A Rate-Disortion Perspective on Local Differential Privacy

Anand D. Sarwate
Rutgers, The State University of New Jersey
asarwate@ece.rutgers.edu

Lalitha Sankar
Arizona State University
lalithasankar@asu.edu

***Abstract*—Local differential privacy is a model for privacy in which an untrusted statistician collects data from individuals who mask their data before revealing it. While randomized response has shown to be a good strategy when the statistician's goal is to estimate a parameter of the population, we consider instead the problem of locally private data publishing, in which the data collector must publish a version of the data it has collected. We model utility by a distortion measure and consider privacy mechanisms that act via a memoryless channnel operating on the data. If we consider a the source distribution to be unknown but in a class of distributions, we arrive at a robust-rate distortion model for the privacy-distortion tradeoff. We show that under Hamming distortions, the differential privacy risk is lower bounded for all nontrivial distortions, and that the lower bound grows logarithmically in the alphabet size.**

I. Introduction

Data sharing is a major challenge for many organizations, and especially for government organizations tasked with providing access to "official statistics." The United States Census Bureau is a canonical example of this: information collected about the population must be shared with the public. Another example is public health monitoring: policies and regulations may require hospitals to share information about their patient population to a regulator. Public health officials receiving such data can help guide the provisioning of health care resources to improve patient outcomes. Finally, regional transmission organizations (RTOs) in the power grid may wish to share information about loads and uses in their networks to facilitate more efficient monitoring and control of the power grid.

In some cases this release may take the form of data derivatives such as histograms, contingency tables, or other statistical summaries of the data. However, in many applications, sharing the raw data, or a version thereof, is necessary for subsequent processing. Sharing data allows for more open-ended investigations and analyses that are not possible with data derivatives alone. However, sharing the data or derivatives raises the danger of privacy violations, especially when the data is sensitive or proprietary. By looking at the data it is often possible to *re-identify* individuals using either identifiers in the data or linking the shared data with auxiliary information [1], [2].

There have been several approaches to mathematically formalizing privacy. In this paper we will examine the relationship between two of these: information-theoretic privacy [3] and differential privacy [4]. In both of these frameworks sharing data is done by a *sanitizer*, which is a function that takes a database X^n of n individuals X_i whose data lies in a set $\mathcal{X}$ and publishes an approximation $\hat{X}^n$ to that database. In this paper we study the relationship between these two definitions of privacy by studying sanitizers that take the form of a channel $Q(\hat{x}|x)$ between an individual data point in the input and its sanitized output. In particular, we study these definitions through the lens of rate distortion theory.

Information theoretic privacy measures privacy in terms of information *leakage*, which is defined as the mutual information $I\left(X^n; \hat{X}^n\right)$. The interpretation is that the output of the sanitizer "leaks" $I\left(X^n; \hat{X}^n\right)$ bits of information about the input. In the basic model, the assumption is that the data X^n is an independent and identically distributed (i.i.d.) sample from a distribution $P(x)$ on $\mathcal{X}$. The goal is to produce a sanitization $\hat{X}^n$ such that the distortion $\sum_{i=1}^{n} d(X_i, \hat{X}_i) \leq nD$, where $d(\cdot, \cdot)$ is a single-letter distortion measure. Thus the information-theoretic version of privacy is a standard rate-distortion problem where the rate is the privacy leakage.

In differential privacy, there is no stochastic assumption on the database, so that data is an individual sequence x^n as opposed to a random variable X^n. However, the mapping $x^n \to \hat{X}^n$ is a random mapping, and privacy is measured as the largest additive gap ϵ between the log likelihood functions for this mapping. In the rate-distortion setting, the goal is to find a single $Q(\hat{x}|x)$ that guarantees a distortion less than D while maintaining a small additive gap.

The main difference between these two privacy models is in the assumptions on the source data; the information-theoretic model assumes we know P and the differential privacy model makes no assumptions on the distribution. To bring these two models closer together, we consider modified versions of the problem. For the information-theoretic problem we assume the source distribution P is unknown but lies in a known set $\mathcal{P}$, and for the differential privacy problem we assume the type of x^n is known to lie in the set $\mathcal{P}$. Other researchers have also studied rate-distortion approaches for differential privacy [5], as well as the structure of optimal mechanisms for function computation [6]–[8]. An interpretation via hypothesis testing forms another information-theoretic understanding of differential privacy [9], [10]. Our work differs from these by considering the modeling issues raised by the knowledge of $\mathcal{P}$ and how it influences the choice of the privacy mechanism and the resulting achievable distortion.

Since the goal is to choose a channel $Q(\hat{x}|x)$, we can think of both privacy definitions in the rate-distortion setting as minimizing the distortion subject to a privacy constraint. That is, given a certain privacy level (measured in terms of leakage or gap), we can find a corresponding set of channels that guarantees that privacy level. We then choose the channel in that set that achieves the lowest distortion possible universally over the set $\mathcal{P}$. Under this formulation, we can ask the following questions: what privacy level ε can be achieved subject to an expected distortion guarantee D that holds universally over $\mathcal{P}$? Contrariwise, for a given privacy level ε, what distortion D Is achievable universally over $\mathcal{P}$? For binary channels under Hamming distortion we characterize this tradeoff and show a surprising result: the minimum privacy risk ε experiences a discontinuity as D approaches its maximum value. This implies that there is a minimum privacy risk for releasing *any* useful version $\hat{x}^n$ of the data. This risk depends on the class of sources $\mathcal{P}$.

II. Problem models

Notation: random variables will be denoted by capital letters and individual sequences or realizations by lower case letters. The shorthand x^n indicates a length-n sequence $(x_1, x_2, \ldots, x_n)$, and x_{-i} indicates $(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, the sequence with the i-th term removed. For a probability distribution $P(x)$ on a space $\mathcal{X}$ and a conditional distribution $Q(\hat{x}|x)$ from a space $\mathcal{X}$ to $\hat{\mathcal{X}}$ the expression $P \times Q$ denotes the joint distribution $P(x)Q(\hat{x}|x)$ on $\mathcal{X} \times \hat{\mathcal{X}}$.

A. Rate distortion

Our investigation in this paper is motivated by the following lossy compression problem. Let $\mathcal{X}$ and $\hat{\mathcal{X}}$ be given alphabets. The encoder's goal is to compress (or publish) is a an n tuple $x^n \in \mathcal{X}^n$. The decoder will take the compressed information provided by the encoder to produce a n tuple $\hat{x}^n \in \hat{\mathcal{X}}^n$. The quality of this reconstruction is measured by a nonnegative distortion function $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}$ applied letter-by-letter to the two sequences in the usual way:

$$d(x_n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

When the tuple x^n is the realization of a sequence of n i.i.d. random variables $X_1, X_2, \ldots, X_n$ with distribution P, the minimum compression rate that that reliably guarantees a reconstruction with distortion less than D is the well-known rate distortion function:

$$R(D) = \min_{Q(\hat{x}|x) : \mathbb{E}_{P \times Q}[d(X,\hat{X})] \leq D} I\left(X; \hat{X}\right). \tag{1}$$

The quantity we are interested in is the channel Q that achieves the minimum mutual information in (1).

B. Differential privacy

One way to interpret the distortion-achieving optimal lossy compression channel Q in (1) is as a mechanism for *private data release* – a holder of (random) private data X^n wishes to release a masked or "sanitized" version $\hat{X}^n$. The distortion represents a loss in utility, and the rate $R(D)$ is the privacy loss or "leakage" associated with the mechanism. A small mutual information (1) means that on average, the output $\hat{X}^n$ does not provide much information about the input X^n.

Differential privacy is another approach to defining privacy which does not make stochastic assumptions on the inputs x^n. Differential privacy is defined for functions of the data x^n; the mapping from x^n to $\hat{x}^n$ is one such function. Because the data is not stochastic, differentially private functions are randomized, and the privacy is a property of the random variable that is the output of the function. These functions are often called mechanisms; a mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ that (randomly) maps data to an arbitrary output space $\mathcal{Y}$ is ε-differentially private if for all $(x^n, \tilde{x}^n) \in \mathcal{X}^n \times \mathcal{X}^n$ such that $d_H(x^n, \tilde{x}^n) \leq 1$,

$$\mathbb{P}\left(M(x^n) \in \mathcal{S}\right) \leq e^{\varepsilon} \mathbb{P}(\, M(\tilde{x}^n) \in \mathcal{S}), \tag{2}$$

for all measurable $\mathcal{S} \subseteq \mathcal{Y}$. One way to interpret (2) is that the distributions of the output of the mechanism with inputs x^n and $\tilde{x}^n$ differing in a single entry are close to each other.

A mechanism that provides ε-differential privacy also guarantees a small leakage as measured by mutual information when the data are random [11]. Let M be a ε-differentially private mechanism and let x^{n-1} be a fixed sequence. Let Y denote the output of M. Suppose X_n and $\tilde{X}_n$ are i.i.d. random variables with distribution P. Then (2) implies that for all realizations $\tilde{x}_n$,

$$\begin{aligned} &\mathbb{P}\left(Y \in \mathcal{S} | (x^{n-1}, x_n)\right) P(\tilde{x}_n) \\ &\qquad \leq e^{\varepsilon} \mathbb{P}\left(Y \in \mathcal{S} | (x^{n-1}, \tilde{x}_n)\right) P(\tilde{x}_n). \end{aligned}$$

Integrating both sides with respect to $\tilde{x}_n$ we see

$$\mathbb{P}\left(Y \in \mathcal{S} | (x^{n-1}, x_n)\right) \leq e^{\varepsilon} \mathbb{P}\left(Y \in \mathcal{S} | x^{n-1}\right).$$

This in turn implies that for all measurable $\mathcal{S}$,

$$\log \frac{\mathbb{P}\left(Y \in \mathcal{S} | (x^{n-1}, x_n)\right)}{\mathbb{P}\left(Y \in \mathcal{S} | x^{n-1}\right)} \leq \varepsilon.$$

That is, the conditional distribution of Y given $X_n = x_n$ is close to the marginal distribution of Y. Taking the expectation over X_n shows that the mutual information $I\left(X_n; Y\right) \leq \varepsilon$.

C. Private channels

For the specific problem of data release, we can consider a restricted class of mechanisms which operate in a single-letter manner by generating a random $\hat{X}_i$ from x_i distributed according to conditional distribution $Q(\hat{x}|x)$. That is, the data holder passes the data x^n through a channel Q to produce $\hat{X}^n$, so $\mathcal{Y} = \hat{\mathcal{X}}^n$. Such a channel provides ε-differential privacy if

$$\log \frac{Q(\hat{x}|x)}{Q(\hat{x}|\tilde{x})} \leq \varepsilon \qquad \forall (x, \tilde{x}, \hat{x}) \in \mathcal{X} \times \mathcal{X} \times \hat{\mathcal{X}}.$$

As above, a channel which provides ε-differential privacy in this way also guarantees a small leakage $I\left(X; \hat{X}\right) \leq \varepsilon$.

For a given rate distortion leakage δ and input distribution P we can define the set of channels which achieve that leakage:

$$\mathcal{Q}_{\mathsf{MI}}(P, \delta) = \left\{Q(\hat{x}|x) : I\left(X; \hat{X}\right) \leq \delta\right\}. \tag{3}$$

Similarly, for a given privacy level ε we can define the set of channels which provide ε-differential privacy:

$$\mathcal{Q}_{\mathsf{DP}}(\varepsilon) = \left\{Q(\hat{x}|x) : \log \frac{Q(\hat{x}|x)}{Q(\hat{x}|\tilde{x})} \leq \varepsilon \ \forall x, \tilde{x} \in \mathcal{X}\right\}. \tag{4}$$

The preceding discussion shows that for any P and any fixed ε, we have $\mathcal{Q}_{\mathsf{DP}}(\varepsilon) \subseteq \mathcal{Q}_{\mathsf{MI}}(P, \varepsilon)$.

D. Classes of sources

We consider rate-distortion problems in which the distribution P governing the data is not known, but it is known that P lies in a set $\mathcal{P}$ of sources on $\mathcal{X}$. We can define a third set of channels containing those channels which guarantee distortion D over the class $\mathcal{P}$:

$$\mathcal{Q}_{\mathsf{RD}}(\mathcal{P}, D) = \left\{Q(\hat{x}|x) : \max_{P \in \mathcal{P}} \mathbb{E}_{P \times Q}[d(X, \hat{X})] \leq D\right\}$$

Because differential privacy is agnostic to the source distribution, $\mathcal{Q}_{\mathsf{DP}}(\varepsilon)$ does not depend on the source distribution P. However, given $\mathcal{P}$ the optimal value of ε subject to a distortion constraint does depend on D:

$$\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D) = \min\left\{\varepsilon : \mathcal{Q}_{\mathsf{RD}}(\mathcal{P}, D) \cap \mathcal{Q}_{\mathsf{DP}}(\varepsilon) \neq \emptyset\right\}. \tag{5}$$

Similarly, for the mutual information we have the optimal value of δ subject to a distortion constraint:

$$\delta^*_{\mathsf{MI}}(\mathcal{P}, D) = \min_{Q \in \mathcal{Q}_{\mathsf{RD}}(\mathcal{P}, D)} \max_{P \in \mathcal{P}} I\left(X; \hat{X}\right). \tag{6}$$

In subsequent sections we will analyze these functions for Hamming distortion.

III. Hamming distortion

A. Binary channels with Hamming distortion

In this section we consider the simplest example of the difference between information-theoretic privacy and differential privacy for binary sources. Let $\mathcal{P} = \{(1-p, p) : p \in [p_{\min}, p_{\max}]\}$ be a set of source distributions on $\{0, 1\}$. We will consider the case where $0 \leq p_{\min} \leq p_{\max} \leq \frac{1}{2}$. We consider memoryless privacy mechanisms Q mapping X to $\hat{X}$.

The set of all possible mechanisms Q can be parameterized by the two crossover probabilities $Q(0|1)$ and $Q(1|0)$. Since in differential privacy the privacy guarantee is just a function of the channel, let

$$\varepsilon(Q) = \max\left\{\left|\log \frac{Q(0|1)}{Q(0|0)}\right|, \left|\log \frac{Q(1|0)}{Q(1|1)}\right|\right\}. \tag{7}$$

The left side of 1 shows the value of $\varepsilon(Q)$ guaranteed by a channel Q as a function of the two crossover probabilities. On the line $Q(0|1) = 1 - Q(1|0)$ the differential privacy risk is 0 because the output of the channel is independent of the input. However, along the edges where $Q(0|1) = 0$ and $Q(1|0) = 0$ the differential privacy risk becomes infinite, since the ratio of the probabilities in (4) becomes infinite.

We measure the utility of a channel Q in terms of the worst-case Hamming distortion over $\mathcal{P}$. The Hamming distortion between two symbols $x, \hat{x} \in \{0, 1\}$ is $d(x, \hat{x}) = \mathbf{1}(x \neq \hat{x})$, and the$d(x^n, \tilde{x}^n)$ between two sequences $x^n, \tilde{x}^n \in \mathcal{X}^n$ is $\sum_{i=1}^n d(x_i, \hat{x}_i)$.

$$\begin{aligned}\Delta &= \max_{p \in \mathcal{P}} \mathbb{E}\left[d(X, \hat{X})\right] \\ &= \max_{p \in \mathcal{P}} \left(pQ(0|1) + (1-p)Q(1|0)\right).\end{aligned}$$

Taking the derivative with respect to p, we see that

$$\frac{d}{dp}\left(pQ(0|1) + (1-p)Q(1|0)\right) = Q(0|1) - Q(1|0).$$

Thus

$$\Delta(Q) = pQ(0|1) + (1-p)Q(1|0),$$

where

$$p = \begin{cases} p_{\min} & \text{if} \quad Q(0|1) < Q(1|0) \\ p_{\max} & \text{if} \quad Q(0|1) > Q(1|0) \end{cases} \tag{8}$$

The maximum distortion is $D_{\max} = p_{\max}$.

The two cases in (8) correspond to above and below the dotted line $Q(0|1) = Q(1|0)$ line shown in Figure 2. The solid line corresponds to the constraint $\Delta(Q) = D$ for some D.

The diagonal line $Q(1|0) = 1 - Q(0|1)$ is the dashed diagonal line in Figure 2. Along this line, we have

$$e^{\varepsilon} = \max\left\{\frac{Q(1|0)}{Q(1|1)}, \frac{Q(1|1)}{Q(1|0)}, \frac{Q(0|0)}{Q(0|1)}, \frac{Q(0|1)}{Q(0|0)}\right\} = 1,$$

or $\varepsilon = 0$. Below the dashed line, in the gray region, we have $Q(1|0) < 1 - Q(0|1)$. This in turn implies that $Q(1|0) < Q(1|1)$ and $Q(0|1) < Q(0|0)$, so we need only calculate two terms for ε:

$$\begin{aligned}e^{\varepsilon} &= \max\left\{\frac{Q(1|1)}{Q(1|0)}, \frac{Q(0|0)}{Q(0|1)}\right\} \\ &= \max\left\{\frac{1 - Q(0|1)}{Q(1|0)}, \frac{1 - Q(1|0)}{Q(0|1)}\right\}.\end{aligned}$$

Now suppose D is fixed and given and we must choose a Q that guarantees $\Delta(Q) \leq D$. What is the smallest privacy risk $\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D)$ over all such Q?

Our first result is a lower bound on $\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D)$ for the class of binary sources

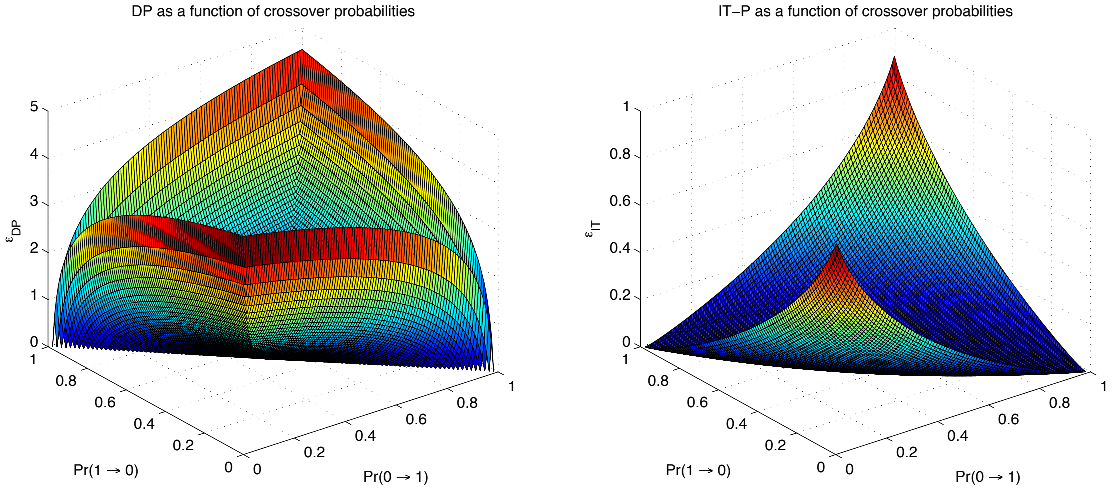

Fig. 1. Surface map of DP and IT privacy guarantees as a function of the channel.

Theorem 1 *Consider the class of binary sources* $\mathcal{P} = \{(1-p, p) : p \in [p_{\min}, p_{\max}]\}$ *for* $0 < p_{\min} \le p_{\max} < 0.5$ *under Hamming distortion. Then for* $D < p_{\max}$,

$$\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D) = \log \frac{1-D}{D}.$$

Moreover,

$$\lim_{D \to p_{\max}} \varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D) = \log \frac{1-p_{\max}}{p_{\max}},$$

but $\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, p_{\max}) = 0$. *That is, the optimal differential privacy risk* $\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D)$ *is discontinuous at* $D = p_{\max}$.

Proof: We claim that ε is minimized along the line $\Delta(Q) = D$. To see this, suppose that we are given a Q such that $\Delta(Q) < D$. If $\frac{1-Q(0|1)}{Q(1|0)} > \frac{1-Q(1|0)}{Q(0|1)}$ we can increase $Q(0|1)$ to lower ε while increasing $\Delta(Q)$ to D. Similarly, if $\frac{1-Q(0|1)}{Q(1|0)} < \frac{1-Q(1|0)}{Q(0|1)}$ we can increase $Q(1|0)$ to lower ε while increasing $\Delta(Q)$ to D. Therefore we need only consider channels Q along the piecewise linear boundary $\Delta(Q) = D$.

We next claim that ε is minimized at $Q(1|0) = Q(0|1)$, which corresponds to the point (D, D) in the plane. At this point, the two terms in ε are equal. To prove the claim, we parameterize the two line segments on the boundary. Let $a = \frac{D}{1-p_{\min}}$. Then the line is parameterized for $t \in [0, 1]$ as

$$\begin{aligned} Q(1|0) &= (1-t)D + ta \\ Q(0|1) &= (1-t)D. \end{aligned}$$

Taking a derivative of one component of ε with respect to t:

$$\begin{aligned} \frac{d}{dt}\frac{1-Q(1|0)}{Q(0|1)} &= \frac{d}{dt}\frac{1-(1-t)D-ta}{(1-t)D} \\ &= \frac{(D-a)(1-t)D + (1-(1-t)D-ta)D}{((1-t)D)^2} \\ &= \frac{-aD+D}{((1-t)D)^2}. \end{aligned}$$

Since $a < 1$ the derivative is positive for $t > 0$, so this term is monotonically increasing in t. Because we have $\frac{1-Q(0|1)}{Q(1|0)} = \frac{1-Q(1|0)}{Q(0|1)}$ at $t = 0$, we must have ε minimized at $Q(1|0) = Q(0|1)$ along the segment (D, D) to $(0, a)$.

Now consider the segment from (D, D) to $(b, 0)$ where $b = D/p_{\max}$. The segment is parameterized as

$$\begin{aligned} Q(1|0) &= (1-t)D \\ Q(0|1) &= (1-t)D + tb. \end{aligned}$$

Taking the derivative of $\frac{1-Q(0|1)}{Q(1|0)}$ with respect to t, we have a similar formula:

$$\begin{aligned} \frac{d}{dt}\frac{1-Q(0|1)}{Q(1|0)} &= \frac{d}{dt}\frac{d}{dt}\frac{1-(1-t)D-tb}{(1-t)D} \\ &= \frac{-bD+D}{((1-t)D)^2}. \end{aligned}$$

And using the same argument, ε is minimized at $Q(1|0) = Q(0|1)$.

Therefore the minimum-ε channel is symmetric, and the corresponding value of ε is

$$\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D) = \log \frac{1-D}{D}.$$

This holds for any $D < D_{\max} = p_{\max}$. In particular,

$$\lim_{D \to D^-_{\max}} \varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D) = \log \frac{1-p_{\max}}{p_{\max}}.$$

However, for $D = p_{\max}$ we can achieve $\varepsilon = 0$ using the channel that sets $Q(0|1) = 1$ and $Q(1|0) = 0$; that is, the channel maps all entries to 0. ■

The theorem says that to guarantee any non-trivial utility, the differential privacy risk is bounded away from 0. In this setting, revealing any approximation of the database incurs a minimum privacy risk of at least $\log \frac{1-p_{\max}}{p_{\max}}$.

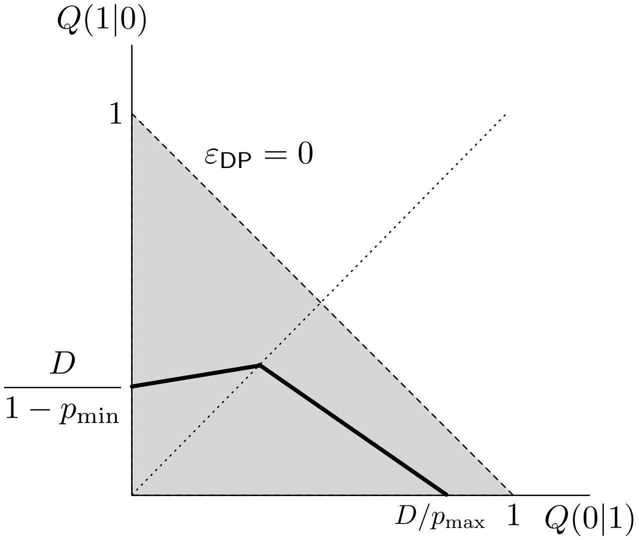

Fig. 2. The $Q(0|1)$-$Q(1|0)$ plane.

B. Discrete alphabets with Hamming distortion

The setup from the binary case generalizes straightforwardly to the case of Hamming distortion for general discrete alphabets. Suppose $\mathcal{X} = [M]$ is a M-ary discrete alphabet. For the rate-distortion problem, we quantify the utility of the published database using the Hamming distortion d. For a given distribution P and a channel Q the expected distortion is

$$\begin{aligned}\mathbb{E}_{P\times Q}[d(X,\hat{X})] &= \Pr\left(X \neq \hat{X}\right)\\ &= \sum_j P\left(j\right) \Pr\left(\hat{X} \neq X | X = i\right)\\ &= \sum_j P\left(j\right) \sum_{i,i\neq j} Q\left(i|j\right)\\ &= \sum_j P\left(j\right)\left(1 - Q\left(j|j\right)\right). \qquad (9)\end{aligned}$$

The worst case Hamming distortion for an M-ary alphabet is defined as

$$\Delta\left(\mathcal{P}, Q\right) = \max_{P\in\mathcal{P}} \mathbb{E}_{P\times Q}[d(X,\hat{X})]. \qquad (10)$$

We have the following conjecture for Hamming distortion for general sources.

Conjecture 1 *Consider a closed convex set of M-ary distributions $\mathcal{P}$ under Hamming distortion. Then the channel achieving the minimum*

$$Q(\hat{x}|x) = \begin{cases} 1 - D & \hat{x} = x \\ \frac{D}{M-1} & \hat{x} \neq x \end{cases}$$

For this channel,

$$\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D) = \log(M-1) + \log\frac{1-D}{D}.$$

Moreover,

$$\lim_{D\to D_{\max}} \varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D) = \log(M-1) + \log\frac{1-D_{\max}}{D_{\max}},$$

but $\varepsilon^*_{\mathsf{DP}}(\mathcal{P}, D_{\max}) = 0$.

The conjecture states that the optimal channel is symmetric and does not depend on the class of sources $\mathcal{P}$. From the distortion constraint, we can see that the set of channels which guarantees distortion D universally over the class $\mathcal{P}$ is a piecewise linear region.

IV. Conclusion

In this paper we initiated the study of differentially private mechanisms from a rate-distortion perspective. The literature on differential privacy is extensive; we refer to a recent tutorial by Dwork and Roth [12] for a more comprehensive treatment. Rate distortion is applicable when the goal of the data collector is to publish an approximation of the data itself. This is different than problems that have been considered previously in the literature, such as releasing the answers to a set of statistical queries [13]–[15], interactive queries [16], releasing synthetic data [17], data marginals [18], or statistical estimation [19]–[21]. Furthermore, we consider the case where the data collector is not trusted, which leads us to use local differential privacy [21] as our measure of privacy.

To capture uncertainty in the source distribution, we consider a robust rate-distortion setting in which the source distribution is unknown but comes from a class $\mathcal{P}$, and ask for a locally differentially private channel $Q(\hat{x}|x)$ that achieves minimum privacy risk while simultaneously guaranteeing distortion no more than D universally over the class $\mathcal{P}$. For the most basic example of binary sources under Hamming distortion, we characterize this channel and show a surprising result: the release of any approximation to the database with non-maximum distortion incurs a minimum privacy risk.

In future work we plan to extend this framework to other source models and distortion measures. For general discrete alphabets with Hamming distortion we conjecture the result for the binary case will extend naturally to prove a similar lower bound on the privacy risk. Given recent interest in mutual information measures of utility via the logarithmic loss function [22], [23], investigating lower bounds for that case would definitely be interesting and may shed some additional insight on the minimax lower bounds for statistical inference [21] that depend on the class $\mathcal{P}$. Extending the framework to continuous sources may be challenging due to to range issues [20]. Another interesting direction is to relax the privacy measure to (ϵ, δ)-differential privacy – in this setting the discontinuity may vanish due to the smoothing effect of δ, establishing another separation between the two privacy measures.

References

[1] L. Sweeney, "k-anonymity: a model for protecting privacy," *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, October 2002. [Online]. Available: http://dx.doi.org/10.1142/S0218488502001648

[2] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proc. 2008 IEEE Symp. on Security and Privacy*, 2008, pp. 111–125.

[3] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoff in databases: An information-theoretic approach," *IEEE Transactions on Information Forensics and Security*, 2013 (to appear). [Online]. Available: http://dx.doi.org/10.1109/TIFS.2013.2253320

[4] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, ser. Lecture Notes in Computer Science, S. Halevi and T. Rabin, Eds., vol. 3876. Berlin, Heidelberg: Springer, March 4–7 2006, pp. 265–284. [Online]. Available: http://dx.doi.org/10.1007/11681878_14

[5] S. Zhou, K. Ligett, and L. Wasserman, "Differential privacy with compression," in *Proceedings of the International Symposium on Information Theory (ISIT)*, Seoul, Korea, June–July 2009.

[6] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy," ArXiV, Tech. Rep. arXiv:1212.1186, October 2012. [Online]. Available: http://arxiv.org/abs/1212.1186

[7] ——, "The optimal mechanism in (ϵ, δ)-differential privacy," ArXiV, Tech. Rep. arXiv:1305.1330, December 2013.

[8] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," ArXiV, Tech. Rep. arXiv:1407.1338 [cs.IT], July 2014.

[9] 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. [Online]. Available: http://dx.doi.org/10.1198/jasa.2009.tm08651

[10] S. Oh and P. Viswanath, "The composition theorem for differential privacy," ArXiV, Tech. Rep. arXiv:1311.0776 [cs.DS], December 2013. [Online]. Available: http://arxiv.org/abs/1311.0776

[11] A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. Vadhan, "The limits of two-party differential privacy," in *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010)*, 2010, pp. 81–90. [Online]. Available: http://dx.doi.org/10.1109/FOCS.2010.14

[12] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014. [Online]. Available: http://dx.doi.org/10.1561/0400000042

[13] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology - EUROCRYPT 2006*, ser. Lecture Notes in Computer Science, S. Vaudenay, Ed., vol. 4004. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 486–503. [Online]. Available: http://dx.doi.org/10.1007/11761679_29

[14] M. Hay, V. Rastogi, G. Miklau, and D. Suciu, "Boosting the accuracy of differentially private histograms through consistency," *Proceedings of the VLDB Endowment*, vol. 3, no. 1, pp. 1021–1032, 2010.

[15] M. Gaboardi, E. J. G. Arias, J. Hsu, A. Roth, and Z. S. Wu, "Dual query: Practical private query release for high dimensional data," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2014.

[16] A. Roth and T. Roughgarden, "Interactive privacy via the median mechanism," in *Proceedings of the 42nd ACM Symposium on Theory of Computing*, 2010.

[17] M. Hardt, K. Ligett, and F. McSherry, "A simple and practical algorithm for differentially private data release," in *Advances in Neural Information Processing Systems 25*, P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., 2012, pp. 2348–2356. [Online]. Available: http://books.nips.cc/papers/files/nips25/NIPS2012_1143.pdf

[18] K. Chandrasekaran, J. Thaler, J. Ullman, and A. Wan, "Faster private release of marginals on small databases," in *Proceedings of Innovations in Theoretical Computer Science*, 2014.

[19] A. Smith, "Privacy-preserving statistical estimation with optimal convergence rates," in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC '11)*. New York, NY, USA: ACM, 2011, pp. 813–822. [Online]. Available: http://dx.doi.org/10.1145/1993636.1993743

[20] K. Chaudhuri and D. Hsu, "Convergence rates for differentially private statistical estimation," in *ICML*, 2012.

[21] J. Duchi, M. J. Wainwright, and M. Jordan, "Local privacy and minimax bounds: Sharp rates for probability estimation," in *Advances in Neural Information Processing Systems 26*, 2013, pp. 1529–1537. [Online]. Available: http://media.nips.cc/nipsbooks/nipspapers/paper_files/nips26/764.pdf

[22] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *Proceedings of the 50th Annual Allerton Conference on Communications, Control and Computing*, 2012.

[23] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *Proceedings of the IEEE Information Theory Workshop*, Hobart, Australia, November 2014.