



**HAL**  
open science

# Honest Fraction Differential Privacy

Imane Taibi, Jan Ramon

► **To cite this version:**

Imane Taibi, Jan Ramon. Honest Fraction Differential Privacy. ACM workshop on information hiding and multimedia security, ACM, Jun 2024, Vigo, Spain. pp.1-5, 10.1145/3658664.3659655 . hal-04610199

**HAL Id: hal-04610199**

**<https://inria.hal.science/hal-04610199v1>**

Submitted on 12 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Honest Fraction Differential Privacy

Imane Taibi and Jan Ramon

Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 - CRIStAL  
Lille, France

June 12, 2024

## Abstract

Over the last decades, differential privacy (DP) has become a standard notion of privacy. It allows to measure how much sensitive information an adversary could infer from a result (statistical model, prediction, etc.) he obtains. In privacy-preserving federated machine learning, one aims to learn a statistical model from data owned by multiple data owners without revealing their sensitive data. A common strategy is to use secure multi-party computation (SMPC) to avoid revealing intermediate results. However, DP assumes a very strong adversary who is able to know all information in the dataset except the targeted secret, while most SMPC methods assume a clearly less strong adversary, e.g., it is common to assume that the adversary has bounded computational power and can corrupt only a minority of the data owners (honest majority). As a chain is not stronger than its weakest part, in such combinations the DP provides an overly strong protection at an unnecessarily high cost in terms of utility. We propose honest fraction differential privacy, which is similar to differential privacy but assumes that the adversary can only collude with data owners covering part of the data. This assumption is very similar to the assumptions made by many SMPC strategies. We illustrate this idea by considering the application to the specific task of unregularized linear regression without bias on sufficiently large datasets.

## 1 Introduction

Recently, differential Privacy (DP), has become a standard model for measuring privacy in data analysis. It allows for estimating the extent to which an adversary can figure out private information from data analysis outputs like statistical models [5]. This approach is essential in fields like healthcare and finance, where it is crucial to handle sensitive data carefully. A common technique is to add some noise to the final result of a computation before publishing it to avoid that a recipient of the result can infer any sensitive data from it.

Federated machine learning [14] involves training algorithms across decentralized devices or servers holding local data samples. In privacy-preserving

federated machine learning, the goal is to learn from datasets distributed across multiple data owners without compromising the confidentiality of the individual datasets [11].

A common strategy in this domain has been the utilization of Secure Multi-Party Computation (SMPC). SMPC [8] aims to compute results based on inputs from multiple parties without revealing any of the input data or intermediate results in collaborative computations, even if the central server is not trusted. Among others, it can readily be used for securely aggregating values contributed by the participating data owners or for generating DP noise in a verifiable way [18].

Standard DP can be seen as assuming an adversary with extensive knowledge, capable of accessing all but one instance in the dataset. This assumption contrasts with the more restrained adversary model in SMPC, where the adversary is often thought to have limited computational power and only able influence at most a certain fraction of the data owners (or at most a subset of data owners corresponding to a certain fraction of the data), as explored by Lindell [13]. The result is an over-protection by DP, leading to a decrease in the utility of the data, a phenomenon widely discussed in the literature [20].

As it is unnecessary in view of an adversary for which one already assumes limited computational power to require information-theoretic differential privacy, the notion of computational differential privacy was introduced [16]. In the computational DP setting, an adversary may have encrypted sensitive information as long as decrypting it is beyond its computational capabilities, e.g., if decryption is infeasible in polynomial time.

Similarly, if the SMPC component of an algorithm already assumes that a certain fraction of the parties are honest, it doesn't help much to assume the adversary has access to all but one instance when considering statistical privacy. To address this issue, we introduce a novel concept termed Honest Fraction Differential Privacy (HFDP). HFDP is similar to traditional DP, but reflects privacy assuming that an adversary can only collude with a certain fraction of the data owners, aligned with many SMPC methods.

We apply HFDP specifically to linear regression models. Linear regression is chosen due to its wide use in various sectors, especially in medical data analysis.

In summary, our contributions are:

- We introduce the novel concept of Honest Fraction Differential Privacy (HFDP), and study its properties.
- We apply the HFDP framework to linear regression and propose a strategy to train  $\gamma$ -HF  $(\epsilon, \delta)$ -DP linear regression models.

The remainder of this paper is structured as follows: first, in Section 2 we review some basic definitions and relevant literature. Next, in Section 3 we present the key ideas of the Honest Fraction Differential Privacy framework, and in Section 4 we apply this to the specific case of simple linear regression. In Section 5 we then provide further discussion, conclusions and directions for future work.

## 2 Background

**Differential Privacy (DP)** We will denote by  $\mathcal{X}$  the space of all possible instances. For a space  $S$ , we will denote by  $\mathcal{P}(S)$  the space of all probability distributions over  $S$ , and we will denote by  $S^*$  the space of all sets of elements of  $S$ . We will denote by  $I_d \in \mathbb{R}^{d \times d}$  the  $d$ -dimensional identity matrix.

The foundational concept of Differential Privacy (DP) was established by Dwork [5, 6], among others with the study of  $(\epsilon, \delta)$ -differential privacy.

**Definition 1** (Adjacent datasets). *We say two datasets  $D, D' \in \mathcal{X}^*$  are adjacent if they differ in at most one instance, i.e., there are  $D_o \in \mathcal{X}^*$  and  $x, x' \in \mathcal{X}$  such that  $D = D_o \cup \{x\}$  and  $D' = D_o \cup \{x'\}$ . We denote by  $\text{adj}_{\mathcal{X}}$  the set of all adjacent datasets in  $\mathcal{X}^*$ .*

**Definition 2** (Mechanism). *For an input space  $\mathcal{X}$  and an output space  $\mathcal{Y}$ , a mechanism  $A : \mathcal{X} \rightarrow \mathcal{Y}$  is a randomized algorithm mapping elements of  $\mathcal{X}$  on elements of  $\mathcal{Y}$ . We denote the space of all mechanisms from  $\mathcal{X}$  to  $\mathcal{Y}$  by  $\mathcal{M}(\mathcal{X}, \mathcal{Y})$ .*

Mechanisms can be (or can be post-processing steps to) a wide variety of algorithms, such as randomized algorithms for learning, optimizing, predicting, inferencing or encoding. A mechanism is differentially private if changing a single instance in the input dataset doesn't lead to a distinguishably different output.

**Definition 3** ( $(\epsilon, \delta)$ -Differential Privacy). *Let  $\epsilon > 0$  and  $\delta \geq 0$ . A mechanism  $A \in \mathcal{M}(\mathcal{X}^*, \mathcal{Y})$  is  $(\epsilon, \delta)$ -differentially private if, for any two adjacent datasets  $D \in \mathcal{X}^*$  and  $D' \in \mathcal{X}^*$ , and for all  $Y \subseteq \mathcal{Y}$ , the inequality  $P(A(D) \in Y) \leq e^\epsilon P(A(D') \in Y) + \delta$  holds.*

To achieve differential privacy (DP), one of the most well-known tools are the Laplace mechanism and the Gaussian mechanism [5, 15]. The Gaussian mechanism adds a Gaussian distributed noise to the output of a function. For functions  $f$  with range  $\mathbb{R}^d$ , we denote by  $M_{G:\sigma^2}[f]$  the mechanism mapping  $x$  on  $f(x) + \eta$  with  $\eta \sim \mathcal{N}(0, \sigma^2 I_d)$ .

**Definition 4** (L2 Sensitivity). *The  $L_2$  sensitivity of a function  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , denoted as  $\Delta_2 f$ , is defined as the maximum Euclidean distance between the outputs of  $f$  on any two adjacent datasets, i.e.,*

$$\Delta_2 f = \max_{(D, D') \in \text{adj}_{\mathcal{X}}} \|f(D) - f(D')\|_2.$$

**Theorem 1** (Gaussian Mechanism (Dwork 2014)). *For any function  $f : \mathcal{X} \rightarrow \mathcal{Y}$  and any  $\epsilon > 0$  and  $\delta > 0$ , the Gaussian mechanism  $M_{G:\sigma^2}[f]$  satisfies  $(\epsilon, \delta)$ -differential privacy if*

$$\sigma^2 \geq 2 \log \left( \frac{1.25}{\delta} \right) \frac{(\Delta_2 f)^2}{\epsilon^2}.$$

**Federated Learning** Federated Learning (FL) is a decentralized approach to machine learning where the training instances are distributed across multiple data owners. Depending on the extent to which parties trust each other, one can assume several security models. While parties can misbehave in many ways, in this paper we focus on parties interested in learning about the sensitive data of other parties. To avoid parties learn each other’s data while computing statistics one can use secure aggregation or more generally secure multi-party computation. While it is possible to design algorithms that are robust against all but one party colluding to reveal the secrets of the last party, such protocols are often expensive. e.g., additive secret sharing is robust against such collusion but the cost is quadratic in the number of parties. It is therefore more common to assume a fraction of the parties are honest, i.e., they follow the protocol and don’t collude, e.g., [17] then has a cost  $O(p \log(p))$  with  $p$  the number of parties. A popular security model is the honest majority model [11, 2] where one assumes that the majority of parties are honest.

**DP in Linear Regression Models** Exploring the application of Differential Privacy (DP) within linear regression models offers a concrete example of how DP principles can be implemented in specific statistical analyses. Zhang et al. [22] introduce the Functional Mechanism, a method that adapts the objective function, ensuring differential privacy for both linear and logistic regression models. Furthering this line of inquiry, Dandekar et al. [4] examine the application of a DP functional mechanism in regularized linear regression models.

**Utility-privacy trade-off** The integration of privacy-preserving techniques in Federated Learning (FL) is essential for safeguarding data across decentralized clients, but it can also impact the utility of the aggregated model. Several research papers have explored this topic, emphasizing the importance of finding a balance between privacy and utility [10, 3, 23, 24]. Zhang et al. [23] delve into the trade-offs between privacy and utility.

**Exploiting external randomness** Some research has sought to exploit intrinsic randomness existing already in some algorithms to reduce the amount of additional noise needed to achieve a certain privacy level. For example, some cryptographic strategies such as fully homomorphic encryption contain randomization which can be exploited as differential privacy noise [19]. Also, some machine learning algorithms such as Stochastic Gradient Descent allow for using their sampling randomness to be exploited to improve privacy [9, 21]. The work described in the current paper can be combined with such strategies exploiting intrinsic randomness from algorithms, as the sources of additional uncertainty for the adversary which these methods and our work exploit are complementary and independent.

**Other privacy notions** Next to classic differential privacy, several other privacy notions have been proposed in literature. Some of these are based on

alternative ways to measure the divergence between the distributions of outputs of a mechanism when provided with two adjacent datasets as input. For example, Rényi Differential Privacy [15] generalizes the notion of  $\epsilon$ -differential privacy by considering Rényi divergence between probability distributions, parameterized by an order parameter  $\alpha$ . Concentrated Differential Privacy [7] is another variant, focusing on bounding the divergence from the expected privacy loss rather than the worst-case scenario. We describe our work in the context of  $(\epsilon, \delta)$ -differential privacy for the simplicity of explanation in this short paper, but the same idea can be also applied to Rényi DP and Concentrated DP.

An interesting general framework is the Pufferfish framework [12] of which all the above privacy notions are instantiations. Our work too can be seen as an instance of the Pufferfish model. Even though the generality decreases the operationality of the framework, [12] shows that a restricted set of properties hold in general for all privacy notions fitting the framework.

### 3 Honest fraction differential privacy

We now define our new concept of Honest Fraction Differential Privacy. The 'honest fraction' refers to the fraction of the parties in federated learning which are assumed to be honest, i.e., not colluding to infer some information. In practice, results in this paper which are secure under a  $\gamma$ -honest fraction assumption remain valid if there are multiple disjoint groups of colluding parties, as long as no such group of colluding parties exceeds a fraction of  $1 - \gamma$  of the parties.

Intuitively, the dataset  $D$  is partitioned as  $D = D_o \cup D_u \cup \{(x_T, y_T)\}$  with  $D_o$  containing instances the adversary can observe (e.g., by colluding with the concerned data owners),  $D_u$  containing instances which are unknown to the adversary, and  $(x_T, y_T)$  the target instance which the adversary can't see but would like to infer more information about. We will sometimes group  $D_o$  and  $(x_T, y_T)$ , setting  $D_- = D_o \cup \{(x_T, y_T)\}$ .

**Definition 5.** Let  $\pi \in \mathcal{P}(\mathcal{X})$  be a probability distribution. We define  $Ext_\pi : \mathbb{N} \times \mathcal{X}^* \rightarrow \mathcal{X}^*$  to be a randomized function which takes as input a number  $n_u$  and a dataset  $D_- \subseteq \mathcal{X}^*$  and outputs a dataset  $D = D_- \cup D_u$  where  $D_u$  contains  $n_u$  instances sampled i.i.d. from  $\pi$ .

**Definition 6** ( $\gamma$ -Honest Fraction  $(\epsilon, \delta)$ -DP). a mechanism  $A \in \mathcal{M}(\mathcal{X}^*, \mathcal{Y})$  is  $\gamma$ -HF  $(\epsilon, \delta)$ -DP if, for any two adjacent datasets  $D_-$  and  $D'_-$ , and for all output sets  $Y \subseteq \mathcal{Y}$ , the inequality

$$P(A(Ext_\pi(n_u, D_-)) \in Y) \leq e^\epsilon P(A(Ext_\pi(n_u, D'_-)) \in Y) + \delta$$

holds for  $n_u = \left\lfloor \frac{\gamma}{1-\gamma} (|D_-| - 1) \right\rfloor - 1$ .

This definition essentially says that the adversary only can see some fraction  $1 - \gamma$  of the dataset, denoted by  $D_o$ , and has no knowledge of the remaining fraction  $\gamma$  of the dataset (except for the output of the algorithm  $A$ ), consisting

of the target instance  $(x_T, y_T)$  and a set  $D_u$  of  $n\gamma - 1$  other unseen instances, except that it is drawn from the same distribution. This means that if an adversary wants to infer something about an instance  $x$ , he does not only take into account the noise added by the mechanism, but also the uncertainty arising from his ignorance of  $D_u$ .

In general the HFDP privacy notion is not composable in the sense that a mechanism  $A$  outputting a pair containing the outputs of a  $\gamma$ -HF  $(\epsilon_1, \delta_1)$ -DP mechanism  $A_1$  and a  $\gamma$ -HF  $(\epsilon_2, \delta_2)$ -DP mechanism  $A_2$ , is itself not necessarily  $\gamma$ -HF  $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ -DP. The reason is that the uncertainty added to the adversary's inference problem by his ignorance of the data  $D_u$  is the same for both mechanisms  $A_1$  and  $A_2$ , as the data owners not colluding with the adversary don't use different data for the evaluations of the mechanisms  $A_1$  and  $A_2$ . The noise added by  $A_1$  and  $A_2$  therefore correlates and simple composition doesn't work. Nevertheless, as we will show in Section 4, HFDP can be used to reduce the amount of noise one must add to a statistical model with multiple parameters to achieve privacy.

## 4 Application to linear regression

In this section, we explore the application of HFDP within the framework of linear regression.

Let  $\mathcal{X} = \mathcal{B}_d = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 1\}$  be the unit ball in  $d$  dimensions,  $\mathcal{Y} = [-1, 1]$  and  $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ . Let  $D \in \mathcal{Z}^n$  be a data set containing  $n$  instances  $z_i$ ,  $i = 1 \dots n$ , where  $z_i = (x_i, y_i)$ . We assume that the adversary knows at most  $\lfloor (1 - \gamma)n \rfloor$  instances. To keep our formulae simple, in the remainder of this work we assume  $\gamma n$  is an integer.

**Definition 7** (Linear regression). *The simple linear regression on  $D$  is the model  $\text{LR}_D : \mathcal{X} \rightarrow \mathcal{Y}$  with  $\text{LR}_D(x) = x^\top \omega^*$ , where  $\omega^* \in \mathbb{R}^d$  minimizes the objective function*

$$F_D(\omega) = \sum_{i=1}^n (y_i - x_i^\top \omega)^2$$

*i.e.*,  $\omega^* = \arg \min_{\omega} F_D(\omega)$ .

In other words, linear regression expresses the value of the target value  $y$  as a linear function of the input values of  $x$ , minimizing the sum of squared errors on the training set. While we consider here for simplicity of our explanation this simple regression variant, our derivation also applies to regularized linear regression. Similarly, it is straightforward to add a bias term to the regression.

**Functional mechanism** In the context of differential privacy, the functional mechanism refers to a class of techniques that provide differential privacy guarantees by adding noise to the objective function that machine learning algorithms try to optimize. To apply the functional mechanism to linear regression,

the first step is to write  $F_D(\omega)$  as a polynomial in  $\omega$ . In particular,

$$F_D(\omega) = \omega^T A \omega + B \omega + C$$

where  $A = \sum_{i=1}^n x_i x_i^T$ ,  $B = -2 \sum_{i=1}^n y_i x_i^T$  and  $C = \sum_{i=1}^n y_i^2$ .

Next, one can add Gaussian noise to the coefficients of the polynomial. Let's denote the noise variables by  $\eta^{(A)} \in \mathbb{R}^{d \times d}$ ,  $\eta^{(B)} \in \mathbb{R}^d$  and  $\eta^{(C)} \in \mathbb{R}$  respectively. The objective function after noise injection is:

$$\hat{f}_D(\omega) = \omega^T (A + \eta^{(A)}) \omega + (B + \eta^{(B)}) \omega + (C + \eta^{(C)})$$

Inspired on Zhang et al. [22], we observe that for all  $i \neq j$ ,  $\Delta_2 A_{i,i} = \Delta_2 C = 1$  and  $\Delta_2 A_{i,j} = \Delta_2 B_i = 2$ . The matrix  $A$  is symmetric, so denoting the upper triangular part of  $A$  by  $A^U$ , the vector  $(A^U, B, C)$  has in total

$$Q = \frac{d(d+1)}{2} + d + 1 = \frac{d^2 + 3d + 2}{2}$$

components, of which  $Q_1 = d + 1$  in the range  $[0, 1]$  have sensitivity 1 and  $Q_2 = (d^2 + d)/2$  in the range  $[-1, 1]$  have sensitivity 2. Hence  $\Delta_2(A^U, B, C) = \sqrt{Q_1 + 4Q_2} = \sqrt{2d^2 + 3d + 1}$ . We can make  $(A^U, B, C)$  differentially private by adding the Gaussian mechanism with variance

$$\sigma_{\text{LR}}^2(\epsilon, \delta) = 2(\Delta_2(A^U, B, C))^2 \log(1.25/\delta)/\epsilon^2$$

Let us now analyze what an adversary can see. An adversary seeing the model (the vector  $\omega^*$ ) may infer something about the vector  $(A^U, B, C)$ . They may also know part of the data set  $D_o$  with  $|D_o| = (1 - \gamma)n$ . For the adversary, the data set  $D$  consists of

$$D = D_o \cup D_u \cup \{(x_T, y_T)\}$$

where  $D_u$  contains  $\gamma n - 1$  instances they can't see and where  $(x_T, y_T)$  is the target instance they would like to infer information about. He knows that  $\Theta = \theta_o + \theta_u + \theta(x_T, y_T) + \eta^{(ABC)}$  with  $\Theta = (A, B, C)$ ,  $\theta(x, y) = (xx^T, xy, y^2)$ ,  $\eta^{(ABC)} = (\eta^{(A)}, \eta^{(B)}, \eta^{(C)})$ ,

$$\theta_o = (A_o, B_o, C_o) = \sum_{(x,y) \in D_o} \theta(x, y)$$

and

$$\theta_u = (A_u, B_u, C_u) = \sum_{(x,y) \in D_u} \theta(x, y).$$

As the adversary may know  $\Theta$  and  $\theta_o$ , the noise prohibiting the adversary from inferring  $(x_T, y_T)$  is the uncertainty on  $\theta_u + \eta^{(ABC)}$ . The variable  $\theta_u$  is a sum of functions of  $|D_u| = \gamma n - 1$  instances, if  $\gamma n$  is sufficiently large,  $\theta_u$  will follow a nearly Gaussian distribution. According to Theorem 1 we can achieve privacy



by ensuring there is Gaussian noise with variance  $\sigma_{LR}^2$ , so the uncertainty on each component of  $\theta_u + \eta^{(ABC)}$  should be at least  $\sigma_{LR}^2$ .

The adversary may estimate  $\theta_u$  as  $\hat{\theta}_u^a = \theta_o |D_u| / |D_o|$ , knowing that both parts of the dataset are drawn from the same distribution. In general, it may not be easy to assess how accurate such estimate  $\hat{\theta}_u^a$  is. Special cases have been described in literature, e.g., if the population distribution is Gaussian, then  $\hat{\theta}_u^a - \mathbb{E}[\theta_u]$  follows a Wishart distribution. Anyway, the adversary may have prior background information and may have a better estimate  $\hat{\theta}_u^a$ . We therefore will assume that  $\hat{\theta}_u^a$  is a good approximation to  $\mathbb{E}[\theta_u]$ .

From the point of view of the party building the DP regression model, to estimate the uncertainty the adversary will have on  $\hat{\theta}_u^a - \theta_u$ , one could compute  $\hat{\theta} = \Theta/n$  and

$$\Sigma = \frac{1}{n-1} \sum_{(x,y) \in D} (\hat{\theta} - \theta_i)(\hat{\theta} - \theta_i)^\top$$

As  $n$  increases,  $\Sigma$  becomes a better estimate for the uncertainty of the adversary on  $\hat{\theta}_u^a - \theta_u$ . In particular, for any  $\iota > 0$ , there is an increasing function  $h : \mathbb{N} \rightarrow [0, 1]$  with  $\lim_{n \rightarrow \infty} h(n) = 1$  such that with probability at least  $1 - \iota$  there holds

$$\mathbb{E}[(\hat{\theta}_u^a - \theta_u)(\hat{\theta}_u^a - \theta_u)^\top] \succeq \Sigma |D_u| h(n)$$

where  $A \succeq B$  means that  $A - B$  is positive semidefinite.

If  $n$  is also sufficiently large to make  $\theta_u$  sufficiently close to a Gaussian distributed random variable (from the point of view of the adversary, even if they would have good knowledge of the population distribution and hence potentially of  $\Sigma$ ), then to achieve  $\gamma$ -HF  $(\epsilon, \delta' + \iota)$ -DP it is sufficient to add to  $\Theta$  noise  $\eta^{(ABC)}$  with  $\eta^{(ABC)} \sim \mathcal{N}(0, \Sigma_\eta)$  such that

$$|D_u| \Sigma h(n) + \Sigma_\eta \succeq \sigma_{LR}^2(\epsilon, \delta') I.$$

This leads us to our main theorem:

**Theorem 2.** *Let  $\gamma \in (0, 1)$ ,  $\epsilon > 0$ ,  $\delta > \delta' > 0$ . There exists a function  $h : \mathbb{N} \rightarrow [0, 1]$  with  $\lim_{n \rightarrow \infty} h(n) = 1$  such that for any  $D \in \mathcal{Z}^n$ , adding Gaussian noise to the  $\Theta = (A^U, B, C)$  vector (before optimizing for  $\omega$ ) where to  $\Theta$  we add  $\eta^{(ABC)}$  with  $\eta^{(ABC)} \sim \mathcal{N}(0, \Sigma_u)$  satisfying  $|D_u| \Sigma h(n) + \Sigma_\eta \succeq \sigma_{LR}^2(\epsilon, \delta') I$ , gives a  $\gamma$ -HF  $(\epsilon, \delta)$ -DP mechanism.*

Here,  $h(n)$  models the extent to which finite values of  $n$  may let the adversary's knowledge of the sum over  $D_u$  deviate from a Gaussian distribution. Deriving good bounds for  $h(n)$  is out of the scope of this short paper.

It is important to observe that for the computation of  $\Sigma$  one only needs to evaluate averages over the instances of the dataset  $D$ . Therefore, in the many federated learning settings where a secure aggregation operation is already available (for example, to compute the coefficients  $A$ ,  $B$  and  $C$ ), the implementation of our technique should be straightforward.

It is possible that the total uncertainty for the adversary, i.e., on  $\theta_u + \eta^{(ABC)}$  is not isotropic, e.g., if no  $\Sigma_\eta$  would exist which makes exactly  $|D_u| \Sigma h(n) + \Sigma_\eta =$

$\sigma_{LR}^2(\epsilon, \delta')I$ . Still, in that case the probability distribution on  $\theta_u - \theta_u^a + \eta^{(ABC)}$  could be written as a sum of the isotropic distribution  $\mathcal{N}(0, \sigma_{LR}^2)$  needed to ensure privacy and some remainder noise in the dimensions where there is high variance in the distribution on  $\theta_u$ .

## 5 Conclusion

In this short paper, we presented a novel idea for guaranteeing privacy while reducing the amount of noise that needs to be added. In particular, we argue that in the context of federated learning where anyway cryptography-based components make stronger security assumptions, e.g., that the adversary is computationally limited or can corrupt only part of the data owners, it makes no sense to adopt a statistical privacy notion assuming a strong adversary having access to all but one instance. We analyzed for the specific case of simple linear regression how assuming that a fraction  $\gamma$  of the data is secured by honest data owners can reduce the noise one needs to add using the functional mechanism. Similar effects can be obtained when applying other privacy mechanisms, e.g., when using the Gaussian mechanism to privatize node information in decision trees or to privatize gradients in DP-SGD [1] to train neural network models.

If all data owners have the same number of instances, our notion of having a fraction  $\gamma$  of the data with honest data owners implies that a fraction  $\gamma$  of the data owners should be honest, which nicely aligns with the typical assumptions of cryptographic algorithms. In case data owners have different numbers of instances, it is possible that more data owners (having fewer instances each) can collude and fewer bigger data owners may collude.

We want to pursue several lines of further work. First, we want to perform systematic experiments on real-world data to better study the effect of changing the security assumptions on the utility of predictive models. Second, we aim to elaborate the several extensions outlined in the body of this paper, e.g., the option to consider HFDP in information-theoretic or computational form depending on the security assumptions, stating formulas for regression with bias terms and the generalizations to regularized linear regression and other privacy mechanisms. Third, we aim to more generally provide techniques to exploit the more limited knowledge of the adversary, even if the fact that the data unknown to the adversary induces correlated noise which is somewhat more difficult to handle in composition of privacy costs. Finally, we want more broadly to investigate whether there are other aspects of real-world scenarios we can exploit to better model potential privacy attacks and better tune the amounts of needed noise to achieve privacy, further improving the achievable utility given a desired level of privacy.

## Acknowledgements

This work was partially supported by the Horizon Europe TRUMPET project grant no 101070038 and the ANR PMR project grant no ANR-20-CE23-0013.

## References

- [1] Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep Learning with Differential Privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318, New York, NY, USA, October 2016. Association for Computing Machinery. arXiv: 1607.00133.
- [2] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS '17*, page 1175–1191, New York, NY, USA, 2017. Association for Computing Machinery.
- [3] Hanxiao Chen, Hongwei Li, Guowen Xu, Yun Zhang, and Xizhao Luo. Achieving privacy-preserving federated learning with irrelevant updates over e-health applications. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pages 1–6, ., 2020. IEEE.
- [4] Ashish Dandekar, Debabrota Basu, and Stéphane Bressan. Differential privacy for regularised linear regression. In Sven Hartmann, Hui Ma, Abdelkader Hameurlain, Günther Pernul, and Roland R. Wagner, editors, *Database and Expert Systems Applications*, pages 483–491, Cham, 2018. Springer International Publishing.
- [5] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [6] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2014.
- [7] Cynthia Dwork and Guy N. Rothblum. Concentrated differential privacy, 2016.
- [8] David Evans, Vladimir Kolesnikov, and Mike Rosulek. A Pragmatic Introduction to Secure Multi-Party Computation. *Foundations and Trends® in Privacy and Security*, 2(2-3):70–246, December 2018. Publisher: Now Publishers, Inc.
- [9] Stephanie L. Hyland and Shruti Tople. An empirical study on the intrinsic privacy of sgd, 2022.
- [10] Yupeng Jiang, Yong Li, Yipeng Zhou, and Xi Zheng. Mitigating sybil attacks on differential privacy based federated learning, 2020.

- [11] Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D’Oliveira, Hubert Eichner, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaid Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrede Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Mariana Raykova, Hang Qi, Daniel Ramage, Ramesh Raskar, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao. Advances and Open Problems in Federated Learning. *arXiv e-prints*, 14:1–210, 2019.
- [12] Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems*, 39(1):3:1–3:36, 2014.
- [13] Yehuda Lindell and Benny Pinkas. Secure multiparty computation for privacy-preserving data mining. *Journal of Privacy and Confidentiality*, 1(1), Apr. 2009.
- [14] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282, Seattle, WA 98103 USA, 20–22 Apr 2017. PMLR.
- [15] Ilya Mironov. Rényi differential privacy. *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, abs/1702.07476:263–275, 2017.
- [16] Ilya Mironov, Omkant Pandey, Omer Reingold, and Salil Vadhan. Computational Differential Privacy. In Shai Halevi, editor, *Advances in Cryptology - CRYPTO 2009*, Lecture Notes in Computer Science, pages 126–142, Berlin, Heidelberg, 2009. Springer.
- [17] César Sabater, Aurélien Bellet, and Jan Ramon. An Accurate, Scalable and Verifiable Protocol for Federated Differentially Private Averaging. *Machine learning Journal*, 111:4249–4293, 2022.
- [18] César Sabater, Florian Hahn, Andreas Peter, and Jan Ramon. Private sampling with identifiable cheaters. *Proc. Priv. Enhancing Technol.*, 2023(2):361–383, 2023.
- [19] Arnaud Grivet Sébert, Martin Zuber, Oana Stan, Renaud Sirdey, and Cédric Gouy-Pailler. When approximate design for fast homomorphic computation provides differential privacy guarantees, 2023.

- [20] Jun Wang, Shubo Liu, and Yongkai Li. A review of differential privacy in individual data release. *International Journal of Distributed Sensor Networks*, 11(10):259682, 2015.
- [21] Wenqi Wei, Ling Liu, Jingya Zhou, Ka-Ho Chow, and Yanzhao Wu. Securing distributed sgd against gradient leakage threats, 2023.
- [22] Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Y. Yang, and Marianne Winslett. Functional mechanism: Regression analysis under differential privacy. *Proc. VLDB Endow.*, 5:1364–1375, 2012.
- [23] Xiaojin Zhang, Hanlin Gu, Lixin Fan, Kai Chen, and Qiang Yang. No free lunch theorem for security and utility in federated learning. *ACM Transactions on Intelligent Systems and Technology*, 14(1):1–35, 2022.
- [24] Jianzhe Zhao, Keming Mao, Chenxi Huang, and Yuyang Zeng. Utility optimization of federated learning with differential privacy. *Discrete Dynamics in Nature and Society*, 2021:1–14, 2021.