Consider the following mechanism:

$$M(D) \sim P_0, \text{ if } a \in D$$
$$M(D) \sim P_1, \text{ if } a \notin D,$$

where P_0 and P_1 are defined in equation (7). Then, after subsampling with ratio p, the induced distributions will be as follows:

$$M(D) \sim pP_0 + (1-p)P_1, \text{ if } a \in D \tag{48}$$
$$M(D) \sim P_1, \text{ if } a \notin D. \tag{49}$$

One can explicitly compute the corresponding privacy region, and obtain the desired result by considering the worst-case among the above region and the one induced (by symmetry) from the pair of distributions

$$pP_1 + (1-p)P_0 \text{ and } P_0. \tag{50}$$

D. Total Variation of the Laplace Mechanism

To determine the total variation of the Laplace mechanism, consider two Laplace distributions, one centered at zero ($P : Lap(\frac{1}{\varepsilon}, 0)$), and one shifted to the right by the sensitivity ($Q : Lap(\frac{1}{\varepsilon}, \Delta)$).

$$\begin{aligned} d_{TV}(P,Q) &= \int_{x:P(x)>Q(x)} P(x) - Q(x)\, dx \\ &= \int_{-\infty}^{\Delta/2} \frac{\varepsilon}{2\Delta} e^{-\frac{\varepsilon|x|}{\Delta}} - \frac{\varepsilon}{2\Delta} e^{-\frac{\varepsilon|x-\Delta|}{\Delta}}\, dx \\ &= \int_{-\infty}^{0} \frac{\varepsilon}{2\Delta}(e^{\varepsilon x} - e^{-\varepsilon(\Delta - x)})\, dx \\ &\quad + \int_{0}^{\frac{\Delta}{2}} \frac{\varepsilon}{2\Delta}(e^{-\varepsilon x} - e^{-\varepsilon(\Delta - x)})\, dx \\ &= 1 - e^{-\frac{\varepsilon}{2}}. \end{aligned}$$

E. Total Variation of the Gaussian Mechanism

Corollary 2.13 in [5] states that a mechanism is μ-GDP if and only if it is (ε,$\delta(\varepsilon)$)-DP for all $\varepsilon \geq 0$, where $\delta(\varepsilon) = \Phi\big(-\frac{\varepsilon}{\mu} + \frac{\mu}{2}\big) - e^{\varepsilon}\Phi\big(-\frac{\varepsilon}{\mu} - \frac{\mu}{2}\big)$. Thus, the total variation distance of the Gaussian mechanism is

$$\delta(0) = \Phi\big(\frac{\mu}{2}\big) - \Phi\big(-\frac{\mu}{2}\big) = 2\Phi\big(\frac{\mu}{2}\big) - 1. \tag{51}$$

F. Analysis of the Staircase Mechanism

1) Total Variation: To derive the total variation of the staircase mechanism, we consider two distributions: $P(x)$ centered at 0, and $Q(x)$ which is $P(x)$ shifted to the right by the sensitivity Δ. We obtain the total variation of the staircase mechanism by subtracting the areas under each distribution for the values of x where $P(x) > Q(x)$. We distinguish the cases of $\gamma \leq \frac{1}{2}$ and $\gamma > \frac{1}{2}$. For $\gamma \leq \frac{1}{2}$, we are interested in the interval $(-\infty, \gamma]$ which we split into two sub-intervals (Figure 6):

- For $(-\infty, -\gamma]$, the difference between the area under the red curve and that under the blue curve is given by $a(\gamma)\Delta\sum_{n=1}^{\infty}(e^{-n\varepsilon} - e^{-n\varepsilon-\varepsilon}) = a(\gamma)\Delta e^{-\varepsilon}$
- For $(-\gamma, \gamma]$, the difference is $a(\gamma)(2\gamma\Delta(1-e^{\varepsilon}))$

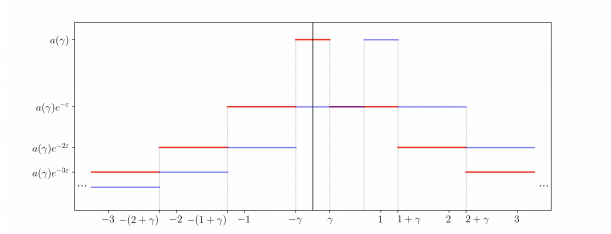


Fig. 6. Two staircase distributions for $\gamma \leq \frac{1}{2}$ shifted by the sensitivity.

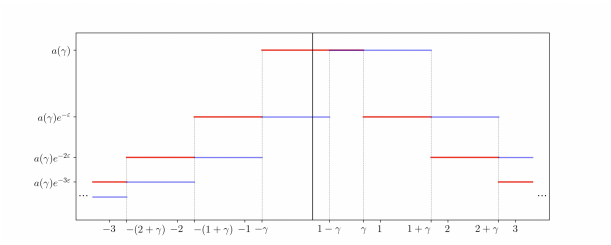


Fig. 7. Two staircase distributions for $\gamma > \frac{1}{2}$ shifted by the sensitivity.

For $\gamma > \frac{1}{2}$, we look at the interval $(-\infty, 1-\gamma)$ (Figure 7). The total variation of the mechanism is given by $a(\gamma)\Delta\sum_{n=0}^{\infty}(e^{-n\varepsilon} - e^{-n\varepsilon-\varepsilon}) = a(\gamma)\Delta$.

Considering two intervals for the values of γ and setting $\Delta = 1$, we derive the following expressions for the total variation of the staircase mechanism

$$d_{TV} = \begin{cases} \dfrac{(1-e^{-\varepsilon})(2\gamma(1-e^{-\varepsilon}) + e^{-\varepsilon})}{2(\gamma + e^{-\varepsilon}(1-\gamma))}, & \gamma \in [0, \frac{1}{2}), \\ \dfrac{1-e^{-\varepsilon}}{2(\gamma + e^{-\varepsilon}(1-\gamma))}, & \gamma \in [\frac{1}{2}, \infty). \end{cases}$$

2) Privacy Region: The staircase mechanism with total variation η achieves exactly the privacy region corresponding to $(\varepsilon, 0)$-DP η-TV (cf. Corollary 1). Consider two staircase-shaped distributions, P_0 and P_1 (where P_1 is P_0 shifted by Δ) for some arbitrary γ. The Neyman-Pearson Lemma states that the optimal decision rule is to reject when the likelihood ratio is above some threshold T. Since the likelihood ratio of P_1 to P_0 is non-decreasing (cf. Figures 6 and 7 in Appendix F1), we can pick a point t and accept whenever $t > x$. The rejection rules $x \geq 1-\gamma$ and $x \geq \gamma$ will yield type I and II values that correspond to the points $(\frac{\eta}{e^{\varepsilon}-1}, 1 - \frac{\eta e^{\varepsilon}}{e^{\varepsilon}-1})$ and $(1 - \frac{\eta e^{\varepsilon}}{e^{\varepsilon}-1}, \frac{\eta}{e^{\varepsilon}-1})$ in Figure 2 for $\delta = 0$. The rest of the tradeoff function is derived by averaging. Therefore, the bound provided in Theorem 2 exactly describes the privacy region of the composed staircase mechanism.

G. Proof of Proposition 2

Recall Dobrushin's classical result [35]:

$$\sup_{\substack{P_0, P_1 \\ P_0 \neq P_1}} \frac{||M_0 - M_1||_{\text{TV}}}{||P_0 - P_1||_{\text{TV}}} = \sup_{x, x'} d_{TV}\big(Q(.|x), Q(.|x')\big). \tag{52}$$

As such,

$$\|M_0 - M_1\|_{TV} \leq \eta_{TV}(Q)\|P_0 - P_1\|_{TV} = \eta\|P_0 - P_1\|,$$

where the inequality follows from (52), and the equality follows from Definition 7. It remains to show that the binary mechanism with erasure achieves

$$||M_0 - M_1||_{\text{TV}} = \eta\, ||P_0 - P_1||_{\text{TV}}. \tag{53}$$

Let $\mathcal{A} \subseteq \mathcal{X}$ such that $\mathcal{A} = \{x \in \mathcal{X} : P_0(x) \geq P_1(x)\}$. Hence

$$\|P_0 - P_1\|_{TV} = P_0(A) - Q_0(A). \tag{54}$$

Now note that

$$M_\nu(0) = \sum_{x\in\mathcal{A}} \eta\frac{e^\varepsilon}{e^\varepsilon-1}P_\nu(x) + \sum_{x\notin\mathcal{A}} \eta\frac{1}{e^\varepsilon-1}P_\nu(x),$$
$$M_\nu(1) = \sum_{x\in\mathcal{A}} \eta\frac{1}{e^\varepsilon-1}P_\nu(x) + \sum_{x\notin\mathcal{A}} \eta\frac{e^\varepsilon}{e^\varepsilon-1}P_\nu(x),$$
$$M_\nu(e) = \left(1-\eta\frac{1+e^\varepsilon}{e^\varepsilon-1}\right)\sum_x P_\nu(x) = \left(1-\eta\frac{1+e^\varepsilon}{e^\varepsilon-1}\right).$$

Therefore,

$$\begin{aligned}
&|M_0(0) - M_1(0)| \\
&= \sum_{x\in\mathcal{A}} \eta\frac{e^\varepsilon}{e^\varepsilon-1}P_0(x) + \sum_{x\notin\mathcal{A}} \eta\frac{1}{e^\varepsilon-1}P_0(x) \\
&\quad - \sum_{x\in\mathcal{A}} \eta\frac{e^\varepsilon}{e^\varepsilon-1}P_1(x) - \sum_{x\notin\mathcal{A}} \eta\frac{1}{e^\varepsilon-1}P_1(x) \\
&= \eta\frac{e^\varepsilon}{e^\varepsilon-1}\sum_{x\in\mathcal{A}}(P_0(x)-P_1(x)) \\
&\quad + \eta\frac{1}{e^\varepsilon-1}\sum_{x\notin\mathcal{A}}(P_0(x)-P_1(x)) \\
&= \eta\frac{e^\varepsilon}{e^\varepsilon-1}||P_0-P_1||_{TV} - \eta\frac{1}{e^\varepsilon-1}||P_0-P_1||_{TV} \\
&= \eta||P_0-P_1||_{TV}
\end{aligned}$$

Since $|M_0(0) - M_1(0)| = |M_0(1) - M_1(1)|$, and $|M_0(e) - M_1(e)| = 0$, we get

$$\|M_0 - M_1\|_{TV} = |M_0(0) - M_1(0)| = \eta\|P_0 - P_1\|_{TV}.$$

H. Proof of Theorem 7

Let $Q_{Y|X}$ be ε-LDP. Let $W_\alpha : \mathcal{Y} \to \mathcal{Y}\cup\{e\}$ be an erasure channel with parameter α, i.e., for all $y \in \mathcal{Y}$, $W(y|y) = 1-\alpha$ and $W(e|y) = \alpha$.

Let $Q'_{Y|X} = W \circ Q_{Y|X}$. Note that for $y \in \mathcal{Y}$, $Q'(y|x) = Q(y|x)(1-\alpha)$ for all $x \in \mathcal{X}$, and $Q(e|x) = \sum_{y\in\mathcal{Y}} Q(y|x)W(e|y) = \alpha$ for all $x \in \mathcal{X}$.

Lemma 3: $d_{TV}(Q'_{Y|X}) \le (1-\alpha)\frac{e^\varepsilon-1}{e^\varepsilon+1}$.

Proof: Fix x, x', and $A \subset \mathcal{Y}\cup\{e\}$. Then,

$$\begin{aligned}
Q'(A|x) - Q'(A|x') &= \sum_{y'\in A} Q'(y|x) - Q'(y|x') \\
&= \sum_{y'\in A\setminus\{e\}} (1-\alpha)(Q(y|x) - Q(y|x')) \\
&\le (1-\alpha)\|Q(.|x) - Q(.|x')\|_{TV}.
\end{aligned}$$

Taking supremum over all x, x', and A, and noting that $d_{TV}(Q) \le \frac{e^\varepsilon-1}{e^\varepsilon+1}$ (by equation (26)), we get the desired result. ∎

As such, choosing $\alpha = 1-\eta\frac{e^\varepsilon+1}{e^\varepsilon-1}$, we get Q' satisfying $d_{TV}(Q') \le \eta$. It is straightforward to check that Q' also satisfies ε-LDP so that Q' is feasible for the $\mathrm{OPT}_{\varepsilon,\eta}$ problem. Let M'_0 and M'_1 be the induced marginals. Then,

$$M'_0(e) = \sum_{x\in\mathcal{X}} P_0(x)Q'(e|x) = \alpha, \tag{55}$$

and for all $y \in \mathcal{Y}$,

$$\begin{aligned}
M'_0(y) &= \sum_{x\in\mathcal{X}} P_0(x)Q'(y|x) \\
&= \sum_{x\in\mathcal{X}} P_0(x)(1-\alpha)Q(y|x) = (1-\alpha)M_0(y).
\end{aligned}$$

Analogous equations hold for M'_1. Hence,

$$\begin{aligned}
&D_f(M'_0||M'_1) \\
&= \sum_{y'\in\mathcal{Y}\cup\{e\}} M'_1(y) f\left(\frac{M'_0(y)}{M'_1(y)}\right) \\
&= \sum_{y\in\mathcal{Y}} (1-\alpha)M_1(y) f\left(\frac{M_0(y)}{M_1(y)}\right) + M_1(e) f\left(\frac{M_0(e)}{M_1(e)}\right) \\
&\overset{(a)}{=} (1-\alpha)D_f(M_0||M_1) \\
&\overset{(b)}{=} \eta\frac{e^\varepsilon+1}{e^\varepsilon-1}D_f(M_0||M_1),
\end{aligned}$$

where (a) follows from the fact that $f(1) = 0$, and (b) follows from the choice of $\alpha = 1-\eta\frac{e^\varepsilon+1}{e^\varepsilon-1}$. Finally, taking supremum on both sides yields

$$\mathrm{OPT}_{\varepsilon,\eta} \ge \eta\frac{e^\varepsilon+1}{e^\varepsilon-1}\mathrm{OPT}_\varepsilon. \tag{56}$$