

# The Laplace Mechanism has optimal utility for differential privacy over continuous queries

Natasha Fernandes<sup>\*†§</sup>, Annabelle McIver<sup>\*</sup>, Carroll Morgan<sup>†‡</sup>

<sup>\*</sup>Department of Computing, Macquarie University, Sydney

<sup>†</sup>School of Computer Science and Engineering, UNSW, Sydney

<sup>‡</sup>Data61, CSIRO, Sydney

<sup>§</sup>Inria and École Polytechnique, IPP, France

**Abstract**—Differential Privacy protects individuals’ data when statistical queries are published from aggregated databases: applying “obfuscating” mechanisms to the query results makes the released information less specific but, unavoidably, also decreases its utility. Yet it has been shown that for *discrete* data (e.g. counting queries), a mandated degree of privacy and a reasonable interpretation of loss of utility, the Geometric obfuscating mechanism is optimal: it loses as little utility as possible [Ghosh et al.[1]].

For *continuous* query results however (e.g. real numbers) the optimality result does not hold. Our contribution here is to show that optimality is regained by using the *Laplace* mechanism for the obfuscation.

The technical apparatus involved includes the earlier discrete result [Ghosh op. cit.], recent work on abstract channels and their geometric representation as hyper-distributions [Alvim et al.[2]], and the dual interpretations of distance between distributions provided by the Kantorovich-Rubinstein Theorem.

**Index Terms**— Differential privacy, utility, Laplace mechanism, optimal mechanisms, quantitative information flow, abstract channels, hyper-distributions.

## I. INTRODUCTION

### A. The existing optimality result, and our extension

Differential Privacy (DP) concerns databases from which (database-) queries produce statistics: a database of information about people can be queried e.g. to reveal their average height, or how many of them are men. But a risk is that from a *general* statistic, specific information might be revealed about individuals’ data: whether a specific person is a man, or his height, or even both. Differentially-private “obfuscating” mechanisms diminish that risk by perturbing their inputs (the raw query results) to produce outputs (the query reported) that are slightly wrong in a probabilistically unpredictable way. That diminishes the personal privacy risk (good) but also diminishes the statistics’ utility (bad).

The existing optimality result is that for a mandated differential privacy parameter, some  $\epsilon > 0$ , and under conditions we will explain, the *Geometric* obfuscating mechanism  $G^\epsilon$  (depending on  $\epsilon$ ) loses the least utility of *any*  $\epsilon$ -Differentially

Private oblivious obfuscating mechanism for the same  $\epsilon$ , that loss being caused by her having to use the perturbed statistic instead of the real one [1].

A conspicuous feature of  $\epsilon$ -DP (that is  $\epsilon$ -differential privacy) is that it is achieved *without* having to know the nature of the individual’s privacy that it is protecting: it is simply made “ $\epsilon$ -difficult” to determine whether *any* of his data is in the database at all. Similarly the minimisation of an observer’s loss (of utility) is achieved by the optimal obfuscation *without* knowing precisely how the obfuscation affects her: instead, the existence of a “loss function” is postulated that monetises her loss (think “dollars”) based on the raw query (which she does not know) and the obfuscated query (which she does know) — and optimality of  $G^\epsilon$  holds wrt. *all* loss functions (within certain realistic constraints) and *all* (other)  $\epsilon$ -DP mechanisms  $M^\epsilon$ .

IN SUMMARY: The existing result states that the  $\epsilon$ -DP Geometric obfuscating mechanism  $G^\epsilon$  minimises loss of utility to an observer when the query results are discrete, e.g. counting queries in some  $(0..N)$ , and certain reasonable constraints apply to the monetisation of loss. But the result does not hold when the query results are continuous, e.g. in the unit interval  $[0, 1]$ . **We show that optimality is regained by using the  $\epsilon$ -DP Laplace mechanism  $L^\epsilon$ .**

## II. DIFFERENTIAL PRIVACY, LOSS OF UTILITY AND OPTIMALITY

### A. Differential privacy

Differential privacy begins with a database that is a multiset of *rows* drawn from some set  $R$  [3]; thus the type of a database is  $\mathbb{B}R$  (using “ $\mathbb{B}$ ” for “bag”). A query  $q$  is a function from database to a query-result in some set  $\mathcal{X}$ , the input of the mechanism, and is thus of type  $\mathbb{B}R \rightarrow \mathcal{X}$ .

A distance function between databases  $\mathbb{D}: \mathbb{B}R \times \mathbb{B}R \rightarrow \mathbb{R}$  measures how different two databases are from each other. Often used is the *Hamming* distance  $\mathbb{D}_H$ ,<sup>1</sup> which gives (as an integer) how many whole rows would have to be removed or inserted to transform one database into another: given two databases  $b_1, b_2: \mathbb{B}R$  we define  $\mathbb{D}_H(b_1, b_2)$  to be the size  $\#(b_1 \Delta b_2)$  of their (multiset-) symmetric difference. Thus in

<sup>1</sup>The Hamming distance is also known as the *symmetric* distance.

particular two databases that differ only because a row has been removed from one of them have Hamming distance 1, and we say that such databases are *adjacent*.

We define also a distance function (metric) between –for the moment– discrete distributions  $\mathbb{D}\mathcal{Y}$  over a set of observations  $\mathcal{Y}$  the output of the mechanism. Given two distributions  $\delta_1, \delta_2$  on  $\mathcal{Y}$ , their distance  $d_D(\delta_1, \delta_2)$  (for “Dwork”) is based on the largest ratio over all  $Y \subseteq \mathcal{Y}$  between probabilities assigned to  $Y$  by  $\delta_1$  and  $\delta_2$  — it is

$$d_D(\delta_1, \delta_2) := \max_{Y \subseteq \mathcal{Y}} |\ln(\delta_1(Y)/\delta_2(Y))| \quad (1)$$

where  $\delta(Y)$  is the probability  $\delta$  assigns to the whole subset  $Y$  of  $\mathcal{Y}$ , and the logarithm is introduced to make the distance satisfy the triangle inequality that metrics require.<sup>2</sup>

Following the presentation of Chatzikokolakis et al. [4], once we have chosen a metric  $D$  on databases, we say that a mechanism  $M$  achieves  $\varepsilon$ -Differential Privacy wrt. that  $D$  and some query  $q$ , i.e. is  $\varepsilon$ -DP for  $D, q$ , just when

$$\text{for all databases } b_1, b_2 \text{ in } \mathbb{B}R \text{ we have} \\ d_D(M(q(b_1)), M(q(b_2))) \leq \varepsilon \cdot D(b_1, b_2) \quad (2)$$

In the special case when  $D$  is the Hamming distance  $D_H$ , the above definition becomes

for all databases  $b_1, b_2$  in  $\mathbb{B}R$  with  $D_H(b_1, b_2) \leq 1$ ,  
i.e. that differ only in the presence/absence of a single row,  
and for all subsets  $Y$  of  $\mathcal{Y}$  we have

$$\Pr(M(q(b_1)) \in Y) \leq e^\varepsilon \cdot \Pr(M(q(b_2)) \in Y) \quad (3)$$

With the above metric-based point of view we can say that an  $\varepsilon$ -DP mechanism is (simply) a  $\varepsilon$ -Lipschitz function from databases  $\mathbb{B}R$  with metric  $D$  to distributions of observations  $\mathbb{D}\mathcal{Y}$  with metric  $d_D$  [4].

*Definition 1:* ( $d_D/D$   $\varepsilon$ -DP for mechanisms) A Lipschitz mechanism  $M$  from  $\mathcal{X}$  raw query outputs to  $\mathcal{Y}$  (observations) is ( $d_D/D$ )  $\varepsilon$ -Differentially Private just when

$$\text{for all inputs } x_1, x_2 \text{ in } \mathcal{X} \text{ to } M \text{ we have} \\ d_D(M(x_1), M(x_2)) \leq \varepsilon \cdot D(x_1, x_2) \quad (4)$$

in which we elide  $d_D$  and  $D$  when they are clear from context. In (2) we gave the special case where  $M$ 's inputs  $x_1, x_2$  were raw query-results  $q(b_1), q(b_2)$ , i.e. with  $b_1, b_2$  two databases acted on by the same query-function  $q$ . And (3) was further specialised to where the two databases were adjacent and the metric was  $D_H$ , the Hamming distance.

## B. “Counting” queries

Counting queries on databases are the special case where the codomain  $\mathcal{X}$  of the query (the mechanism input) is the non-negative integers and the query  $q$  returns the number of database rows satisfying some criterion, like “being a man”. The “average height” query is not a counting query.

When the database metric is the Hamming distance  $D_H$ , a counting query can be characterised more generally as one that is a 1-Lipschitz function wrt.  $d_H$  and the usual metric (absolute

difference) on the integers, i.e. one whose result changes by at most 1 between adjacent databases. Since composition of Lipschitz functions (merely) multiplies their Lipschitz factors, the composition of a counting query and an obfuscating mechanism is  $\varepsilon$ -DP as a whole if the mechanism on its own (i.e. without the query, acting “obliviously” on  $x=q(b)$ ) is  $\varepsilon$ -Lipschitz. That is why for counting queries we can concentrate on the mechanisms alone (whose type is  $\mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ ) rather than including the databases and their type  $\mathbb{B}R$  in our analysis.

## C. Prior knowledge, open-source and the observer

Although the database contents are not (generally) known, often the distribution of its query results *is* known: this is “prior knowledge”, where e.g. it is known that a database of heights in the Netherlands is likely to contain higher values than a similar database in other countries — and that knowledge is different from the (unknown) fact we are trying to protect, i.e. whether a *particular* person’s height is in that database.

We abstract from prior knowledge of the database by concentrating instead on the prior knowledge  $\pi$  of the distribution  $\mathcal{X}$  of raw queries, the inputs  $x$  to the mechanism, that is *induced* as the push-forward of the query-function (an “open source” aggregating function) over the known distribution of possible databases themselves. Knowing  $\pi$  on the input in  $\mathcal{X}$  the observer can use her knowledge of the mechanism (also open source) to deduce a distribution on the output observations in  $\mathcal{Y}$  that will result from applying it and –further– she can also deduce a posterior distribution on  $\mathcal{X}$  based on any particular  $y$  in  $\mathcal{Y}$  that she observes.

## D. The Geometric mechanism is $\varepsilon$ -DP for $D_H$

1) *Specialising to  $D_H$ :* Recall from §II-B that  $D_H$ , the Hamming distance, is what is typically used for counting queries. In that case we see as follows from (2) that the Geometric mechanism  $G$  can be made  $\varepsilon$ -DP.

The Geometric distribution centred on 0 with parameter  $\alpha$  assigns (discrete) probability

$$G_\alpha(n) := 1^{-\alpha/1+\alpha} \cdot \alpha^{|n|} \quad (5)$$

to any integer  $n$  (positive or negative) [1]. It implements an  $\varepsilon$ -DP Geometric mechanism by obfuscating the query according to (5) above: thus set  $\alpha := e^{-\varepsilon}$  and define

$$G^\varepsilon(n)(n') := G_\alpha(n'-n) = 1^{-\alpha/1+\alpha} \cdot \alpha^{|n'-n|} \quad (6)$$

to be the probability that integer  $n$  is input and  $n'$  is output. Thus applied to some  $n$ , the effect of  $G^\varepsilon$  with  $\varepsilon := -\ln \alpha$ <sup>3</sup> is to leave  $n$  as it is with probability  $1^{-\alpha/1+\alpha}$  and to split the remaining probability  $2\alpha/1-\alpha$  equally between adding 1’s or subtracting them:  $G^\varepsilon$  continues (in the same direction) with repeated probability  $\alpha$  until, with probability  $1-\alpha$ , it stops.

As explained in §II-C, we now concentrate on  $G^\varepsilon$  alone and how it perturbs its input (a query result), i.e. no longer considering the database from which the query came.<sup>4</sup>

<sup>3</sup>The  $\alpha$  in  $G_\alpha$  is  $<1$ , so  $\varepsilon > 0$ .

<sup>4</sup>Note that although the (raw) query is *output* from the database, it is *input* to the obfuscating mechanism. That is why we refer to  $\mathcal{X}$  as “input”.

<sup>2</sup>This distance is also known as “max divergence”.

2) *The geometric mechanism truncated:* In (6) the mechanism  $G^\varepsilon$  can effect arbitrary large perturbations. But in practice its output is constrained (in the discrete case) to a finite set  $(0..N)$  by (re-)assigning all probabilities for negative observations to observation 0, and all probabilities for  $\geq N$  observations to  $N$ . For example with  $e^{-\varepsilon}=\alpha=1/2$  and restricting to (0..2) we have  $G^\varepsilon(0)(0) = \dots + 1/12 + 1/6 + 1/3 = 2/3$  and  $G^\varepsilon(0)(1) = 1/6$  and  $G^\varepsilon(0)(2) = 1/12 + 1/24 + \dots = 1/6$ . It can be shown [5] however that truncation makes no difference to our results, and so from here on we will assume that truncation has been applied to  $G^\varepsilon$ .

### E. Discrete optimality

It has been shown [1] that when  $\mathcal{X}$  is discrete (and hence the prior  $\pi$  on  $\mathcal{X}$  is also), and when the obfuscation is via  $G^\varepsilon$ , and when the observer applies a “loss function”  $\ell(w, x)$  of her choice to monetise in  $\mathbb{R}^\geq$  the loss of utility to her if the raw query was  $x$  but she assumes it was  $w$ , then *any other*  $\varepsilon$ -DP mechanism  $M^\varepsilon$  acting on  $\mathcal{X}$  can only lose *more* utility (on average) according to that  $\pi$  and  $\ell$  than  $G^\varepsilon$  does. That is, the  $\varepsilon$ -DP Geometric mechanism is *optimal* for minimising loss (maximising utility) over all priors  $\pi$  and all (“legal”) loss functions  $\ell$  under a mandated  $\varepsilon$ -DP obfuscation. A loss function is said to be *legal* if it is monotone (increasing) wrt. to the difference between the guess ( $w$ ) and the actual value  $x$  of the query. As explained in [1] this means that the loss  $\ell(w, x)$  takes the form of a function  $m(|w-x|, x)$ , which must be monotone (increasing) in its first argument.

### F. The geometric mechanism is never $\varepsilon$ -DP on dense continuous inputs, e.g. when $\mathbb{D}$ on $\mathcal{X}$ is Euclidean

If the input metric for  $G$  is not the Hamming distance  $\mathbb{D}_H$ , e.g. when the  $G$ 's input  $\mathcal{X}$  is continuous, still  $G$ 's output remains discrete, taking some number of steps, each of fixed length say  $\lambda > 0$ , in either direction. That is, any  $G$  input  $x$  is perturbed to  $x+i\lambda$  for some integer  $i$ .

Now because  $\mathcal{X}$  is continuous and dense, we can vary the input  $x$  itself, by a tiny amount, to some  $x'$  so that  $\mathbb{D}(x, x') < \lambda$  no matter how small  $\lambda$  might be, producing perturbations  $x'+i\lambda$  each of which is distant that same (constant)  $\mathbb{D}(x, x')$  from the original  $x+i\lambda$  and, precisely because  $\mathbb{D}(x, x') < \lambda$ , those new perturbations cannot overlap the ones based on the original  $x$ .

Thus the two distributions produced by  $G$  acting on  $x$  and on  $x'$  have supports that do not intersect at all. And therefore the  $\mathbb{D}_D$  distance between the two distributions is infinite, meaning that  $G$  cannot be  $\varepsilon$ -DP for any (finite)  $\varepsilon$ . That is, for a database producing truly real query results  $\mathcal{X}$ , a standard (discrete)  $G$  cannot establish  $\varepsilon$ -DP for any  $\varepsilon$ , however large  $\varepsilon$  might be.

There are two possible solutions. The first solution, both obvious and practical, is to “discretise” the input and to scale appropriately: a person’s height of 1.75m would become 175cm instead. A second solution however is motivated by taking a more theoretical approach. Rather than discretise the type of the query results, we *leave* it continuous — and

seek our optimal mechanism among those that —unlike the Geometric— do not take only discrete steps. It will turn out to be the Laplace distribution.

### G. Our result — continuous optimality

In the discrete case typically the set  $\mathcal{X}$  of raw queries is  $(0..N)$  for some  $N \geq 0$ , and the prior knowledge  $\pi$  is a (discrete) distribution on that. For our continuous setting we will use  $\mathcal{X}=[0, 1]$  for raw queries, the unit interval  $\mathcal{U}$ , and the discrete distribution  $\pi$  will become a proper measure on  $[0, 1]$  expressed as a probability density function. The  $\varepsilon$ -DP obfuscating mechanisms, now  $K^\varepsilon$  for “kontinuuous”, will take a raw query  $x$  from a continuous set  $\mathcal{X}$  rather than a discrete one. And the metric on  $\mathcal{X}=\mathcal{U}$  will be Euclidean.

**Our (continuous) optimality result** formalised at Thm. 5 is that  $\varepsilon$ -DP Laplace mechanism  $L^\varepsilon$  minimises loss over all continuous priors  $\pi$  on  $\mathcal{X}=\mathcal{U}$  and all legal loss functions  $\ell$  under a mandated  $\varepsilon$ -DP obfuscation with respect to the Euclidean metric on the continuous input  $\mathcal{X}=[0, 1]$ .

The theorem requires that *all* mechanisms satisfy (2) with  $\mathbb{D}$  the Euclidean distance on continuous inputs. We write  $\varepsilon$ -DP for such mechanisms. The argument in §II-F above shows therefore that Geometric mechanisms are no longer suitable (for optimality) because on continuous  $\mathcal{X}$  they are no longer  $\varepsilon$ -DP.

## III. AN OUTLINE OF THE PROOF

We access the existing discrete results in §I-A, and §II-E from within the continuous  $\mathcal{U}$  by “pixelating” it, that is defining  $\mathcal{U}_N = \{0, 1/N, 2/N, \dots, N-1/N, 1\}$  for integer  $N > 0$ , and mapping  $(0..N)$  isomorphically onto that discrete subset. We then establish near optimality for a similarly pixelated Laplace mechanism, showing that “near” becomes “equal to” when  $N$  tends to infinity. In more detail:

- (a) (We show in §VI-B that) Any (discrete) prior on  $\mathcal{U}_N$  corresponds to some prior on the original  $\mathcal{U}$ , but can also be obtained by pixelating some *continuous* prior  $\pi$  on all of  $\mathcal{U}$ , concentrating its (now discrete) probabilities onto elements of  $\mathcal{U}_N$  only: e.g. the probability  $\pi[n/N, n+1/N)$  of the entire  $1/N$ -sized interval is moved onto the point  $n/N$ . We write it  $\pi_N$ .
- (b) (§VI-C) Any function  $f$  acting on all of  $\mathcal{U}$  can be made into an  $N$ -step function by first restricting its inputs to  $\mathcal{U}_N$  and then filling in the “missing” values  $f(x)$  for  $x$  in  $(n/N, n+1/N)$  by copying the value for  $f(n/N)$ . If  $f$  is an  $\varepsilon$ -DP mechanism  $K^\varepsilon$ , we write its  $N$ -stepped version as  $K_N^\varepsilon$ , and note that  $K_N^\varepsilon$  remains  $\varepsilon$ -DP when restricted to the points in  $\mathcal{U}_N$  only.  
If  $f$  is a loss function on  $(\mathcal{W}$  and)  $\mathcal{X}$  we write  $\ell_N$  for its stepped version.
- (c) (§VII-C; Lem. 10; Lem. 16) Now for any  $N$ , mechanism  $K_N^\varepsilon$ , prior  $\pi_N$ , and legal  $N$ -step loss function  $\ell_N$  we can appeal to the discrete optimality result: for the pixelated prior  $\pi_N$  and the  $N$ -step and legal  $\ell_N$  the loss due to  $G_N^\varepsilon$  is  $\leq$  the loss due to  $K_N^\varepsilon$ .

- (d) (§VII-B; Thm. 13) The replacement of  $G_N^\varepsilon$  by  $L_N^\varepsilon$  (both  $N$ -step functions on  $[0, 1]$ ) is via pixelating the *output* (continuous) distribution of  $L_N^\varepsilon$  to a multiple  $T$  of  $N$ : we write that  ${}^T L_N^\varepsilon$ . The Kantorovich-Rubinstein Theorem, provided additionally that  $\ell_N$  is  $p$ -Lipschitz for some  $p > 0$  independent of  $N$ , shows that the (additive) difference between the  $G_N^\varepsilon$ -loss and the  ${}^T L_N^\varepsilon$ -loss, for any  $\pi_N$  and  $\ell_N$  and  $T$  a multiple of  $N$ , tends to zero as  $N$  increases.
- (e) (§VI-D) Then we remove the subscript  $\pi_N$  on the prior, and on the mechanisms  $K_N^\varepsilon$  and  ${}^T L_N^\varepsilon$ , relying now on the  $\varepsilon$ -DP of the two mechanisms to make the (multiplicative) ratio between the losses they cause tend to 1.
- (f) (§VIII) The final step, removing the subscript  $N$  from  $\ell_N$ , is that the loss-calculating procedure is continuous and that  $\ell_N$  tends to  $\ell$  as  $N$  tends to infinity.

#### IV. CHANNELS; LOSS FUNCTIONS; HYPER-DISTRIBUTIONS; REFINEMENT

In this section we provide a summary of the more general Quantitative Information Flow techniques that we will need for the subsequent development.

##### A. Channels, priors, marginals, posteriors

The standard treatment of information flow is via Shannon’s (unreliable) channels: they take an input  $x$  from say  $\mathcal{X}$  and deliver an output that for a perfect channel will be  $x$  again, but for an imperfect channel might be some other  $x'$  in  $\mathcal{X}$  instead. For example, an imperfect channel transmitting bits might “flip” input bits so that with probability say  $1/4$  an input 0 becomes an output 1 and vice versa [6]. In the discrete case, and generalising to allow outputs of possibly a different type  $\mathcal{Y}$ , such channels are  $\mathcal{X} \times \mathcal{Y}$  matrices  $C$  whose row- $x$ , column- $y$  element  $C_{x,y}$  is the probability that input  $x$  will produce output  $y$ . A perfect channel would be the identity matrix on  $\mathcal{X} \times \mathcal{X}$ ; a completely broken channel on  $\mathcal{X} \times \mathcal{Y}$  for any  $\mathcal{Y}$  would have  $C_{x,y} = 1/\#\mathcal{Y}$ .

The  $x$ -th row of a (channel) matrix  $C$  is  $C_{x,-}$ ; and the  $y$ -th column is  $C_{-,y}$ . Since each row sums to 1 (making  $C$  a *stochastic* matrix), the row  $C_{x,-}$  determines a discrete distribution in  $\mathbb{D}\mathcal{Y}$ ; for the “broken” channel it would be the uniform distribution, which we write  $\odot$ .

As a matrix, a channel has type  $\mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  (but with 1-summing rows); isomorphically it also has type  $\mathcal{X} \rightarrow \mathbb{D}\mathcal{Y}$ . We’ll write  $\mathcal{X} \rightarrow \mathcal{Y}$  for both, provided context makes it clear which one we are using.

If a *prior* distribution  $\pi: \mathbb{D}\mathcal{X}$  on  $\mathcal{X}$  is known, then the channel  $C$  can be applied to  $\pi$  to create a joint distribution  $J$  in  $\mathbb{D}(\mathcal{X} \times \mathcal{Y})$  on both input and output together, written  $\pi \triangleright C$  and where  $J_{x,y} := \pi_x C_{x,y}$ . For that  $J$ , the left-marginal  $\sum_y J_{x,y}$  gives the prior  $\pi$  again (no matter what  $C$  might be), i.e. the probability that the input was  $x$  — thus  $\pi_x = J_{x,\Sigma}$  if we use that notation for the marginal. The right marginal  $J_{\Sigma,y}$  is the probability that the output is  $y$ , given both  $\pi$  and  $C$ .

The *y*-posterior distribution on  $\mathcal{X}$ , given  $\pi, C$  and a particular  $y$ , is the *conditional* distribution on  $\mathcal{X}$  if that  $y$  was output:

it is the  $y$ -th column divided by the marginal probability of that  $y$ , that is  $J_{-,y}/J_{\Sigma,y}$  (provided the marginal is not zero).

If we fix  $\pi$  and  $C$ , and use the conventional abbreviation  $p_{XY}$  for the resulting joint distribution ( $\pi \triangleright C$ ), then the usual notations for the above are  $p_X$  for left marginal ( $= \pi$ ) and  $p_X(x)$  for its value  $\pi_x$  at a particular  $x$ , with  $p_Y$  and  $p_Y(y)$  similarly for the right marginal. Then  $p_{X|y}(x)$  is the posterior probability of the original observation’s being  $x$  when  $y$  has been observed. Further, we can write just  $p(x)$  and  $p(y)$  and  $p(x|y)$  when context makes the (missing) subscripts clear.

##### B. Loss functions; remapping

Our obfuscating mechanisms  $M$  and  $G^\varepsilon$  are channels like  $C$  in the discrete case — the result of the query is the channel’s input  $x$ , and the (perturbed) value the observer sees is the channel’s output  $y$ . The loss functions  $\ell(w, x)$  will quantify the loss to her of seeing (only)  $y$ , and then choosing  $w$ , when what she really wants to know is  $x$ . Such  $\varepsilon$ -DP mechanisms have earlier been modelled this way, i.e. as channels by Alvim et al. [7] and Chatzikokolakis et al. [4], who observed that for  $\varepsilon$ -DP the ratios of their entries must satisfy the  $\varepsilon$ -DP constraints, because the definition at §II-A(4) reduces to comparing (multiplicatively) adjacent entries in channel columns.

The connection between the observation  $y$  and the loss-function parameter  $w$  is that the observer does not necessarily have to “take what she sees” — there might be good reasons for her making a different choice. For example, in a word-guessing game where the last, obfuscated letter ? in a word SA? is shown on the board, the observer might have to guess what it really is. Even if it looks like a blurry Y (value 4 in Scrabble), she might instead guess X (value 8) because that would earn more points on average if from prior knowledge she knows that X strictly *more* than half as likely as Y is — i.e. it’s worth her taking the risk. Thus rather than mandating that the observer must accept what she thinks the letter is most likely to be, she uses the obfuscated query  $y$  to deduce information about the *whole* posterior distribution of the *actual* query... and might suggest that she guess some  $w \neq y$ , because the expected loss of doing that is less than (the expected utility is greater than) it would be if she simply accepted the  $y$  she saw. That rational strategy is called “remapping” [1]. Thus she sees  $y$ , but  $y$  tells her that  $w$  is what she should choose as her least-loss inducing guess for  $x$ . That is, the *simplest* strategy is “take what you see”; but it might not be the best one. In general (and now using  $M$  again for mechanism), we write  $\$(\pi, M, \ell)$  for the *expected* loss to a rational observer, given the  $\pi, M$  she knows and the loss function  $\ell$  she has chosen: it is

$$\sum_y p(y) \min_w \sum_x \ell(w, x) p(x|y) \quad , \quad (7)$$

that is the expected value, over all possible observations  $y$  and their marginal probabilities, of the *least* loss she could rationally achieve over *all* her possible choices  $w$  given the knowledge that  $y$  will have provided about the posterior distribution  $p(X|y)$  of the actual raw input  $x$ . Note that  $M$  and

$\pi$  determine (from §IV-A) the  $p(y)$  and  $p(x|y)$  that appear in (7). We remark that this formulation for measuring expected loss corresponds precisely to the formulation used by Ghosh et al. in the optimality theorem.

### C. The relevance of hyper-distributions, abstract channels

It is important to remember that the expected-loss formula (7) does not use the *actual* mechanism-output values  $y$  in any way directly: instead it takes the only expected value of what they *might be*. All that matters is their marginal probabilities  $p(y)$  and the a-posteriori distributions  $p(X|y)$  that they induce. That allows us to abstract from  $\mathcal{Y}$  altogether.

A *hyper-distribution* expresses that abstraction: it is a distribution of distributions on  $\mathcal{X}$  alone, that is of type  $\mathbb{D}\mathbb{D}\mathcal{X}$ ; abbreviate those as “hyper” and “ $\mathbb{D}^2\mathcal{X}$ ”. Given a joint distribution  $J: \mathbb{D}(\mathcal{X} \times \mathcal{Y})$ , we write  $[J]$  for the hyper-distribution whose support is posterior distributions<sup>5</sup>  $p(X|y)$  on  $\mathcal{X}$  and which assigns the corresponding marginal distribution  $p(y)$  to each. (Zero-valued marginals are left out.) We now re-express (7) in those terms.

If we write  $\ell(w, -)$  for the function on  $\mathcal{X}$  that  $\ell$  determines once  $w$  is fixed, and write  $\mathcal{E}_{\text{DIST RV}}$  for expected value of random-variable RV with distribution DIST, then  $\min_w \mathcal{E}_{p(X|y)} \ell(w, -)$  is the inner part of (7). Then fix some  $\ell$  and define for general distribution  $\delta: \mathbb{D}\mathcal{X}$  that

$$Y_\ell(\delta) := \min_w \mathcal{E}_\delta \ell(w, -) \quad , \quad (8)$$

(using  $Y$  for “entrop $Y$ ”) so that  $Y_\ell$  is itself a real-valued function on distributions  $\delta$  (as e.g. Shannon entropy is). With that preparation, the expression (7) becomes the expected value of  $Y_\ell$  over the hyper produced by abstracting from  $J = \pi \triangleright M$  as above. That is (7) gives equivalently

$$\$(\pi, M, \ell) = \mathcal{E}_{[\pi \triangleright M]} Y_\ell \quad , \quad (9)$$

in which the  $M$  and  $\pi$  now explicitly appear and where –we recall– the brackets  $[-]$  convert the joint distribution  $\pi \triangleright M$  to a hyper. (If  $Y_\ell$  were in fact Shannon entropy, then (9) would be the *conditional* Shannon entropy. But  $Y_\ell$ ’s are much more general than Shannon entropy alone [2], [9].)

Finally, using hypers we define an *abstract channel* to be a function from prior to hyper, i.e. of type  $\mathcal{X} \rightarrow \mathbb{D}^2\mathcal{X}$ , realised from some concrete channel  $M: \mathcal{X} \rightarrow \mathcal{Y}$  as  $\pi \mapsto [\pi \triangleright M]$ . It is “abstract” because the type  $\mathcal{Y}$  no longer appears: it is unnecessary because if  $M(\pi)$  is the application of  $M$  as a function applied to prior  $\pi$ , then from (9) the worst rational expected loss is written simply  $\mathcal{E}_{M(\pi)} Y_\ell$ .

(Recall from §IV-B that this naturally takes into account the “rational observers” and the remapping they might perform, as described in [1].)

#### 1) Example of a channel representation of a mechanism:

If we have a discrete input  $\mathcal{X} := \{x_0, x_1, x_2\}$ , and discrete output  $\mathcal{Y} := \{y_0, y_1, y_2, y_3, y_4\}$ , we can represent an obfuscating mechanism  $M$  with the channel  $M$  below.

$$M = \begin{bmatrix} 2/3 & 1/6 & 1/12 & 1/24 & 1/24 \\ 1/6 & 1/6 & 1/3 & 1/6 & 1/6 \\ 1/24 & 1/24 & 1/12 & 1/6 & 2/3 \end{bmatrix}$$

As described in §IV-A above, the row  $M_{x,-}$  corresponds to the probability distribution of outputs  $y$  in  $\mathcal{Y}$  for that  $x$ . For example the top left number  $2/3 = M_{x_0, y_0}$  is the probability that output  $y_0$  is observed when the input is  $x_0$ . We can interpret this as an  $\varepsilon$ -DP mechanism once we know the metric  $\mathbb{D}$  on  $\mathcal{X}$ . In particular §II-A(1) simplifies to comparing ratios of entries in the same column, and when we do that we find that for example  $\mathfrak{d}_D(M(x_0), M(x_1)) = \ln 4$ . Thus from §II-A(4) now applied to  $\mathcal{X}$  we can say that if  $M$  is  $\varepsilon$ -DP then  $\varepsilon$  satisfies

$$\mathfrak{d}_D(M(x_0), M(x_1)) = \ln 4 \leq \varepsilon \cdot \mathbb{D}(x_0, x_1) \quad .$$

2) *Example of a loss function calculation:* Now suppose that we choose a loss function known as “Bayes Risk”, br defined on  $\mathcal{X} := \{x_0, x_1, x_2\}$  as above:

$$\text{br}(w, x) := 1 \text{ if } x \neq w \text{ else } 0 \quad ,$$

where  $\mathcal{W} := \mathcal{X}$ . Letting the input prior be the uniform distribution  $\odot$  over  $\mathcal{X}$ , we can compute the loss  $\$(\odot, M, \text{br})$  by selecting for each output  $y$ , the  $w$  which makes the expected value of  $\text{br}(w, -)$  over the posterior  $p_{X|y}$  the least. We then take the expected value of these least values over the marginal  $p_Y$ . For  $y_0$  for example, that least expected value occurs at  $w = x_0$ , and for  $y_1$  it occurs *either* for  $w = x_0$  or  $w = x_0$ . Overall the total expected loss is  $1/3$ .

### D. Refinement of hypers and mechanisms

The hypers  $\mathbb{D}^2\mathcal{X}$  on  $\mathcal{X}$  have a partial order ( $\sqsubseteq$ ) “refinement” [10] that we will need in the proof of our main result. It admits several equivalent interpretations in this context. Below, we write  $\Delta$  etc. for general hypers in  $\mathbb{D}^2\mathcal{X}$ .

We have that  $\Delta \sqsubseteq \Delta'$ , that hyper  $\Delta$  is refined by hyper  $\Delta'$ , under any of these equivalent conditions:

- (a) when  $\mathcal{E}_\Delta Y_\ell \leq \mathcal{E}_{\Delta'} Y_\ell$  for *all* loss functions  $\ell$  (i.e. whether legal or not).
- (b) when considered as distributions on posteriors  $\mathbb{D}\mathcal{X}$  it is possible to convert  $\Delta$  into  $\Delta'$  via a Wasserstein-style “earth move” of probability from one posterior to another [11], [12], [8].
- (c) when generated from joint-distribution matrices  $D$  in  $\mathbb{D}(\mathcal{X} \times \mathcal{Y})$  generating  $\Delta$ , and  $D'$  in  $\mathbb{D}(\mathcal{X} \times \mathcal{Y}')$  generating  $\Delta'$ , there is a “post-processing matrix”  $R$  of type  $\mathcal{Y} \rightarrow \mathcal{Y}'$  such that as matrices we have  $D \cdot R = D'$  via matrix multiplication.

And we say that one mechanism  $M$  is refined by another  $M'$  just when  $[\pi \triangleright M] \sqsubseteq [\pi \triangleright M']$  for all priors  $\pi$ . When this occurs we also write  $M \sqsubseteq M'$ . From formulation (a) we will use the fact that the ( $\sqsubseteq$ )-infimum of the  ${}^T L_N^\varepsilon$ ’s (indexed over a sequence of  $T$ ’s) is just  $L^\varepsilon$  itself [13] and [15, Lem. 20, Appendix §B].

Formulation (b) is particularly useful. If we find a specific earth move from  $\Delta$  to  $\Delta'$  that defines a refinement we can then

<sup>5</sup>In the hyper-distribution literature these are called “inners” [8].

use the equivalent (a) to deduce that  $\mathcal{E}_\Delta Y_\ell \leq \mathcal{E}_{\Delta'} Y_\ell$ . However if we can also compute the cost <sup>6</sup> of the particular earth move we can conclude in addition that the difference  $|\mathcal{E}_\Delta Y_\ell - \mathcal{E}_{\Delta'} Y_\ell|$  must be bounded above by an amount we can compute. This follows from the well-known Kantorovich-Rubinstein duality [11] which says that  $|\mathcal{E}_\Delta Y_\ell - \mathcal{E}_{\Delta'} Y_\ell|$  is no more than minimal cost incurred by any earth move transforming  $\Delta$  to  $\Delta'$  scaled by the ‘‘Lipschitz constant’’ <sup>7</sup> of  $Y_\ell$ . We use these ideas in Lem. 11 and Thm. 13.

## V. MEASURES ON CONTINUOUS $\mathcal{X}$ AND $\mathcal{Y}$

### A. Measures via probability density functions

Continuous analogues of the  $\pi$ ,  $M$  and  $\ell$  will be our principal concern here: ultimately we will use  $\mathbb{M}[0, 1]$  for our measurable spaces  $\mathcal{X}$  and  $\mathcal{Y}$ , and will suppose for simplicity that  $\mathcal{X}=\mathcal{Y}=[0, 1]$ . (More generality is achieved by simple scaling).

Measures  $\mathbb{M}[0, 1]$  (that is  $\mathbb{M}\mathcal{X}$  and  $\mathbb{M}\mathcal{Y}$ ) will be given as probability density functions, where a PDF say  $\mu: [0, 1] \rightarrow \mathbb{R}^{\geq}$  determines the probability  $\int_a^b \mu$  assigned to the sample  $[a, b] \subseteq [0, 1]$  using the standard Borel measure on  $[0, 1]$ , and more generally the expected value of some random variable  $V$  on  $[a, b]$  given by PDF  $\mu$  is  $\int_a^b \mu(x)V(x)dx$ .

Even though  $\mu$  is of type PDF, we abuse notation to write for example  $\mu[a, b]$  for the probability  $\int_a^b \mu$  that  $\mu$  assigns to that interval, and  $\mu_a$  for the probability  $\mu$  assigns to the point  $a$  alone, i.e. some  $r$  just when when the actual PDF-value of  $\mu(a)$  is the Dirac delta-function scaled by  $r$ , written  $\delta_r$ .

### B. Continuous mechanisms over continuous priors

Our mechanisms  $M$ , up to now discrete, will now become ‘‘kontinuous’’, renamed  $K$  as a mnemonic. Thus a continuous mechanism  $K: \mathcal{X} \rightarrow \mathbb{M}\mathcal{Y}$  given input  $x$  produces measure  $K(x)$  on the observations  $\mathcal{Y}=[0, 1]$ . And given a whole continuous prior  $\pi: \mathbb{M}[0, 1]$ , that same  $K$  therefore determines a joint measure over  $\mathcal{X} \times \mathcal{Y}$ .<sup>8</sup>By analogy with (8,9) we have

**Definition 2: Continuous version of (7)** The expected loss  $\$(\pi, K, \ell)$  due to continuous prior  $\pi$ , continuous mechanism  $K$  and loss function  $\ell$  is given by <sup>9</sup>

$$\int_0^1 (\inf_w (\int_0^1 \ell(w, x)\pi(x)K(x)(y) dx)) dy. \quad (10)$$

The continuous version of uncertainty (8) is now

$$Y_\ell(\delta) := \inf_{w: \mathcal{W}} \int_0^1 \ell(w, x)\delta(x) dx$$

and the continuous version of expected loss (9) is now

$$\$(\pi, K, \ell) = \int_{y: \mathcal{Y}} Y_\ell dK(\pi) \quad .$$

<sup>6</sup>The cost is determined by the amount of ‘‘earth’’ to be moved, and the distance it must be moved. See for example [14].

<sup>7</sup>The Lipschitz constant of a function is the amount by which the difference in outputs can vary when compared to the difference in inputs.

<sup>8</sup>See [15, Appendix §A2].

<sup>9</sup>This is well defined whenever the  $\mathcal{W}$ -indexed family of functions of  $y$  given by  $\int_0^1 \ell(w, x)\pi(x)K(x)(y) dx$  contains a countable subset  $\mathcal{W}'$  such that the  $\inf$  over  $\mathcal{W}$  is equal to the  $\inf$  over  $\mathcal{W}'$  [16]. This is clear if  $\mathcal{W}$  is finite, and whenever  $\mathcal{W}'$  can be taken to be the rationals.

### C. The truncated Laplace mechanism

As for the Geometric mechanism, the Laplace mechanism is based on the Laplace distribution. It is defined as follows:

**Definition 3: (Laplace distribution)** The  $\varepsilon$ -Laplace mechanism with with input  $x$  in  $\mathcal{X}=[0, 1]$  and probability density in  $\mathbb{R}^{\geq}$  for output  $y$  in  $\mathcal{Y}$  is usually written as a PDF in  $y$  (for given  $x$ ) as [17]

$$L^\varepsilon(x, y) := \varepsilon/2 \cdot e^{-\varepsilon|y-x|} \quad .$$

The  $\varepsilon/2$  is a normalising factor. It is known [4] that the mechanism  $L^\varepsilon$  satisfies  $\varepsilon$ -DP over  $[0, 1]$  (where the underlying metric on  $\mathcal{X}$  is Euclidean). Just as for the Geometric mechanism we truncate  $L^\varepsilon$ 's outputs so that they also lie inside  $\mathcal{U}$ . We do so in the same manner, by remapping all outputs greater than 1 to 1, and all outputs less than 0 to 0.

**Definition 4: (truncated Laplace mechanism)** As earlier for  $G^\varepsilon$ , we *truncate* the Laplace mechanism  $L^\varepsilon$  to  $L^\varepsilon$  for inputs restricted to  $[0, 1]$ , and output restricted to  $[0, 1]$ , in the following way (as a PDF):

$$L^\varepsilon(x)(y) := \begin{array}{ll} \delta_a & \text{if } y=0 \\ L^\varepsilon(x, y) & \text{if } 0 < y < 1 \\ \delta_b & \text{if } y=1 \end{array} \quad ,$$

where the constants  $a, b$  are  $\int_{-\infty}^0 L^\varepsilon(x, y)dy = e^{\varepsilon x}/2$  and  $\int_1^\infty L^\varepsilon(x, y)dy = e^{\varepsilon(1-x)}/2$  respectively, and  $\delta_r$  is the Dirac delta-function with weight  $r$ .

We can now state our **principal contribution**. It is to show that the *truncated* Laplace  $L^\varepsilon$  is universally optimal, in this continuous setting, in the same way that  $G^\varepsilon$  was optimal in the discrete setting:

**Theorem 5: (truncated Laplace is optimal)** Let  $K^\varepsilon$  be any continuous  $\varepsilon$ -DP mechanism with input and output both  $[0, 1]$ , and let  $\pi$  be any continuous (prior) probability distribution over  $[0, 1]$  and  $\ell$  any Lipschitz continuous <sup>10</sup>, legal loss function on  $\mathcal{X}=\mathcal{U}$ .

Then  $\$(\pi, L^\varepsilon, \ell) \leq \$(\pi, K, \ell)$  .

As we foreshadowed in the proof outline in §III, Thm. 5 relies ultimately on the earlier-proven optimality  $G^\varepsilon$  in the discrete case: we must show how we can approximate continuous  $\varepsilon$ -DP mechanisms in discrete form, each one satisfying the conditions under which the earlier result applies, and in §VI we fill in the details. Along the way we show how the Laplace mechanism provides a smooth approximation to the Geometric- with discrete inputs.

## VI. APPROXIMATING CONTINUITY FOR $\mathcal{X}$

### A. Connecting continuous and discrete

Our principal tool for connecting the discrete and continuous settings is the evenly-spaced discrete subset  $\mathcal{U}_N =$

<sup>10</sup>Lipschitz continuous is less general than continuous. It means that the difference in outputs is within a constant  $\kappa > 0$  scaling factor of the difference between the inputs.

$\{0, 1/N, 2/N \dots, N-1/N, 1\}$  of the unit interval  $\mathcal{U}=[0, 1]$  for ever-increasing  $N>0$ .

The separation  $1/N$  is the *interval width*.

### B. Approximations of continuous priors

The  $N$ -approximation of prior  $\pi: \mathbb{M}\mathcal{U}$  of type  $\mathbb{D}\mathcal{U}_N$ , i.e. yielding actual probabilities (not densities), and is defined

$$\pi_N(n/N) := \begin{cases} \pi[n/N, n+1/N] & \text{if } n < N-1 \\ \pi[n/N, 1] & \text{if } n = N-1 \\ 0 & \text{otherwise} \end{cases} .$$

The discrete  $\pi_N$  gathers each of the continuous  $\pi$ -interval's measure onto its left point, with as a special case  $[1, 1]$  from  $\pi$  included onto the point  $N-1/N$  of  $\pi_N$ .

As an example take  $N$  to be 2, and  $\pi$  to be the uniform (continuous) distribution over  $\mathcal{U}$ , which can be represented by the constant 1 PDF. Since the interval width is  $1/2$ , we see that  $\pi_N$  assigns probability  $1/2$  to both 0 and  $1/2$  and zero to all other points in  $\mathcal{U}$ .

### C. $N$ -step mechanisms and loss functions

In the other direction, we can lift discrete  $M$  and loss-function  $\ell$  on  $\mathcal{U}_N$  into the continuous  $\mathcal{X}=\mathcal{U}$  by replicating their values for the  $x$ 's *not* in  $\mathcal{U}_N$  in a way that constructs  $N$ -step functions: we have

*Definition 6:* For  $x$  in  $\mathcal{U}=[0, 1]$  define  $\lfloor x \rfloor_N := \lfloor Nx \rfloor / N$ .

*Definition 7:* Given mechanism  $M: \mathcal{U}_N \rightarrow \mathcal{Y}$ , define  $M_N: [0, 1] \rightarrow \mathbb{R}^{\geq}$  so that

$$M_N(x) := \begin{cases} M(\lfloor x \rfloor_N) & \text{if } 0 \leq x < 1 \\ M(N-1/N) & \text{if } x = 1 \end{cases} .$$

Note that we have not yet committed here to whether  $M$  produces discrete or continuous distributions on its *output*  $\mathcal{Y}$ . We are concentrating only on its *input* (from  $\mathcal{X}$ ).

Similarly, given loss function  $\ell: \mathcal{W} \times \mathcal{U}_N \rightarrow \mathbb{R}^{\geq}$ , define  $\ell_N: \mathcal{W} \times [0, 1] \rightarrow \mathbb{R}^{\geq}$  so that

$$\ell_N(w, x) := \begin{cases} \ell(w, \lfloor x \rfloor_N) & \text{if } 0 \leq x < 1 \\ \ell(w, N-1/N) & \text{if } x = 1 \end{cases} .$$

Say that mechanisms and loss functions over  $[0, 1]$  are  *$N$ -step functions* just when they are constructed as above.

The important property enabled by the above definitions is the correspondence between loss functions' values in their pixelated and original versions, which will allow us to apply the earlier discrete-optimality result, based on Lem. 9 to come. That is, we have

*Lemma 8:* For any continuous prior  $\pi$  in  $\mathbb{M}\mathcal{U}$ , mechanism  $M$  in  $\mathcal{U} \rightarrow \mathcal{Y}$  and loss function  $\ell$  in  $\mathcal{W} \times \mathcal{U} \rightarrow \mathbb{R}^{\geq}$  we have

$$\underbrace{\$(\pi, M_N, \ell_N)}_{\text{continuous } \mathcal{X}} = \underbrace{\$(\pi_N, M, \ell)}_{\text{discrete } \mathcal{X}} .$$

That is, the loss realised via a pixelated  $\pi_N$ , and (already discrete)  $M$  and  $\ell$ , all operating on  $\mathcal{U}_N$ , is the same as the loss realised via the original continuous  $\pi$  and the lifted (and thus  $N$ -step) mechanism  $M_N$  and  $\ell_N$ , now operating over all of  $\mathcal{X}=\mathcal{U}$ .

*Proof:* We interpret the losses using Def. 2, focussing on the integrand of the inner integral. Note that we can split it up into a finite sum of integrals of the form  $\int_{n/N}^{n+1/N} \pi(x) V(x) dx$ . When we do that for the left-hand formula  $\$(\pi, M_N, \ell_N)$  we see that throughout the interval  $[n/N, n+1/N)$  the contribution of the mechanism and the loss is constant, i.e.  $M_N(x)(y) \cdot \ell_N(w, x) = M(n/N)(y) \cdot \ell(w, n/N)$ . This means the integral becomes

$$M(n/N)(y) \cdot \ell(w, n/N) \cdot \int_{n/N}^{n+1/N} \pi(x) dx$$

which is equal to  $M(n/N)(y) \cdot \ell(w, n/N) \cdot \pi_N(n/N)$ . A similar argument applies to the last interval (which includes 1), compensated for by the definitions of  $\ell_N$  and  $M_N$  to take their corresponding values from  $1-N/N$ .

Looking now at the right-hand formula,  $\$(\pi_N, M, \ell)$  we see that it is now exactly the finite sum of the integrals just described. ■

### D. Approximating continuous $\varepsilon$ -DP mechanisms

The techniques above give good discrete approximations for continuous-input  $\varepsilon$ -DP mechanisms  $M$  acting on continuous priors simply by considering  $M_N$ 's for increasing  $N$ 's, using §VI-C. As a convenient abuse of notation, when we *start* with a continuous-input mechanism  $M$  on  $[0, 1]$  we write  $M_N$  to mean the  $N$ -step mechanism that is made by first restricting  $M$  to the subset  $\mathcal{U}_N$  of  $[0, 1]$  and then lifting that restriction “back again” as in Def. 7, effectively converting it into an  $N$ -step function. When we do this we find that the posterior loss wrt.  $N$ -step loss functions can be bounded above and below by using pixelated priors and  $N$ -stepped mechanisms.

*Lemma 9:* Let  $K$  be a continuous-input  $\varepsilon$ -DP mechanism, and  $\pi$  in  $\mathbb{M}[0, 1]$  a continuous prior and  $\ell$  a (non-negative)  $N$ -step function. Then the following inequalities hold:

$$e^{-\frac{\varepsilon}{N}} \cdot \underbrace{\$(\pi_N, K_N, \ell)}_{\text{discrete } \mathcal{X}} \leq \underbrace{\$(\pi, K, \ell)}_{\text{continuous } \mathcal{X}} \leq e^{\frac{\varepsilon}{N}} \cdot \underbrace{\$(\pi_N, K_N, \ell)}_{\text{discrete } \mathcal{X}} .$$

(Notice that the middle formula  $\$(\pi, K, \ell)$ , the mechanism  $K$  is not  $N$ -stepped, but in the formulae on either side they are as in Lem. 8.)

*Proof:* The proof is as for Lem. 8, but noting also that  $K$ 's being  $\varepsilon$ -DP implies that for all  $N$  we have  $K(\lfloor x \rfloor_N)(y) \times e^{-\frac{\varepsilon}{N}} \leq K(x)(y) \leq K(\lfloor x \rfloor_N)(y) \times e^{\frac{\varepsilon}{N}}$ .<sup>11</sup> ■

With Lem. 8 and Lem. 9 we can study optimality of  $\mathcal{L}^\varepsilon$  on finite discrete inputs  $\mathcal{U}_N$ . We will see that, although Geometric mechanisms are still optimal for the (effectively) discrete inputs  $\mathcal{U}_N$ , the Laplace mechanism provides increasingly good *approximate optimality* for  $\mathcal{U}_N$  as  $N$  increases, and is in fact (truly) optimal in the limit.

<sup>11</sup>Here we are using the  $\varepsilon$ -DP-constraints applied to the PDF  $K(x)(y)$ .

## VII. THE LAPLACE AND GEOMETRIC MECHANISMS

In this section we make precise the restriction of the Geometric mechanism  $G^\varepsilon$  (§II-D1) to inputs and outputs both in  $\mathcal{U}_N$  (a subset of  $[0, 1]$ ): for both  $x, y$  in  $\mathcal{U}_N$  we define

$$\underbrace{G_N^\varepsilon(x)(y)}_{\text{on } \mathcal{U}_N} := \underbrace{G_{\frac{\varepsilon}{N}}(Nx)(Ny)}_{\text{on } (0..N)} . \quad (11)$$

As an illustration, we take  $\varepsilon=2\ln 4$  and input  $\mathcal{X}=\mathcal{U}_2$ , in which the 2 comes from  $\mathcal{U}_2$  and the  $\ln 4$  comes from the  $\alpha=1/4$  of the Geometric distribution used to make the mechanism  $G^\varepsilon$ . Using the *three* points 0,  $1/2$  and 1 of the input, we compute the truncated geometric mechanism  $G_2^\varepsilon$  as the channel below, where the rows' labels are (invisibly) the inputs  $\mathcal{U}_2$ , and the columns are similarly labelled by the outputs (also  $\mathcal{U}_2$  in this case). This means that if the input was 0, then the output (after truncation) will be 0 with probability  $4/5$ , and  $1/2$  with probability  $3/20$  etc:

$$G_2^\varepsilon = \begin{bmatrix} 4/5 & 3/20 & 1/20 \\ 1/5 & 3/5 & 1/5 \\ 1/20 & 3/20 & 4/5 \end{bmatrix} .$$

Notice now that the ratio of adjacent probabilities that are in the same column satisfy the  $\varepsilon$ -DP constraint, so for example  $4/5 \div 1/5 = 3/5 \div 3/20 = 4 \leq e^{(2\ln 4)/2}$ . Notice also that the distance between adjacent inputs in  $\mathcal{U}_2$  under the Euclidean distance is  $1/2$ , not 1 as it would be in the conventional  $\mathcal{X}=(0, 1, 2)$ .

Suppose now that we consider  $\mathcal{U}_4$  instead, consisting of the five points 0,  $1/4, 1/2, 3/4$  and 1, and we adjust the  $\alpha$  in the underlying Geometric distribution  $G_\alpha$  from §II-D1(5). The  $\varepsilon$ -DP parameter  $\varepsilon$ , now  $4\ln 2$ , is the *same* as before — and the resulting matrix is

$$G_4^\varepsilon = \begin{bmatrix} 2/3 & 1/6 & 1/12 & 1/24 & 1/24 \\ 1/3 & 1/3 & 1/6 & 1/12 & 1/12 \\ 1/6 & 1/6 & 1/3 & 1/6 & 1/6 \\ 1/12 & 1/12 & 1/6 & 1/3 & 1/3 \\ 1/24 & 1/24 & 1/12 & 1/6 & 2/3 \end{bmatrix}$$

As before though, the ratio of adjacent probabilities that are in the same column satisfy the  $\varepsilon$ -DP-constraint over all of  $\mathcal{U}_4$ : now we have  $2/3 \div 1/3 = 1/3 \div 1/2 = 2 \leq e^{(4\ln 2)/4}$ .

This amplifies the explanation in (2) that the  $\varepsilon$ -DP constraints over discrete inputs  $\mathcal{U}_N$  must take into account the underlying metric on the input space. More generally, whenever we double  $N$  in  $\mathcal{U}_N$ , the  $\alpha$ -parameter must become  $\sqrt{\alpha}$ .

At this point, we have enough to be able to appeal to the discrete optimality result, to bound below the losses for continuous mechanisms, provided that the loss  $\ell_N$  is  $N$ -legal, i.e. that its legality obtains at least for the distinct points in  $\mathcal{U}_N$ .

*Lemma 10:* For any continuous prior  $\pi$  in  $\mathbb{M}\mathcal{U}$ ,  $\varepsilon$ -DP-mechanism  $M:\mathcal{U}\rightarrow\mathcal{Y}$  and loss function  $\ell:\mathcal{W}\times\mathcal{U}\rightarrow\mathbb{R}^\geq$  such that  $\ell_N$  is  $N$ -legal, we have:

$$\$(\pi_N, G_N^\varepsilon, \ell_N) \leq \$(\pi, M_N, \ell_N)$$

*Proof:* Follows from Lem. 8 and noting that  $M$  restricted to  $\mathcal{U}_N$  satisfies the conditions for universal discrete optimality [1]. ■

Our next task is to study the relationship between the Geometric- and Laplace mechanisms. We show first that  $G_N^\varepsilon$  is refined (§IV-D) by the truncated Laplace mechanism also restricted to  $\mathcal{U}_N$ . Since  $\mathcal{L}^\varepsilon$  is already defined over the whole of  $\mathcal{U}$  we continue to write its restriction to  $\mathcal{U}_N$  as  $\mathcal{L}^\varepsilon$ . This will immediately show that losses under the Geometric are no more than those under the Laplace (§IV-D(1)), consistent with observations that, on discrete inputs, Laplace obfuscation does not necessarily minimise the loss. Since the output  $\mathcal{Y}$  of  $\mathcal{L}^\varepsilon$  is continuous, we proceed by first approximating it using post-processing to make Laplace-based mechanisms  ${}^T\mathcal{L}^\varepsilon$ , defined below, which have discrete output, and which can form an anti-refinement chain converging to  $\mathcal{L}^\varepsilon$ . We are then able to show separately the refinements between  $G_N^\varepsilon$  and  ${}^T\mathcal{L}^\varepsilon$ , using methods designed for finite mechanisms.

The  $T, N$ -Laplace mechanisms approximate  $\mathcal{L}^\varepsilon$  by  $T$ -pixelation of their outputs. Here  $x$  is (still) in  $\mathcal{U}_N$  but  $y$  is in  $\mathcal{U}_T$ .

$${}^T\mathcal{L}^\varepsilon(x)(y) := \begin{cases} \mathcal{L}^\varepsilon(x)[y, y+1/T] & \text{if } y < 1-1/T \\ \mathcal{L}^\varepsilon[1-1/T, 1] & \text{otherwise.} \end{cases} \quad (12)$$

That is, we pixelate the  $\mathcal{Y}$  using  $T$  for the Laplace (independently of the  $N$  we use for  $\mathcal{X}$ .) This is illustrated in Fig. 1a.

Observe that as this is a post-processing (§IV-D(3)) of the output of  $\mathcal{L}^\varepsilon$ , the refinement  $\mathcal{L}^\varepsilon \sqsubseteq {}^T\mathcal{L}^\varepsilon$  follows.

### A. Refinement between $N$ -Geometric and $T, N$ -Laplace mechanisms

We now demonstrate the crucial fact that  $G_N^\varepsilon$  is refined by  ${}^T\mathcal{L}^\varepsilon$ . We use version (b) of refinement, described in §IV-D, and establish a Wasserstein-style earth-move between hypers  $[\odot \triangleright G_N^\varepsilon]$  and  $[\odot \triangleright {}^T\mathcal{L}^\varepsilon]$  (i.e. for uniform prior  $\odot$ ).

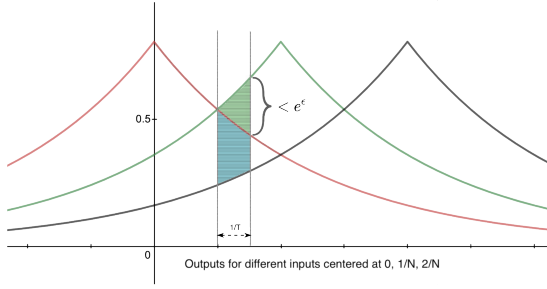
*Lemma 11:* For all integer  $T > 0$  we have that  $G_N^\varepsilon \sqsubseteq {}^T\mathcal{L}^\varepsilon$ .

*Proof:* Take  $\Delta, \Delta'$  in  $\mathbb{D}^2\mathcal{U}_N$  as hypers both with finite supports. We can depict such hypers in  $\mathbb{R}^{N+1}$ -space by locating their supports, each of which is a point in  $\mathbb{R}^{N+1}$ , where the axes of the diagram correspond to each point in  $\mathcal{U}_N$ . For example if we take  $\Delta$  to be the hyper-distribution  $[\odot \triangleright G_2^\varepsilon]$ , it has three posterior distributions, which are 1-summing triples in  $\mathbb{R}^3$ . They are depicted by the orange points in Fig. 1. Similarly the supports of the a hyper-distribution  $\Delta'$  taken to be  $[\odot \triangleright {}^T\mathcal{L}^\varepsilon]$  are represented by the blue dots. Notice that the blue dots are contained in the convex hull of the orange dots, and this observation allows us to prove that the mechanisms  $G_2^\varepsilon$  and  ${}^8\mathcal{L}^\varepsilon$  are in a refinement relation.

We use the following fact [8, Lem. 12.2] about refinement ( $\sqsubseteq$ ).

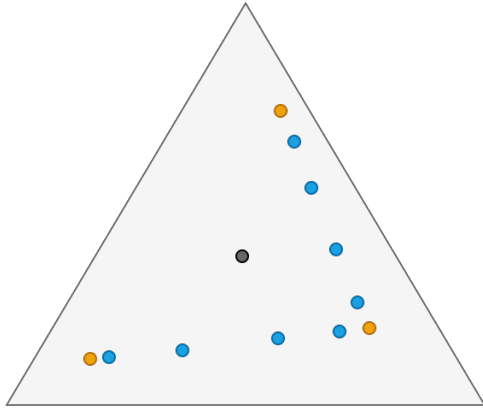
Let  $C, C':\mathcal{U}_N \rightarrow \mathcal{U}_T$  be channels and let  $\odot$  be the uniform prior. If the supports of  $[\odot \triangleright C]$  are linearly independent when considered as vectors in  $\mathbb{R}^N$ , and their convex hull encloses the supports of  $[\odot \triangleright C']$ , then  $C \sqsubseteq C'$ .<sup>12</sup>





The width of the central “vertical slice” is  $1/T$ .

(a) Illustrates batching the output for  $T\mathcal{L}$  (similar for  $T\mathcal{L}^\varepsilon$ ). The outputs (shown here as PDF plots) are batched into output segments of length  $1/T$  in this example, for  $T=8$ . The segment from  $[x, x+1/T)$  is indicated by the two vertical lines. The probability assigned to this segment is the area under the relevant curves. For the red curve it is the sum of the white and blue regions; the green curve it is the sum of the white, blue and green regions and for the black curve it is only the white region.



(b) The supports of hypers  $[\circlearrowright G_2^\varepsilon]$  (orange) and  $[\circlearrowright 8\mathcal{L}^\varepsilon]$  (blue) for inputs  $\{0, 1/2, 1\}$  placed within the (triangular) probability simplex. The blue points are within in the convex hull of the orange points.

Fig. 1:  $N$ -geometric and  $T, N$ -Laplace mechanisms.

To apply this result, we let  $C$  be  $G_N^\varepsilon$  recall that indeed the supports of  $[\circlearrowright G_N^\varepsilon]$  are linearly independent (see for example [5]). Moreover in general, the supports of  $[\circlearrowright T\mathcal{L}^\varepsilon]$  are also contained in the convex hull. We provide details of this latter fact in [15, Appendix §B]. ■

Finally we can show full refinement between the Laplace and the Geometric mechanism, which follows from continuity of refinement [13].

*Theorem 12:*  $G_N^\varepsilon \sqsubseteq \mathcal{L}^\varepsilon$ .

*Proof:* We first form an anti-refinement chain  $\dots \sqsubseteq T_1\mathcal{L}^\varepsilon \sqsubseteq T_0\mathcal{L}^\varepsilon$  so that (a)  $\mathcal{L}^\varepsilon \sqsubseteq T_i\mathcal{L}^\varepsilon$  for all  $i$ , and (b) the chain converges to  $\mathcal{L}^\varepsilon$ .

<sup>12</sup>The lemma applies to channels because of the direct correspondence between channels and the supports of hyper-distributions formed from uniform priors.

We reason as follows:

$$\begin{aligned} & G_N^\varepsilon \sqsubseteq \mathcal{L}^\varepsilon \\ \text{iff} & \quad \quad \quad \text{“}\sqsubseteq\text{ is continuous; (a), (b) above”} \\ & G_N^\varepsilon \sqsubseteq T_i\mathcal{L}^\varepsilon \quad \text{for all } i \geq 0 \end{aligned}$$

which follows from Lem. 11. We provide details of (a), (b) just above in [15, Appendix §B]. ■

We have shown that the Laplace mechanism is a refinement of the Geometric mechanism. This means that it genuinely leaks less information than does the Geometric mechanism and therefore affords greater privacy protections. On the other hand this means that we have lost utility with respect to the aggregated information. In the next section we turn to the comparison of the Laplace and Geometric mechanisms with respect to that loss.

### B. The Laplace approximates the Geometric

The geometrical interpretation of the Laplace and Geometric mechanisms set out above indicates how the Laplace approximates the Geometric as  $\mathcal{U}_N$ 's interval-width approaches 0. In particular the refinement relationship established in Thm. 12 describes how the posteriors of  $[\circlearrowright T\mathcal{L}^\varepsilon]$  all lie in between pairs of posteriors of  $[\circlearrowright G_N^\varepsilon]$ . This relationship between posteriors translates to a bound between the corresponding expected losses  $\$(\circlearrowright, \mathcal{L}^\varepsilon, \ell)$  and  $\$(\circlearrowright, G_N^\varepsilon, \ell)$  via the Kantorovich-Rubinstein theorem applied to the hypers  $[\circlearrowright T\mathcal{L}^\varepsilon]$  and  $[\circlearrowright G_N^\varepsilon]$ . We sketch the argument in the next theorem, and provide full details in [15, Appendix §D].

*Theorem 13:* Let  $\ell$  be a  $\kappa$ -Lipschitz loss function, and  $\circlearrowright$  the uniform distribution over  $\mathcal{U}_N$ . Then

$$\$(\circlearrowright, \mathcal{L}^\varepsilon, \ell) - \$(\circlearrowright, G_N^\varepsilon, \ell) \leq c\kappa/N, \quad (13)$$

where  $c = 3/(1-e^{-\varepsilon})^2$  is constant for fixed  $\varepsilon$ .

*Proof:* We appeal to the Kantorovich-Rubinstein theorem which states that the “Kantorovich distance” between probability distributions  $\Delta, \Delta'$  bounds above the difference between expected values  $|\mathcal{E}_\Delta f - \mathcal{E}_{\Delta'} f|$  whenever  $f$  satisfies the  $\kappa$ -Lipschitz condition. In our case the relevant distributions are the *hyper*-distributions  $[\circlearrowright T\mathcal{L}^\varepsilon]$  and  $[\circlearrowright G_N^\varepsilon]$ , and the relevant Lipschitz functions are  $Y_\ell$  for loss functions  $\ell$ .<sup>13</sup>

We write  $\mathbb{W}(\Delta, \Delta')$  for the Wasserstein distance between hyper-distributions  $\Delta, \Delta'$  which is determined by the minimal *earth-moving* cost to transform  $\Delta$  to  $\Delta'$ . For any such earth move each posterior  $\delta$  of  $\Delta$  is reassigned to a selection of posteriors of  $\Delta'$  in proportion to the probability mass that  $\Delta$  assigns to  $\delta$ . The cost of the move is the expected value of the distance between posterior reassignment (weighted by the proportion of the reassignment). Thus the cost of any specific earth move provides an upper bound to  $\mathbb{W}(\Delta, \Delta')$ .<sup>14</sup> Importantly for us, the relation of refinement  $\sqsubseteq$  determines a specific earth move [8] whose cost we can calculate.

<sup>13</sup>Some  $f: \mathbb{D}\mathcal{X} \rightarrow \mathbb{R}$  is  $\kappa$ -Lipschitz if  $|f(\delta) - f(\delta')| \leq \kappa\mathbb{W}(\delta, \delta')$ , for  $\kappa > 0$ , and  $\mathbb{W}(\delta, \delta')$  is the Kantorovich distance between  $\delta, \delta'$ .

<sup>14</sup>All the costs are determined by the underlying metric used to define the probability distributions. For us this is determined by the Euclidean distance on the interval  $[0, 1]$ .

Referring to Lem. 11 and Fig. 1, we see that the refinement between the approximation to the Laplace  $[\circlearrowright^T \mathcal{L}^\varepsilon]$  and  $[\circlearrowright^T G_N^\varepsilon]$ , reassigns the Geometric's posteriors (the orange dots) to the Laplace's posteriors (the blue dots). Crucially though the Geometric's posteriors form a linear order according to their distance from one another, and the refinement described in Lem. 11 shows how each Laplace posterior lies in between adjacent pairs of Geometric posteriors (according to the linear ordering), provided that  $N$  divides  $T$ . Therefore any redistribution of a Geometric posterior is bounded above by the distance to one or other of its adjacent posteriors. We show in [15, Appendix §D] that distances between adjacent pairs is bounded above by  $c/N$ , and therefore  $\mathbb{W}([\circlearrowright^T \mathcal{L}^\varepsilon], [\circlearrowright^T G_N^\varepsilon]) \leq c/N$ .

Next we observe that if  $\ell(w, x)$  is a  $\kappa$ -Lipschitz function on  $[0, 1]$  (as a function of  $x$ ), then  $Y_\ell$  is a  $\kappa$ -Lipschitz function, and so by the Kantorovich-Rubinstein theorem we must have (recalling from (9)) that  $\$(\pi, M, \ell) = \mathcal{E}_{[\pi \triangleright M]} Y_\ell$ :

$$\$(\circlearrowright^T \mathcal{L}^\varepsilon, \ell) - \$(\circlearrowright^T G_N^\varepsilon, \ell) \leq c\kappa/N. \quad (14)$$

By Thm. 12 and post-processing we see that  $G_N^\varepsilon \sqsubseteq \mathcal{L}^\varepsilon \sqsubseteq T \mathcal{L}^\varepsilon$ . Recall from (a) that refinement means that the corresponding losses are also ordered, i.e.

$$\$(\circlearrowright^T G_N^\varepsilon, \ell) \leq \$(\circlearrowright^T \mathcal{L}^\varepsilon, \ell) \leq \$(\circlearrowright^T T \mathcal{L}^\varepsilon, \ell)$$

and so the difference  $\$(\circlearrowright^T \mathcal{L}^\varepsilon, \ell) - \$(\circlearrowright^T G_N^\varepsilon, \ell)$  must be no more than the difference  $\$(\circlearrowright^T T \mathcal{L}^\varepsilon, \ell) - \$(\circlearrowright^T G_N^\varepsilon, \ell)$ , thus (13) follows from (14). Full details are set out in [15, Appendix §D].  $\blacksquare$

More generally, (13) holds whatever the prior.

*Theorem 14:* Let  $\ell$  be a  $\kappa$ -Lipschitz loss function, and  $\pi$  any prior over  $\mathcal{U}_N$ . Then

$$\$(\pi, \mathcal{L}^\varepsilon, \ell) - \$(\pi, G_N^\varepsilon, \ell) \leq c\kappa/N. \quad (15)$$

*Proof:* This follows as for Thm. 13, by direct calculation, noting that for discrete distributions we have  $\$(\circlearrowright^T M, \ell^*) = \$(\pi_N, M, \ell)$ , where  $\ell^*(w, x) := \ell(w, x) \times \pi_N(x) \times N$ . Details are given in [15, Appendix §D].  $\blacksquare$

### C. Approximating monotonic functions

The final piece needed to complete our generalisation of the Ghosh et al.'s optimality theorem is monotonicity. We describe here how to approximate continuous monotonic functions, and expose the limitations for the Laplace mechanism.

*Definition 15:* The loss function  $\ell : \mathcal{W} \times \mathcal{X} \rightarrow \mathbb{R}$  is said to be *monotone* if: there is some mapping  $\theta : \mathcal{W} \rightarrow [0, 1]$ , such that

$$\ell(w, x) := m(|\theta(w) - x|, x),$$

where  $m : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is monotone in its first argument.

Notice how  $\theta$  takes care of any remapping that might need to be applied for computing expected losses. Interestingly step functions are not in general monotone on the whole of the continuous input  $[0, 1]$ , but fortunately they are for the restrictions to  $\mathcal{U}_N$  that we need.

*Lemma 16:* Let  $\ell$  be monotone. Then the approximation  $\ell_T$  restricted to  $\mathcal{U}_N$  is monotone whenever  $T$  is a multiple of  $N$ .

*Proof:* If  $x \in \mathcal{U}_N$  then  $\lfloor x \rfloor_T = x$  since  $N$  divides  $T$ .  $\blacksquare$

Examples of continuous monotone loss functions include  $\ell_{\text{en}}$  and  $\ell_{\text{en}^2}$ , where  $x, w \in [0, 1]$ , and

$$\ell_{\text{en}}(w, x) := |x - w|. \quad (16)$$

Note that  $\ell_{\text{en}}$  is 1-Lipschitz and  $\ell_{\text{en}^2}$  is 2-Lipschitz.

We note finally that as the pixellation of  $N$  of  $\ell$  increases the approximations  $\ell_N$  converge to  $\ell$ .

## VIII. UNIVERSAL OPTIMALITY FOR THE LAPLACE MECHANISM

We finally have all the pieces in place to prove our main result, Thm. 5 from §V-C — the generalisation of discrete optimality [1].

Let  $K^\varepsilon$  be any continuous  $\varepsilon$ -DP mechanism with input  $[0, 1]$ , and let  $\pi$  be a (continuous) probability distribution over  $[0, 1]$  and  $\ell$  a legal (i.e. continuous, monotone,  $\kappa$ -Lipschitz) loss function. Then:

$$\$(\pi, \mathcal{L}^\varepsilon, \ell) \leq \$(\pi, K^\varepsilon, \ell). \quad (17)$$

*Proof:* We use the above results to approximate the expected posterior loss by step functions; these approximations are equivalent to posterior losses over discrete mechanisms satisfying  $\varepsilon$ -DP enabling appeal to Ghosh et al.'s universal optimality result on discrete mechanisms. We reason as follows:

$$\begin{aligned} & \$(\pi, K^\varepsilon, \ell_N) \times e^{\varepsilon/N} \\ & \geq \$(\pi_N, K_N^\varepsilon, \ell_N) && \text{“Lem. 9”} \\ & \geq \$(\pi_N, G_N^\varepsilon, \ell_N) && \text{“Lem. 10: } \ell_N \text{ is legal by Lem. 16”} \\ & \geq && \text{“Thm. 14; } \ell_N \text{ is } \kappa\text{-Lipschitz”} \\ & \geq \$(\pi_N, \mathcal{L}^\varepsilon, \ell_N) - c\kappa/N \\ & \geq \$(\pi, \mathcal{L}^\varepsilon, \ell_N) \times e^{-\varepsilon/N} - c\kappa/N. && \text{“Lem. 9”} \end{aligned}$$

The result now follows as above by taking  $N$  to  $\infty$ , and noting that  $e^{\varepsilon/N}$ ,  $e^{-\varepsilon/N}$ ,  $c\kappa/N$  and  $\ell_N$  converge to 1, 1, 0,  $\ell$  respectively, and that taking expected values over fixed distributions is continuous.  $\blacksquare$

Note that Thm. 5 only holds for mechanisms that are  $\varepsilon$ -DP. An arbitrary embedding  $K_N$  is not necessarily  $\varepsilon$ -DP, and in particular Thm. 5 does not apply to  $G_N^\varepsilon$ . Also the continuous property on  $\ell$  is required, because  $\ell_N$  must be monotone for all  $N$ . Thus arbitrary step functions do not satisfy this property, and so the Laplace mechanism is not in general universally optimal wrt. arbitrary step functions. Two popular loss functions however are continuous, and thus we have the following corollary.

*Corollary 17:* The Laplace mechanism is universally optimal for  $\ell_{\text{en}}$  and  $\ell_{\text{en}^2}$ .

## IX. RELATED WORK

The study of (universally) optimal mechanisms is one way to understand the cost of obfuscation, needed to implement privacy, but with a concomitant loss of utility of queries to databases. Pai and Roth [18] provide a detailed survey of the

principles underlying the design of mechanisms including the need to trade utility with privacy, and Dinur et al. [19] explore the relationship between how much noise needs to be added to database queries relative to the usefulness of the data released. Our use of loss functions to measure utility follows both that of Ghosh et al. [1] and Alvim et al. [9], and concerns optimality for entire mechanisms that satisfy a particular level of  $\epsilon$ -DP. For example, the mean error  $len$  and the mean squared error  $len^2$  can be used to evaluate loss, as described by Ghosh et al. [1] and mentioned in §VII-C.

The Laplace mechanism as a way to implement differential privacy has been shown for example by Dwork and Roth [20]. Moreover Chatzikokolakis et al. [4] showed how it satisfied  $\epsilon$ -DP-privacy as formulated here using the Euclidean metric.

Whilst rare, optimality results avoid the need to design bespoke implementations of privacy mechanisms that are tailored to particular datasets. The Geometric mechanism appears to be special for discrete inputs, as Ghosh et al. [1] showed when utility is measured using their “legal” loss functions. On the other hand, although the Laplace mechanism continues to be a popular obfuscation mechanism, its deficiencies in terms of utility have been demonstrated by others when the inputs to the obfuscation are discrete [21], and where the optimisation is based on minimising the probability of reporting an incorrect value, subject to the  $\epsilon$ -DP-constraint. Similarly Geng et al. [22] show that adding noise according to a kind of “pixellated” distribution appears to produce the best utility for arbitrary discrete datasets. Such examples are consistent with our Thm. 12 showing where the Laplace mechanism is a refinement of the Geometric mechanism (loses more utility) when restricted to a discrete input (to the obfuscation). We mention also that optimal mechanisms have also been studied by Gupte et al. [23] wrt. minimax agents, rather than average-case minimising agents.

Other works have shown the Laplace mechanism is optimal for metric differential privacy in particular non-Bayesian scenarios. Koufogiannis et al. [24] show that the Laplace mechanism is optimal for the mean-squared error function under Lipschitz privacy, equivalent to metric differential privacy; and Wang et al. [25] show that the Laplace mechanism minimises loss measured by Shannon entropy, again for metric differential privacy. Our result on  $\mathbb{R}$  includes those results as specific cases; however, those works do go further in that they demonstrate optimality for their respective loss functions on  $\mathbb{R}^n$ . We leave the study of these domains in the Bayesian setting to future work.

We also note that the linear ordering of the underlying query results seems to be important for finding optimality results. For example Brenner and Nissim [26] have demonstrated that for non-linearly ordered inputs, there are no optimal  $\epsilon$ -DP-mechanisms for a fixed level of  $\epsilon$ . Although their result finds that only counting queries have optimal mechanisms, their context (oblivious mechanisms on database queries) does not include the possibility of continuous valued query results with a linear order; our result does not contradict their impossibility, it can be seen rather as an extension of this result to a

continuous setting.

Alvim et al. [27] also use the framework of Quantitative Information Flow to study the relationship between the privacy and the utility of  $\epsilon$ -DP mechanisms. In their work they model utility in terms of a leakage measure, where leakage is defined as the ratio of input gain to output gain wrt. a mechanism modelled as a channel. Their *gain* is entirely dual to our *loss* here, and is a model of an adversary trying to infer as much as possible about the secret input. Other notions of optimality have also been studied in respect of showing that the Laplace mechanism is not optimal, including [28] who work with “near instance” optimality, and Geng and Viswanath [22] show how to scale the Laplace in various ways to obtain good utility. Note also that these definitions of optimality do not use a prior, and therefore represent the special case of utility per exact input, rather than in a scenario where the observer’s prior knowledge is included.

The use of the Laplace mechanism in real privacy applications has been demonstrated by Chatzikokolakis et al. [4] for geolocation privacy, and [29] for for privacy-preserving graph analysis, and Phan et al. [30] in deep learning.

Information-theoretic channel models for studying differential privacy were originally proposed by Alvim et al. [7], [27], and extended to arbitrary metrics in [4].

## X. CONCLUSION

We have studied the relationship between differential privacy (good) and loss of utility (bad) when the input  $\mathcal{X}$  can be over an interval of the reals, rather than having  $\mathcal{X}$  described as in the optimality result of Ghosh et al. [1], [31], i.e. as a discrete space. Here we have instead used as input space the continuous interval  $[0, 1]$ ; but we note that the result extends straightforwardly to any finite interval  $[a, b]$  of  $\mathbb{R}$ . Our result also imposes the condition that the loss functions must be  $\kappa$ -Lipschitz for some finite  $\kappa$ . We do not know whether this condition can be removed in general.

We observe that for  $N$ -step loss functions, the Laplace mechanism is not optimal, and in fact a bespoke Geometric mechanism *will* be optimal for such loss functions. However our Thm. 14 provides a way to estimate the error, relative to the optimal loss, when using the Laplace mechanism.

Finally we note that the space of  $\epsilon$ -DP mechanisms is very rich, even for discrete inputs, suggesting that the optimality result given here will be useful whenever the input domain can be linearly ordered.

### Acknowledgements

We thank Catuscia Palamidessi for suggesting this problem to us.

## APPENDIX

The appendices may be found at [15].

## REFERENCES

- [1] A. Ghosh, T. Roughgarden, and M. Sundarajan, "Universally utility-maximising privacy mechanisms," *SIAM J. COMPUT.*, vol. 41, no. 6, pp. 1673–1693, 2012.
- [2] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, "Additive and multiplicative notions of leakage, and their capacities," in *IEEE 27th Computer Security Foundations Symposium, CSF 2014, Vienna, Austria, 19-22 July, 2014*. IEEE, 2014, pp. 308–322. [Online]. Available: <http://dx.doi.org/10.1109/CSF.2014.29>
- [3] C. Dwork, F. McSherry, K. Nissim, and A. D. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, ser. Lecture Notes in Computer Science, S. Halevi and T. Rabin, Eds., vol. 3876. Springer, 2006, pp. 265–284. [Online]. Available: [https://doi.org/10.1007/11681878\\_14](https://doi.org/10.1007/11681878_14)
- [4] K. Chatzikokolakis, M. Andrés, N. Bordenabe, and C. Palamidessi, "Broadening the scope of differential privacy using metrics," in *International Symposium on Privacy Enhancing Technologies Symposium*, ser. LNCS, vol. 7981. Springer, 2013.
- [5] K. Chatzikokolakis, N. Fernandes, and C. Palamidessi, "Comparing systems: Max-case refinement orders and application to differential privacy," in *Proc. CSF*. IEEE Press, 2019.
- [6] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [7] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, and C. Palamidessi, "On the relation between differential privacy and quantitative information flow," in *Automata, Languages and Programming - 38th International Colloquium, ICALP 2011, Zurich, Switzerland, July 4-8, 2011, Proceedings, Part II*, 2011, pp. 60–76. [Online]. Available: [https://doi.org/10.1007/978-3-642-22012-8\\_4](https://doi.org/10.1007/978-3-642-22012-8_4)
- [8] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, *The Science of Quantitative Information Flow*, ser. Information Security and Cryptography. Springer International Publishing, 2020.
- [9] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, "Measuring information leakage using generalized gain functions," in *Proc. 25th IEEE Computer Security Foundations Symposium (CSF 2012)*, Jun. 2012, pp. 265–279.
- [10] A. McIver, C. Morgan, L. Meinicke, G. Smith, and B. Espinoza, "Abstract channels, gain functions and the information order," in *FCS 2013 Workshop on Foundations of Computer Security*, 2013.
- [11] S. Rachev and L. Ruschendorf, *Mass transportation problems*. Springer, 1998, vol. 1.
- [12] Y. Deng and W. Du, "The Kantorovich Metric in computer science: A brief survey," *Electron. Notes Theor. Comput. Sci.*, vol. 253, no. 3, pp. 73–82, Nov. 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.entcs.2009.10.006>
- [13] A. McIver, L. Meinicke, and C. Morgan, "A Kantorovich-monadic powerdomain for information hiding, with probability and nondeterminism," in *Proc. LiCS 2012*, 2012.
- [14] E. Lawler, *Combinatorial optimization: Networks and Matroids*. Holt, Rinehart and Winston, 1976.
- [15] N. Fernandes, A. McIver, and C. Morgan, "The Laplace Mechanism has optimal utility for differential privacy over continuous queries," April 2021, full version of this paper with appendices. [Online]. Available at <http://www.cse.unsw.edu.au/~carrollm/LiCS2021-210420.pdf>
- [16] P. Meyer-Nieberg, *Banach Lattices*. Springer-Verlag, 1991.
- [17] E. Wilson, "First and second laws of error," *JASA*, vol. 18, no. 143, 1923.
- [18] M. M. Pai and A. Roth, "Privacy and mechanism design," *SIGecom Exch.*, vol. 12, no. 1, pp. 8–29, 2013. [Online]. Available: <https://doi.org/10.1145/2509013.2509016>
- [19] I. Dinur and K. Nissim, "Revealing information while preserving privacy," in *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 9-12, 2003, San Diego, CA, USA*, F. Neven, C. Beeri, and T. Milo, Eds. ACM, 2003, pp. 202–210. [Online]. Available: <https://doi.org/10.1145/773153.773173>
- [20] 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.
- [21] J. Soria-Comas and J. Domingo-Ferrer, "Optimal data-independent noise for differential privacy," *Information Sciences*, vol. 250, pp. 200–214, 2012.
- [22] 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, 2015.
- [23] M. Gupte and M. Sundararajan, "Universally optimal privacy mechanisms for minimax agents," in *Proc. Symp. Principles of Database Systems*, ser. PODS '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 135–146. [Online]. Available: <https://doi.org/10.1145/1807085.1807105>
- [24] F. Koufogiannis, S. Han, and G. J. Pappas, "Optimality of the laplace mechanism in differential privacy," *arXiv preprint arXiv:1504.00065*, 2015.
- [25] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud, "Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems," in *53rd IEEE conference on decision and control*. IEEE, 2014, pp. 2130–2135.
- [26] H. Brenner and K. Nissim, "Impossibility of differentially private universally optimal mechanisms," in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. Los Alamitos, CA, USA: IEEE Computer Society, oct 2010, pp. 71–80. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/FOCS.2010.13>
- [27] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, P. Degano, and C. Palamidessi, "On the information leakage of differentially-private mechanisms," *Journal of Computer Security*, vol. 23, no. 4, pp. 427–469, 2015. [Online]. Available: <https://doi.org/10.3233/JCS-150528>
- [28] H. Asi and J. C. Duchi, "Near instance-optimality in differential privacy," 2020, [arXiv:2005.10630v1](https://arxiv.org/abs/2005.10630v1), 2020.
- [29] Y. Wang, X. Wu, and L. Wu, "Differential privacy preserving spectral graph analysis," in *Advances in Knowledge Discovery and Data Mining*, J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 329–340.
- [30] N. Phan, X. Wu, H. Hu, and D. Dou, "Adaptive laplace mechanism: Differential privacy preservation in deep learning," in *2017 IEEE International Conference on Data Mining (ICDM)*, 2017, pp. 385–394.
- [31] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 351–360. [Online]. Available: <https://doi.org/10.1145/1536414.1536464>