Received: 25 May 2019 | Accepted: 1 April 2020

DOI: 10.1111/rssb.12454

ORIGINAL ARTICLE

Discussion Paper

Gaussian differential privacy

Jinshuo Dong[1] | **Aaron Roth**[2] | **Weijie J. Su**[3]

[1]Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[2]Department of Computer Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Wharton Statistics and Data Science Department, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Weijie Su, Department of Statistics, Wharton School, University of Pennsylvania, USA.
Email: suw@wharton.upenn.edu

Funding information
NSF, Grant/Award Number: CAREER DMS-1847415, CCF-1763314 and CNS-1253345

Abstract

In the past decade, differential privacy has seen remarkable success as a rigorous and practical formalization of data privacy. This privacy definition and its divergence based relaxations, however, have several acknowledged weaknesses, either in handling composition of private algorithms or in analysing important primitives like privacy amplification by subsampling. Inspired by the hypothesis testing formulation of privacy, this paper proposes a new relaxation of differential privacy, which we term 'f-differential privacy' (f-DP). This notion of privacy has a number of appealing properties and, in particular, avoids difficulties associated with divergence based relaxations. First, f-DP faithfully preserves the hypothesis testing interpretation of differential privacy, thereby making the privacy guarantees easily interpretable. In addition, f-DP allows for lossless reasoning about composition in an algebraic fashion. Moreover, we provide a powerful technique to import existing results proven for the original differential privacy definition to f-DP and, as an application of this technique, obtain a simple and easy-to-interpret theorem of privacy amplification by subsampling for f-DP. In addition to the above findings, we introduce a canonical single-parameter family of privacy notions within the f-DP class that is referred to as 'Gaussian differential privacy' (GDP), defined based on hypothesis testing of two shifted Gaussian distributions. GDP is the focal privacy definition among the family of f-DP guarantees due to a central limit theorem for differential privacy that we

[Read before The Royal Statistical Society at an online meeting organized by the Research Section on Wednesday, December 16th, 2020, Professor G. P. Nason in the Chair]

prove. More precisely, the privacy guarantees of *any* hypothesis testing based definition of privacy (including the original differential privacy definition) converges to GDP in the limit under composition. We also prove a Berry–Esseen style version of the central limit theorem, which gives a computationally inexpensive tool for tractably analysing the exact composition of private algorithms. Taken together, this collection of attractive properties render *f*-DP a mathematically coherent, analytically tractable and versatile framework for private data analysis. Finally, we demonstrate the use of the tools we develop by giving an improved analysis of the privacy guarantees of noisy stochastic gradient descent.

KEYWORDS

Blackwell theorem, central limit theorem, composition, differential privacy, primal-dual perspective, privacy amplification, private stochastic gradient descent, subsampling, trade-off function

1 | INTRODUCTION

Modern statistical analysis and machine learning are overwhelmingly applied to data concerning *people*. Valuable data sets generated from personal devices and online behaviour of billions of individuals contain data on location, web search histories, media consumption, physical activity, social networks and more. This is on top of continuing large-scale analysis of traditionally sensitive data records, including those collected by hospitals, schools and the Census. This reality requires the development of tools to perform large-scale data analysis in a way that still protects the *privacy* of individuals represented in the data.

Unfortunately, the history of data privacy for many years consisted of ad hoc attempts at 'anonymizing' personal information, followed by high profile de-anonymizations. This includes the release of AOL search logs, de-anonymized by the *New York Times* (Barbaro & Zeller, 2006), the Netflix Challenge data set, de-anonymized by Narayanan and Shmatikov (2008), the realization that participants in genome-wide association studies could be identified from aggregate statistics such as minor allele frequencies that were publicly released (Homer et al., 2008), and the reconstruction of individual-level census records from aggregate statistical releases (Abowd, 2018).

Thus, we urgently needed a rigorous and principled privacy-preserving framework to prevent breaches of personal information in data analysis. In this context, *differential privacy* has put private data analysis on firm theoretical foundations (Dwork et al., 2006a,b). This definition has become tremendously successful; in addition to an enormous and growing academic literature, it has been adopted as a key privacy technology by Google (Erlingsson et al., 2014), Apple (Apple, 2017), Microsoft (Ding et al., 2017) and the US Census Bureau (Abowd, 2018). The definition of this concept involves privacy parameters $\varepsilon \geq 0$ and $0 \leq \delta \leq 1$.

Definition 1 (Dwork et al., 2006a, b). A randomized algorithm M that takes as input a data set consisting of individuals is (ε, δ)-differentially private (DP) if for any pair of data sets S, S' that differ in the record of a single individual, and any event E,

$$\mathbb{P}\left[M(S) \in E\right] \leq \mathrm{e}^{\varepsilon}\, \mathbb{P}\left[M(S') \in E\right] + \delta. \tag{1}$$

When $\delta = 0$, the guarantee is simply called ε-DP.

In this definition, data sets are *fixed* and the probabilities are taken *only* over the randomness of the mechanism. In particular, the event E can take any measurable set in the range of M. To achieve differential privacy, a mechanism is necessarily randomized. For example, consider the problem of privately releasing the average cholesterol level of individuals in the data set $S = (x_1, \ldots, x_n)$, where x_i corresponds to the cholesterol level of individual i. A privacy-preserving mechanism may take the form

$$M(S) = \frac{x_1 + \cdots + x_n}{n} + \text{noise}.$$

The level of the noise has to be large enough to mask the *characteristics* of any individual's cholesterol level, while *not* being too large to distort the population average for accuracy purposes. Consequently, the probability distributions of $M(S)$ and $M(S')$ are close to each other for any data sets S, S' that differ in only one individual record.

Differential privacy is most naturally defined through a hypothesis testing problem from the perspective of an attacker who aims to distinguish S from S' based on the output of the mechanism. This statistical viewpoint was first observed by Wasserman and Zhou (2010) and then further developed by Kairouz et al. (2017), which is a direct inspiration for our work. In short, consider the hypothesis testing problem

$$H_0\colon \text{ the underlying data set is } S \quad \text{versus} \quad H_1\colon \text{ the underlying data set is } S' \tag{2}$$

and call Alice the only individual that is in S but not S'. As such, rejecting the null hypothesis corresponds to the detection of the absence of Alice, whereas accepting the null hypothesis means to detect the presence of Alice in the data set. Using the output of an (ε, δ)-DP mechanism, the power of any test at significance level $0 < \alpha < 1$ has an upper bound of $\mathrm{e}^{\varepsilon}\alpha + \delta$. This bound is only slightly larger than α provided that ε, δ are small and, therefore, *any* test is essentially powerless. Put differently, differential privacy with small privacy parameters protects against any inferences of the presence of Alice, or any other individual, in the data set.

Despite its apparent success, there are good reasons to want to relax the original definition of differential privacy, which has led to a long line of proposals for such relaxations. The most important shortcoming is that (ε, δ)-DP does not tightly handle composition. Composition concerns how privacy guarantees degrade under repetition of mechanisms applied to the same data set, rendering the design of differentially private algorithms *modular*. Without compositional properties, it would be near impossible to develop complex differentially private data analysis methods. Although it has been known since the original papers defining differential privacy (Dwork et al., 2006a,b) that the composition of an (ε_1, δ_1)-DP mechanism and an (ε_2, δ_2)-DP mechanism yields an ($\varepsilon_1 + \varepsilon_2$, $\delta_1 + \delta_2$)-DP mechanism, the corresponding upper bound $\mathrm{e}^{\varepsilon_1+\varepsilon_2}\alpha + \delta_1 + \delta_2$ on the power of any test at significance level α no longer tightly characterizes the trade-off between significance level and power for the testing between S and S'. In Dwork et al. (2010), the authors gave an improved composition theorem, but it fails to capture the correct hypothesis testing trade-off. This is for a fundamental reason: (ε, δ)-DP is mis-parameterized in the sense that the guarantees of the composition of $(\varepsilon_i, \delta_i)$-DP mechanisms cannot be characterized by any single pair of parameters (ε, δ). Even worse, given any δ, finding the smallest parameter ε for composition of a sequence of differentially private algorithms is computationally hard (Murtagh & Vadhan, 2016), and so in practice, one must resort to approximations. Given that composition and modularity are first-order desiderata for a useful privacy definition, these are substantial drawbacks and often continue to push practical algorithms with meaningful privacy guarantees out of reach.

In the light of this, substantial recent effort has been devoted to developing relaxations of differential privacy for which composition can be handled exactly. This line of work includes several variants of 'concentrated differential privacy' (Bun & Steinke, 2016; Dwork & Rothblum, 2016), 'Rényi differential privacy' (Mironov, 2017) and 'truncated concentrated differential privacy' (Bun et al., 2018a). These definitions are tailored to be able to exactly and easily track the 'privacy cost' of compositions of the most basic primitive in differential privacy, which is the perturbation of a real valued statistic with Gaussian noise.

While this direction of privacy relaxation has been quite fruitful, there are still several places one might wish for improvement. First, these notions of differential privacy no longer have hypothesis testing interpretations, but are rather based on studying divergences that satisfy a certain information processing inequality. There are good reasons to prefer definitions based on hypothesis testing. Most immediately, hypothesis testing based definitions provide an easy way to interpret the guarantees of a privacy definition. More fundamentally, a theorem due to Blackwell (see Theorem 2) provides a formal sense in which a tight understanding of the trade-off between type I and type II errors for the hypothesis testing problem of distinguishing between $M(S)$ and $M(S^{'})$ contains only more information than any divergence between the distributions $M(S)$ and $M(S^{'})$ (so long as the divergence satisfies the information processing inequality).

Second, certain simple and fundamental primitives associated with differential privacy—most notably, *privacy amplification by subsampling* (Kasiviswanathan et al., 2011)—either fail to apply to the existing relaxations of differential privacy, or require a substantially complex analysis (Wang et al., 2018). This is especially problematic when analysing privacy guarantees of stochastic gradient descent—arguably the most popular present-day optimization algorithm—as subsampling is inherent to this algorithm. At best, this difficulty arising from using these relaxations could be overcome by using complex technical machinery. For example, it necessitated Abadi et al. (2016) to develop the numerical *moments accountant* method to sidestep the issue.

1.1 | Our contributions

In this work, we introduce a new relaxation of differential privacy that avoids these issues and has other attractive properties. Rather than giving a 'divergence' based relaxation of differential privacy, we start fresh from the hypothesis testing interpretation of differential privacy, and obtain a new privacy definition by allowing the *full* trade-off between type I and type II errors in the simple hypothesis testing problem (2) to be governed by some function f. The functional privacy parameter f is to this new definition as (ε, δ) is to the original definition of differential privacy. Notably, this definition that we term f-differential privacy (f-DP)—which captures (ε, δ)-DP as a special case—is accompanied by a powerful and elegant toolkit for reasoning about composition. Here, we highlight some of our contributions:

1.1.1 | An algebra for composition

We show that our privacy definition is *closed* and *tight* under composition, which means that the trade-off between type I and type II errors that results from the composition of an f_1-DP mechanism with an f_2-DP mechanism can always be *exactly* described by a certain function f. This function can be expressed via f_1 and f_2 in an algebraic fashion, thereby allowing for losslessly reasoning about composition. In contrast, (ε, δ)-DP or any other privacy definition artificially restricts itself to a small number of parameters. By allowing for a *function* to keep track of the privacy guarantee of the mechanism, our new privacy definition avoids the pitfall of premature summarization (to quote Holmes, 2019, 'Premature summarization is the root of all evil in statistics'.) in intermediate steps and, consequently, yields a comprehensive delineation of the overall privacy guarantee. See more details in Section 3.

1.1.2 | A central limit phenomenon

We define a single-parameter family of f-DP that uses the type I and type II error trade-off in distinguishing the standard normal distribution $\mathcal{N}(0, 1)$ from $\mathcal{N}(\mu, 1)$ for $\mu \geq 0$. This is referred to as Gaussian differential privacy (GDP). By relating to the hypothesis testing interpretation of differential privacy (2), the GDP guarantee can be interpreted as saying that determining whether or not Alice is in the data set is at least as difficult as telling apart $\mathcal{N}(0, 1)$ and $\mathcal{N}(\mu, 1)$ based on one draw. Moreover, we show that GDP is a 'canonical' privacy guarantee in a fundamental sense: for any privacy definition that retains a hypothesis testing interpretation, we prove that the privacy guarantee of composition with an appropriate scaling converges to GDP in the limit. This central limit theorem type of result is remarkable not only because of its profound theoretical implication, but also for providing a computationally tractable tool for analytically approximating the privacy loss under composition. Figure 1 demonstrates that this tool yields surprisingly accurate approximations to the exact trade-off in testing the hypotheses (2) or substantially improves on the existing privacy guarantee in terms of type I and type II errors. See Sections 2.2 and 3 for a thorough discussion.

1.1.3 | A primal-dual perspective

We show a general duality between f-DP and infinite collections of (ε, δ)-DP guarantees. This duality is useful in two ways. First, it allows one to analyse an algorithm in the framework of f-DP, and then convert back to an (ε, δ)-DP guarantee at the end, if desired. More fundamentally, this duality provides an approach to import techniques developed for (ε, δ)-DP to the framework of f-DP. As an important application, we use this duality to show how to reason simply about privacy amplification by subsampling for f-DP, by leveraging existing results for (ε, δ)-DP. This is in contrast to divergence based notions of privacy, in which reasoning about amplification by subsampling is difficult.

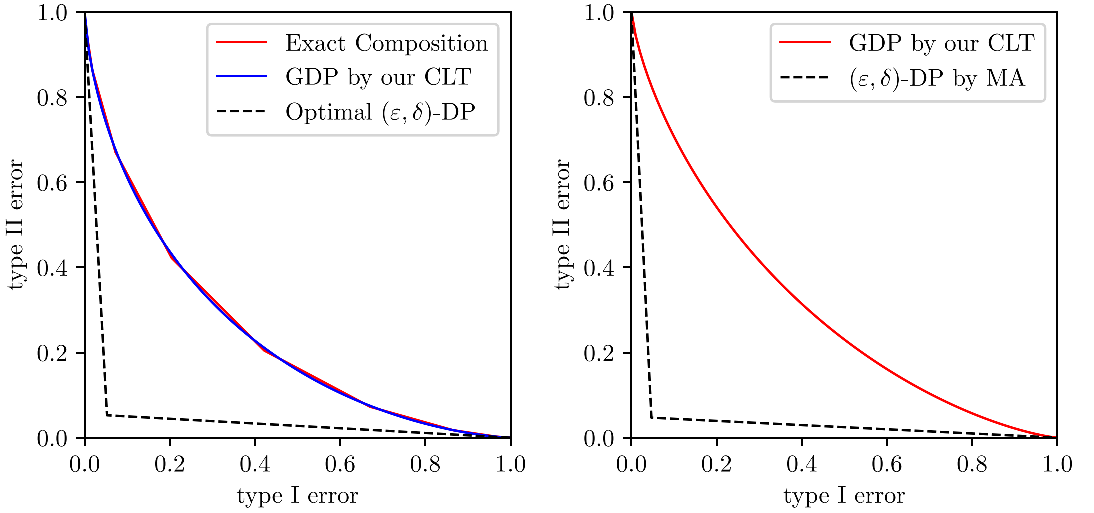

FIGURE 1 Left: Our central limit theorem based approximation (in blue) is very close to the composition of just 10 mechanisms (in red). The tightest possible approximation via an (ε, δ)-DP guarantee (in back) is substantially looser. See Figure 5 for parameter setup. Right: Privacy analysis of stochastic gradient descent used to train a convolutional neural network on MNIST (LeCun & Cortes, 2010). The f-DP framework yields a privacy guarantee (in red) for this problem that is significantly better than the optimal (ε,δ)-DP guarantee (in black) that is derived from the moments accountant (MA) method (Abadi et al., 2016). Put simply, our analysis shows that stochastic gradient descent releases less sensitive information than expected in the literature. See Section 5 for more plots and details [Colour figure can be viewed at wileyonlinelibrary.com]

Taken together, this collection of attractive properties render f-DP a mathematically coherent, computationally efficient and versatile framework for privacy-preserving data analysis. To demonstrate the practical use of this hypothesis testing-based framework, we give a substantially sharper analysis of the privacy guarantees of noisy stochastic gradient descent, improving on previous special-purpose analyses that reasoned about divergences rather than directly about hypothesis testing (Abadi et al., 2016). This application is presented in Section 5.

2 | f-DIFFERENTIAL PRIVACY AND ITS BASIC PROPERTIES

In Section 2.1, we give a formal definition of f-DP. Section 2.2 introduces Gaussian differential privacy, a special case of f-DP. In Section 2.3, we highlight some appealing properties of this new privacy notation from an information-theoretic perspective. Next, Section 2.4 offers a profound connection between f-DP and (ε, δ)-DP. Finally, we discuss the group privacy properties of f-DP.

Before moving on, we first establish several key pieces of notation from the differential privacy literature.

- **Data set.** A data set S is a collection of n records, each corresponding to an individual. Formally, we write the data set as $S = (x_1, \dots, x_n)$, and an individual $x_i \in X$ for some abstract space X. Two data sets $S' = (x'_1, \dots, x'_n)$ and S are said to be *neighbours* if they differ in exactly one record, that is, there exists an index j such that $x_i = x'_i$ for all $i \neq j$ and $x_j \neq x'_j$.

- **Mechanism.** A mechanism M refers to a randomized algorithm that takes as input a data set S and releases some (randomized) statistics $M(S)$ of the data set in some abstract space Y. For example, a mechanism can release the average salary of individuals in the data set plus some random noise.

2.1 | Trade-off functions and *f*-DP

All variants of differential privacy informally require that it be hard to *distinguish* any pairs of neighbouring data sets based on the information released by a private a mechanism M. From an attacker's perspective, it is natural to formalize this notion of 'indistinguishability' as a hypothesis testing problem for two neighbouring data sets S and $S^{'}$:

$$H_0 \ : \ \text{the underlying data set is } S \quad \text{versus} \quad H_1 \ : \ \text{the underlying data set is } S'.$$

The output of the mechanism M serves as the basis for performing the hypothesis testing problem. Denote by P and Q the probability distributions of the mechanism applied to the two data sets, namely $M(S)$ and $M(S^{'})$, respectively. The fundamental difficulty in distinguishing the two hypotheses is best delineated by the *optimal* trade-off between the achievable type I and type II errors. More precisely, consider a rejection rule $0 \leq \phi \leq 1$ that takes as input the released results of the mechanism, with its type I and type II errors defined as

$$\alpha_\phi = \mathbb{E}_P[\phi], \quad \beta_\phi = 1 - \mathbb{E}_Q[\phi],$$

respectively. The two errors satisfy, for example, the well-known constraint $\alpha_\phi + \beta_\phi \geq 1 - \mathrm{TV}(P, Q)$, where the total variation distance $\mathrm{TV}(P, Q)$ is the supremum of $|P(A) - Q(A)|$ over all measurable sets A. Instead of this rough constraint, we seek to characterize the fine-grained trade-off between the two errors. Explicitly, fixing the type I error at *any* level, we consider the minimal achievable type II error. This motivates the following definition.

Definition 2 (trade-off function). For any two probability distributions P and Q on the same space, define the trade-off function $T(P, Q) : [0, 1] \to [0, 1]$ as

$$T(P, Q)(\alpha) = \inf \left\{ \beta_\phi : \alpha_\phi \leq \alpha \right\},$$

where the infimum is taken over all (measurable) rejection rules.

The trade-off function serves as a clear-cut boundary of the achievable and unachievable regions of type I and type II errors, rendering itself the *complete* characterization of the fundamental difficulty in testing between the two hypotheses. The greater this function is, the harder it is to distinguish the two distributions. In particular, the greatest trade-off function is the identity trade-off function $\mathrm{Id}(\alpha) := 1 - \alpha$. Notably, $1 - f$ is the ROC curve for classifying the output as being from the null or alternative hypothesis. For completeness, the minimal β_ϕ can be achieved by the likelihood ratio test—a fundamental result known as the Neyman–Pearson lemma, which is stated in Appendix A for convenience.

A function is called a trade-off function if it is equal to $T(P, Q)$ for some distributions P and Q. Below we give a necessary and sufficient condition for f to be a trade-off function and relegate its proof to Appendix A. This characterization reveals, for example, that $\max\{f, g\}$ is a trade-off function if both f and g are trade-off functions.

Proposition 1 *A function $f: [0, 1] \to [0, 1]$ is a trade-off function if and only if f is convex, continuous, non-increasing and $f(x) \leq 1 - x$ for $x \in [0, 1]$.*

Now, we propose a new generalization of differential privacy built on top of trade-off functions. Below, we write $g \geq f$ for two functions defined on $[0, 1]$ if $g(x) \geq f(x)$ for all $0 \leq x \leq 1$, and we abuse notation by identifying $M(S)$ and $M(S')$ with their corresponding probability distributions. Note that if $T(P, Q) \geq T(\widetilde{P}, \widetilde{Q})$, then in a very strong sense, P and Q are harder to distinguish than $\widetilde{P}$ and $\widetilde{Q}$ at *any* level of type I error.

Definition 3 (f-differential privacy). Let f be a trade-off function. A mechanism M is said to be f-differentially private if

$$T\left(M(S), M(S')\right) \geq f$$

for all neighbouring data sets S and S'.

A graphical illustration of this definition is shown in Figure 2. Letting P and Q be the distributions such that $f = T(P, Q)$, this privacy definition amounts to saying that a mechanism is f-DP if distinguishing any two neighbouring data sets based on the released information is at least as difficult as distinguishing P and Q based on a single draw. In contrast to existing

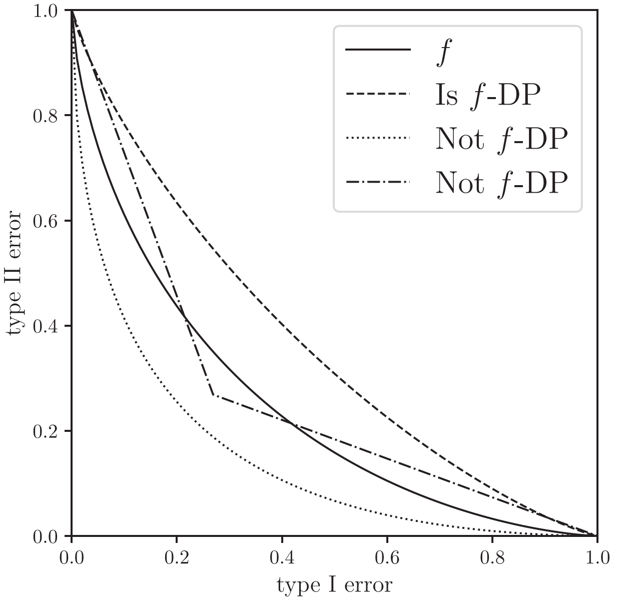

FIGURE 2 Three different examples of $T(M(S), M(S'))$. Only the dashed line corresponds to a trade-off function satisfying f-DP

definitions of differential privacy, our new definition is parameterized by a function, as opposed to several real-valued parameters (e.g., ε and δ). This functional perspective offers a complete characterization of 'privacy', thereby avoiding the pitfall of summarizing statistical information too early. This fact is crucial to the development of a composition theorem for f-DP in Section 3. Although this completeness comes at the cost of increased complexity, as we will see in Section 2.2, a simple family of trade-off functions can often closely capture privacy loss in many scenarios.

Naturally, the definition of f-DP is symmetric in the same sense as the neighbouring relationship, which by definition is symmetric. Observe that this privacy notion also requires

$$T\left(M(S'), M(S)\right) \geq f$$

for any neighbouring pair S, S'. Therefore, it is desirable to restrict our attention to 'symmetric' trade-off functions. Proposition 2 shows that this restriction does not lead to any loss of generality.

Proposition 2 *Let a mechanism M be f-DP. Then, M is f^{S}-DP with $f^{\mathrm{S}} = \max\{f, f^{-1}\}$, where the inverse function is defined as*

$$f^{-1}(\alpha) := \inf\{t \in [0,1] : f(t) \leq \alpha\} \tag{3}$$

for $\alpha \in [0, 1]$.

We prove Proposition 2 in Appendix A. Writing $f = T(P, Q)$, we can express the inverse as $f^{-1} = T(Q, P)$, which therefore is also a trade-off function. As a consequence of this, f^{S} continues to be a trade-off function by making use of Proposition 1 and, moreover, is *symmetric* in the sense that

$$f^{\mathrm{S}} = (f^{\mathrm{S}})^{-1}.$$

Importantly, this symmetrization gives a tighter bound in the privacy definition since $f^{\mathrm{S}} \geqslant f$. In the remainder of the paper, therefore, trade-off functions will always be assumed to be symmetric unless otherwise specified.

We conclude this subsection by showing that f-DP is a generalization of (ε, δ)-DP. This foreshadows a deeper connection between f-DP and (ε, δ)-DP that will be discussed in Section 2.4. Denote

$$f_{\varepsilon,\delta}(\alpha) = \max\left\{0, 1 - \delta - \mathrm{e}^{\varepsilon}\alpha, \mathrm{e}^{-\varepsilon}(1 - \delta - \alpha)\right\} \tag{4}$$

for $0 \leq \alpha \leq 1$, which is a trade-off function. Figure 3 shows the graph of this function and its evident symmetry. The following result is adapted from Wasserman and Zhou (2010).

Proposition 3 (Wasserman & Zhou, 2010). *A mechanism M is (ε, δ)-DP if and only if M is $f_{\varepsilon,\delta}$-DP.*

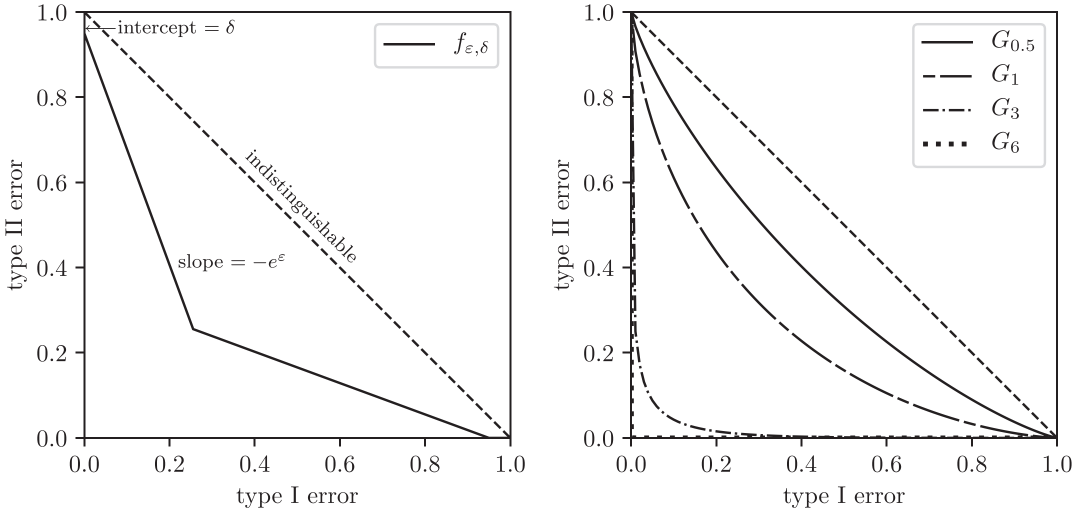

FIGURE 3 Left: $f_{\varepsilon,\delta}$ is a piecewise linear function and is symmetric with respect to the line $y = x$. It has (nontrivial) slopes $-e^{\pm\varepsilon}$ and intercepts $1 - \delta$. Right: Trade-off functions of unit-variance Gaussian distributions with different means. The case of $\mu = 0.5$ is reasonably private, $\mu = 1$ is borderline private, and $\mu = 3$ is basically non-private: an adversary can control type I and type II errors simultaneously at only 0.07. In the case of $\mu = 6$ (almost coincides with the axes), the two errors both can be as small as 0.001

2.2 | Gaussian differential privacy

This subsection introduces a parametric family of f-DP guarantees, where f is the trade-off function of two normal distributions. We refer to this specialization as Gaussian differential privacy (GDP). GDP enjoys many desirable properties that lead to its central role in this paper. Among others, we can now precisely define the trade-off function with a single parameter. To define this notion, let

$$G_\mu : = T\left(\mathcal{N}(0,1), \mathcal{N}(\mu,1)\right)$$

for $\mu \geq 0$. An explicit expression for the trade-off function G_μ reads

$$G_\mu(\alpha) = \Phi\left(\Phi^{-1}(1-\alpha) - \mu\right), \tag{5}$$

where Φ denotes the standard normal CDF. For completeness, we provide a proof of (5) in Appendix A. This trade-off function is decreasing in μ in the sense that $G_\mu \leq G_{\mu'}$ if $\mu \geq \mu'$. We now define GDP:

Definition 4 A mechanism M is said to satisfy μ-Gaussian Differential Privacy (μ-GDP) if it is G_μ-DP. That is,

$$T(M(S), M(S')) \geq G_\mu$$

for all neighbouring data sets S and $S^{'}$.

GDP has several attractive properties. First, this privacy definition is fully described by the single mean parameter of a unit-variance Gaussian distribution, which makes it easy to describe and interpret the privacy guarantees. For instance, one can see from the right panel of Figure 3 that $\mu \le 0.5$ guarantees a reasonable amount of privacy, whereas if $\mu \geqslant 6$, almost nothing is being promised. Second, loosely speaking, GDP occupies a role among all hypothesis testing based notions of privacy that is similar to the role that the Gaussian distribution has among general probability distributions. We formalize this important point by proving central limit theorems for f-DP in Section 3, which, roughly speaking, says that f-DP converges to GDP under composition in the limit. Lastly, as shown in the remainder of this subsection, GDP *precisely* characterizes the Gaussian mechanism, one of the most fundamental building blocks of differential privacy.

Consider the problem of privately releasing a univariate statistic $\theta(S)$ of the data set S. Define the sensitivity of θ as

$$\mathrm{sens}(\theta) = \sup_{S,S'} |\theta(S) - \theta(S')|,$$

where the supremum is over all neighbouring data sets. The Gaussian mechanism adds Gaussian noise to the statistic θ in order to obscure whether θ is computed on S or S'. The following result shows that the Gaussian mechanism with noise properly scaled to the sensitivity of the statistic satisfies GDP.

Theorem 1 *Define the Gaussian mechanism that operates on a statistic θ as $M(S) = \theta(S) + \xi$, where $\xi \sim \mathcal{N}(0, \mathrm{sens}(\theta)^2/\mu^2)$. Then, M is μ-GDP.*

Proof of Theorem 1 Recognizing that $M(S)$, $M(S')$ are normally distributed with means $\theta(S)$, $\theta(S')$, respectively, and common variance $\sigma^2 = \mathrm{sens}(\theta)^2/\mu^2$, we get

$$T(M(S), M(S')) = T(\mathcal{N}(\theta(S), \sigma^2), \mathcal{N}(\theta(S'), \sigma^2)) = G_{|\theta(S)-\theta(S')|/\sigma}.$$

By the definition of sensitivity, $|\theta(S) - \theta(S')|/\sigma \le \mathrm{sens}(\theta)/\sigma = \mu$. Therefore, we get

$$T(M(S), M(S')) = G_{|\theta(S)-\theta(S')|/\sigma} \geqslant G_\mu.$$

This completes the proof.

As implied by the proof above, GDP offers the tightest possible privacy bound of the Gaussian mechanism. More precisely, the Gaussian mechanism in Theorem 1 satisfies

$$G_\mu(\alpha) = \inf_{\text{neighbouring } S,S'} T(M(S), M(S'))(\alpha), \tag{6}$$

where the infimum is (asymptotically) achieved at the two neighbouring data sets such that $|\theta(S) - \theta(S')| = \mathrm{sens}(\theta)$ *irrespective* of the type I error α. As such, the characterization by GDP is precise in the pointwise sense. In contrast, the right-hand side of Equation (6) in general is not necessarily a convex function of α and, in such case, is not a trade-off function according to Proposition 1. This nice property of Gaussian mechanism is related to the log-concavity of Gaussian distributions. See Proposition A.3 for a detailed treatment of log-concave distributions.

2.3 | Post-processing and the informativeness of *f*-DP

Intuitively, a data analyst cannot make a statistical analysis more disclosive only by processing the output of the mechanism M. This is called the post-processing property, a natural requirement that any notion of privacy, including our definition of f-DP, should satisfy.

To formalize this point for f-DP, denote by $\mathrm{Proc} : Y \to Z$ a (randomized) algorithm that maps the input $M(S) \in Y$ to some space Z, yielding a new mechanism that we denote by $\mathrm{Proc} \circ M$. The following result confirms the post-processing property of f-DP.

Proposition 4 *If a mechanism M is f-DP, then its post-processing $\mathrm{Proc} \circ M$ is also f-DP.*

Proposition 4 is a consequence of the following lemma. Let $\mathrm{Proc}(P)$ be the probability distribution of $\mathrm{Proc}(\zeta)$ with ζ drawn from P. Define $\mathrm{Proc}(Q)$ likewise.

Lemma 1 *For any two distributions P and Q, we have $T(\mathrm{Proc}(P), \mathrm{Proc}(Q)) \geqslant \mathrm{T}(P, Q)$.*

This lemma means that post-processed distributions can only become more difficult to tell apart than the original distributions from the perspective of trade-off functions. While the same property holds for many divergence based measures of indistinguishability such as the Rényi divergences used by the concentrated differential privacy family of definitions (Bun & Steinke, 2016; Bun et al., 2018a; Dwork & Rothblum, 2016; Mironov, 2017), a consequence of the following theorem is that trade-off functions offer the most informative measure among all. This remarkable inverse of Lemma 1 is due to Blackwell (see also Theorem 2.5 in Kairouz et al., 2017).

Theorem 2 (Blackwell, 1950, Theorem 10). *Let P, Q be probability distributions on Y and P', Q' be probability distributions on Z. The following two statements are equivalent:*

1. *$T(P, Q) \le T(P', Q')$.*
2. *There exists a randomized algorithm* $\mathrm{Proc}: Y \to Z$ *such that* $\mathrm{Proc}(P) = P'$, $\mathrm{Proc}(Q) = Q'$.

To appreciate the implication of this theorem, we begin by observing that post-processing induces an order on pairs of distributions, which is called the Blackwell order (see, e.g., Raginsky, 2011). Specifically, if the above condition (b) holds, then we write $(P, Q) \preceq_{\mathrm{Blackwell}} (P', Q')$ and interpret this as '(P, Q) is easier to distinguish than (P', Q') in the Blackwell sense'. Similarly, when $T(P, Q) \le T(P', Q')$, we write $(P, Q) \preceq_{\mathrm{tradeoff}} (P', Q')$ and interpret this as '(P, Q) is easier to distinguish than (P', Q') in the testing sense'. In general, any privacy measure used in defining a privacy notion induces an order $\preceq$ on pairs of distributions. Assuming the post-processing property for the privacy notion, the induced order $\preceq$ must be consistent with $\preceq_{\mathrm{Blackwell}}$. Concretely, we denote by $\mathrm{Ineq}(\preceq) = \{(P, Q; P', Q') : (P, Q) \preceq (P', Q')\}$ the set of all comparable pairs of the order $\preceq$. As is clear, a privacy notion satisfies the post-processing property if and only if the induced order $\preceq$ satisfies $\mathrm{Ineq}(\preceq) \supseteq \mathrm{Ineq}(\preceq_{\mathrm{Blackwell}})$.

Therefore, for any reasonable privacy notion, the set $\mathrm{Ineq}(\preceq)$ must be large enough to contain $\mathrm{Ineq}(\preceq_{\mathrm{Blackwell}})$. However, it is also desirable to have a not too large $\mathrm{Ineq}(\preceq)$. For example, consider the privacy notion based on a trivial divergence D_0 with $D_0(P \| Q) \equiv 0$ for any P, Q. Note that $\mathrm{Ineq}(\preceq_{D_0})$ is the largest possible and, meanwhile, it is not informative at all in terms of measuring the indistinguishability of two distributions.

The argument above suggests that going from the 'minimal' order $\mathrm{Ineq}(\preceq_{\text{Blackwell}})$ to the 'maximal' order $\mathrm{Ineq}(\preceq_{D_0})$ would lead to information loss. Remarkably, f-DP is the most informative differential privacy notion from this perspective because its induced order $\preceq_{\text{tradeoff}}$ satisfies $\mathrm{Ineq}(\preceq_{\text{tradeoff}}) = \mathrm{Ineq}(\preceq_{\text{Blackwell}})$. In stark contrast, this is not true for the order induced by other popular privacy notions such as Rényi differential privacy and (ε, δ)-DP. We prove this claim in Appendix B and further justify the informativeness of f-DP by providing general tools that can losslessly convert f-DP guarantees into divergence based privacy guarantees.

2.4 | A primal-dual perspective

In this subsection, we show that f-DP is equivalent to an infinite *collection* of (ε, δ)-DP guarantees via the convex conjugate of the trade-off function. As a consequence of this, we can view f-DP as the *primal* privacy representation and, accordingly, its *dual* representation is the collection of (ε, δ)-DP guarantees. Taking this powerful viewpoint, many results from the large body of (ε, δ)-DP work can be carried over to f-DP in a seamless fashion. In particular, this primal-dual perspective is crucial to our analysis of 'privacy amplification by subsampling' in Section 4. All proofs are deferred to Appendix A.

First, we present the result that converts a collection of (ε, δ)-DP guarantees into an f-DP guarantee. This result is self-evidence and its proof is, therefore, omitted.

Proposition 5 (Dual to primal). *Let I be an arbitrary index set such that each $i \in I$ is associated with $\varepsilon_i \in [0, \infty)$ and $\delta_i \in [0, 1]$. A mechanism is $(\varepsilon_i, \delta_i)$-DP for all $i \in I$ if and only if it is f-DP with*

$$f = \sup_{i \in I} f_{\varepsilon_i, \delta_i}.$$

This proposition follows easily from the equivalence of (ε, δ)-DP and $f_{\varepsilon,\delta}$-DP. We remark that the function f constructed above remains a symmetric trade-off function.

The more interesting direction is to convert f-DP into a collection of (ε, δ)-DP guarantees. Recall that the convex conjugate of a function g defined on $(-\infty, \infty)$ is defined as

$$g^*(y) = \sup_{-\infty < x < \infty} yx - g(x). \tag{7}$$

To define the conjugate of a trade-off function f, we extend its domain by setting $f(x) = \infty$ for $x < 0$ and $x > 1$. With this adjustment, the supremum is effectively taken over $0 \le x \le 1$.

Proposition 6 (Primal to dual). *For a symmetric trade-off function f, a mechanism is f-DP if and only if it is $(\varepsilon, \delta(\varepsilon))$-DP for all $\varepsilon \geqslant 0$ with $\delta(\varepsilon) = 1 + f^*(-e^{\varepsilon})$.*

For example, taking $f = G_\mu$, the following corollary provides a lossless conversion from GDP to a collection of (ε, δ)-DP guarantees. This conversion is exact and, therefore, any other (ε, δ)-DP guarantee derived for the Gaussian mechanism is implied by this corollary. See Figure 4 for an illustration of this result.

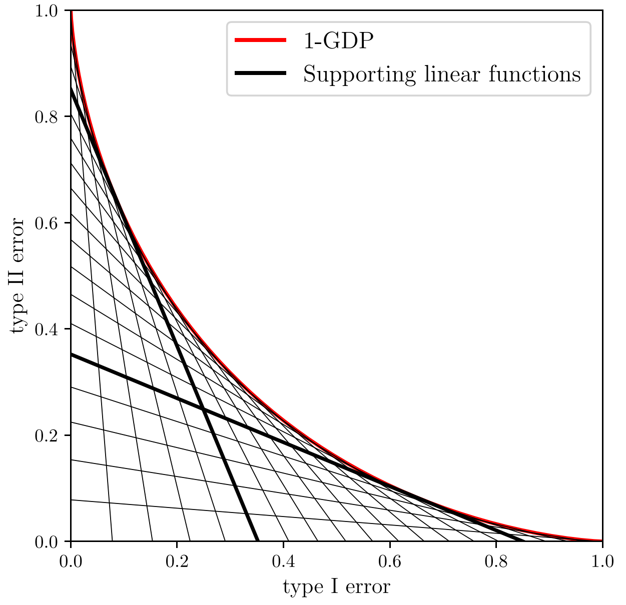

FIGURE 4 Each (ε, $\delta(\varepsilon)$)-DP guarantee corresponds to two supporting linear functions (symmetric to each other) to the trade-off function describing the complete f-DP guarantee. In general, characterizing a privacy guarantee using only a subset of (ε, δ)-DP guarantees (for example, only those with small δ) would result in information loss [Colour figure can be viewed at wileyonlinelibrary.com]

Corollary 1 *A mechanism is μ-GDP if and only if it is (ε, $\delta(\varepsilon)$)-DP for all $\varepsilon \geqslant 0$, where*

$$\delta(\varepsilon) = \Phi\left(-\frac{\varepsilon}{\mu} + \frac{\mu}{2}\right) - \mathrm{e}^{\varepsilon}\Phi\left(-\frac{\varepsilon}{\mu} - \frac{\mu}{2}\right).$$

This corollary has appeared earlier in Balle and Wang (2018). Along this direction, Balle et al. (2018) further proposed ‘privacy profile’, which in essence corresponds to an infinite collection of (ε, δ). The notion of privacy profile mainly serves as an analytical tool in Balle et al. (2018).

The primal-dual perspective provides a useful tool through which we can bridge the two privacy definitions. In some cases, it is easier to work with f-DP by leveraging the interpretation and informativeness of trade-off functions, as seen from the development of composition theorems for f-DP in Section 3. Meanwhile, (ε, δ)-DP is more convenient to work with in the cases where the lower complexity of two parameters ε, δ is helpful, for example, in the proof of the privacy amplification by subsampling theorem for f-DP. In short, our approach in Section 4 is to first work in the dual world and use existing subsampling theorems for (ε, δ)-DP, and then convert the results back to f-DP using a slightly more advanced version of Proposition 6.

2.5 | Group privacy

The notion of f-DP can be extended to address privacy of a *group* of individuals, and a question of interest is to quantify how privacy degrades as the group size grows. To set up the notation, we say that two data sets S, S' are k-neighbours (where $k \geq 2$ is an integer) if there exist data sets $S = S_0, S_1, \ldots, S_k = S'$ such that S_i and S_{i+1} are neighbouring or identical for all $i = 0, \ldots, k-1$.

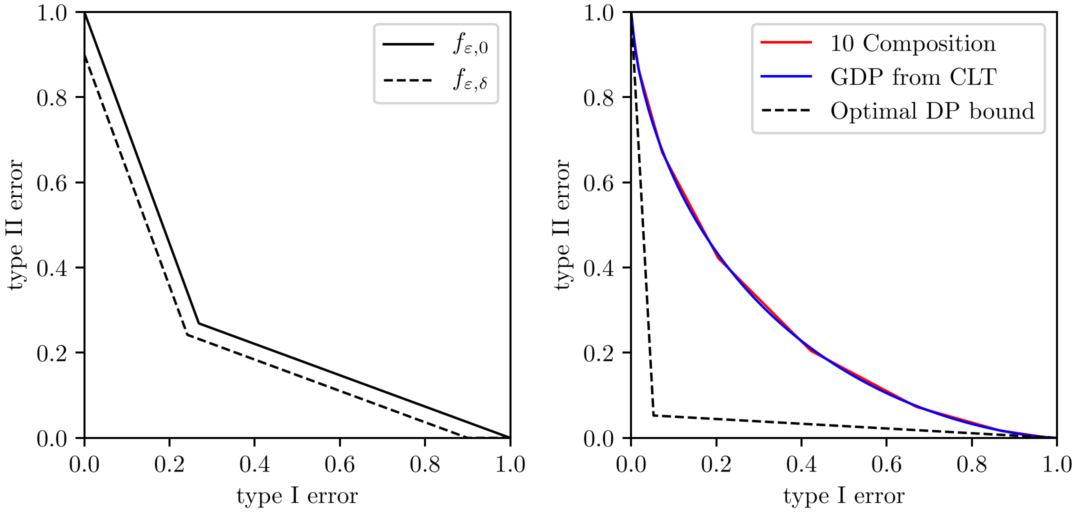

FIGURE 5 Left: Tensoring with $f_{0,\delta}$ scales the graph towards the origin by a factor of $1-\delta$. Right: Tenfold composition of $(1/\sqrt{10}, 0)$-DP mechanisms, that is, $f_{\varepsilon,0}^{\otimes n}$ with $n = 10$, $\varepsilon = 1/\sqrt{n}$. The dashed curve corresponds to $\varepsilon = 2.89$, $\delta = 0.001$. These values are obtained by first setting δ=0.001 and finding the smallest ε such that the composition is (ε, δ)-DP. Note that the central limit theorem approximation to the true trade-off curve is almost perfect, whereas the tightest possible approximation via (ε, δ)-DP is substantially looser [Colour figure can be viewed at wileyonlinelibrary.com]

Equivalently, S, S' are k-neighbours if they differ by at most k individuals. Accordingly, a mechanism M is said to be f-DP for *groups of size k* if

$$T(M(S), M(S')) \geqslant f$$

for all k-neighbours S and S'.

In the following theorem, we use $h^{\circ k}$ to denote the k-fold iterative composition of a function h. For example, $h^{\circ 1} = h$ and $h^{\circ 2}(x)=h(h(x))$.

Theorem 3 *If a mechanism is f-DP, then it is* $[1-(1-f)^{\circ k}]$*-DP for groups of size k. In particular, if a mechanism is μ-GDP, then it is* kμ*-GDP for groups of size k.*

For completeness, $1-(1-f)^{\circ k}$ is a trade-off function and, moreover, remains symmetric if f is symmetric. These two facts and Theorem 3 are proved in Appendix A. As revealed in the proof, the privacy bound $1-(1-f)^{\circ k}$ in general cannot be improved, thereby showing that the group operation in the f-DP framework is *closed* and *tight*. In addition, it is easy to see that $1-(1-f)^{\circ k} \leq 1-(1-f)^{\circ(k-1)}$ by recognizing that the trade-off function f satisfies $1-f(x) \geqslant x$. This is consistent with the intuition that detecting changes in groups of k individuals becomes easier as the group size increases.

As an interesting consequence of Theorem 3, the group privacy of ε-DP in the limit corresponds to the trade-off function of two Laplace distributions. Recall that the density of Lap(μ, b) is $\frac{1}{2b}e^{-|x-\mu|/b}$.

Proposition 7 *Fix $\mu \ge 0$ and set $\varepsilon = \mu/k$. As $k \to \infty$, we have*

$$1-(1-f_{\varepsilon,0})^{\circ k} \to T(\mathrm{Lap}(0,1), \mathrm{Lap}(\mu,1)).$$

The convergence is uniform over [0, 1].

Two remarks are in order. First, $T(\mathrm{Lap}(0, 1), \mathrm{Lap}(\mu, 1))$ is not equal to $f_{\varepsilon,\delta}$ for any ε, δ and, therefore, (ε, δ)-DP is not expressive enough to measure privacy under the group operation. Second, the approximation in this theorem is very accurate even for small k. For example, for $\mu = 1, k = 4$, the function $1-(1-f_{\varepsilon,0})^{\circ k}$ is within 0.005 of $T(\mathrm{Lap}(0, 1), \mathrm{Lap}(\mu, 1))$ uniformly over [0, 1]. The proof of Proposition 7 is deferred to Appendix A.

3 | COMPOSITION AND LIMIT THEOREMS

Imagine that an analyst performs a sequence of analyses on a private data set, in which each analysis is informed by prior analyses on the same data set. Provided that every analysis alone is private, the question is whether all analyses collectively are private, and if so, how the privacy degrades as the number of analyses increases, namely under composition. It is essential for a notion of privacy to gracefully handle composition, without which the privacy analysis of complex algorithms would be almost impossible.

Now, we describe the composition of two mechanisms. For simplicity, this section writes X for the space of data sets and abuse notation by using n to refer to the number of mechanisms in composition (the use of n is consistent with the literature on central limit theorems). Let $M_1 : X \to Y_1$ be the first mechanism and $M_2 : X \times Y_1 \to Y_2$ be the second mechanism. In brief, M_2 takes as input the output of the first mechanism M_1 in addition to the data set. With the two mechanisms in place, the joint mechanism $M : X \to Y_1 \times Y_2$ is defined as

$$M(S) = (y_1, M_2(S, y_1)), \tag{8}$$

where $y_1 = M_1(S)$. Roughly speaking, the distribution of $M(S)$ is constructed from the marginal distribution of $M_1(S)$ on Y_1 and the conditional distribution of $M_2(S, y_1)$ on Y_2 given $M_1(S) = y_1$. The composition of more than two mechanisms follows recursively. In general, given a sequence of mechanisms $M_i : X \times Y_1 \times \cdots \times Y_{i-1} \to Y_i$ for $i = 1, 2, \ldots, n$, we can recursively define the joint mechanism as their composition:

$$M : X \to Y_1 \times \cdots \times Y_n.$$

Put differently, $M(S)$ can be interpreted as the trajectory of a Markov chain whose initial distribution is given by $M_1(S)$ and the transition kernel $M_i(S, \ldots)$ at each step.

Using the language above, the goal of this section is to relate the privacy loss of M to that of the n mechanisms $M_1, \ldots, M_n$ in the f-DP framework. In short, Section 3.1 develops a general composition theorem for f-DP. In Section 3.2, we identify a central limit theorem phenomenon of composition in the f-DP framework, which can be used as an approximation tool, just like we use the central limit theorem for random variables. This approximation is extended to and improved for (ε, δ)-DP in Section 3.3.

3.1 | A general composition theorem

The main thrust of this subsection is to demonstrate that the composition of private mechanisms is closed and tight in the f-DP framework. This result is formally stated in Theorem 4, which shows that the composed mechanism remains f-DP with the trade-off function taking the form of a certain product. To define the product, consider two trade-off functions f and g that are given as $f = T(P, Q)$ and $g = T(P', Q')$ for some probability distributions P, P', Q, Q'.

Definition 5 The tensor product of two trade-off functions $f = T(P, Q)$ and $g = T(P', Q')$ is defined as

$$f \otimes g := T(P \times P', Q \times Q').$$

Throughout the paper, write $f \otimes g(\alpha)$ for $(f \otimes g)(\alpha)$, and denote by $f^{\otimes n}$ the n-fold tensor product of f. The well-definedness of $f^{\otimes n}$ rests on the associativity of the tensor product, which we will soon illustrate.

By definition, $f \otimes g$ is also a trade-off function. Nevertheless, it remains to be shown that the tensor product is well defined: that is, the definition is independent of the choice of distributions used to represent a trade-off function. More precisely, assuming $f = T(P, Q) = T(\tilde{P}, \tilde{Q})$ for some distributions $\tilde{P}, \tilde{Q}$, we need to ensure that

$$T(P \times P', Q \times Q') = T(\tilde{P} \times P', \tilde{Q} \times Q').$$

We defer the proof of this intuitive fact to Appendix C. Below we list some other useful properties of the tensor product of trade-off functions, whose proofs are placed in Appendix D.

1. The product $\otimes$ is commutative and associative.
2. If $g_1 \geqslant g_2$, then $f \otimes g_1 \geqslant f \otimes g_2$.
3. $f \otimes \mathrm{Id} = \mathrm{Id} \otimes f = f$, where the identity trade-off function $\mathrm{Id}(x) = 1 - x$ for $0 \le x \le 1$.
4. $(f \otimes g)^{-1} = f^{-1} \otimes g^{-1}$. See the definition of inverse in Equation (3).

Note that Id is the trade-off function of two identical distributions. Property 4 implies that when f, g are symmetric trade-off functions, their tensor product $f \otimes g$ is also symmetric.

Now we state the main theorem of this subsection. Its proof is given in Appendix C.

Theorem 4 *Let $M_i(\cdot, y_1, \ldots, y_{i-1})$ be f_i-DP for all $y_1 \in Y_1, \ldots, y_{i-1} \in Y_{i-1}$. Then the n-fold composed mechanism $M : X \to Y_1 \times \cdots \times Y_n$ is $f_1 \otimes \cdots \otimes f_n$-DP.*

This theorem shows that the composition of mechanisms remains f-DP or, put differently, composition is closed in the f-DP framework. Moreover, the privacy bound $f_1 \otimes \cdots \otimes f_n$ in Theorem 4 is *tight* in the sense that it cannot be improved in general. To see this point, consider the case where the second mechanism completely ignores the output of the first mechanism. In that case, the composition obeys

$$\begin{aligned} T(M(S), M(S')) &= T(M_1(S) \times M_2(S), M_1(S') \times M_2(S')) \\ &= T(M_1(S), M_1(S')) \otimes T(M_2(S), M_2(S')). \end{aligned}$$

Next, taking neighbouring data sets such that $T(M_1(S), M_1(S')) = f_1$ and $T(M_2(S), M_2(S')) = f_2$, one concludes that $f_1 \otimes f_2$ is the tightest possible bound on the twofold composition. For comparison, the advanced composition theorem for (ε, δ)-DP does not admit a single pair of optimal parameters ε, δ (Dwork et al., 2010). In particular, no pair of ε, δ can exactly capture the privacy of the composition of (ε, δ)-DP mechanisms. See Section 3.3 and Figure 5 for more elaboration.

In the case of GDP, composition enjoys a simple and convenient formulation due to the identity

$$G_{\mu_1} \otimes G_{\mu_2} \otimes \cdots \otimes G_{\mu_n} = G_{\mu},$$

where $\mu = \sqrt{\mu_1^2 + \cdots + \mu_n^2}$. This formula is due to the rotational invariance of Gaussian distributions with identity covariance. We provide the proof in Appendix D. The following corollary formally summarizes this finding.

Corollary 2 *The n-fold composition of μ_i-GDP mechanisms is $\sqrt{\mu_1^2 + \cdots + \mu_n^2}$-GDP.*

On a related note, the pioneering work Kairouz et al. (2017) is the first to take the hypothesis testing viewpoint in the study of privacy composition and to use Blackwell's theorem as an analytic tool therein. In particular, the authors offered a composition theorem for (ε, δ)-DP that improves on the advanced composition theorem (Dwork et al., 2010). Following this work, Murtagh and Vadhan (2016) provided a self-contained proof by essentially proving the '(ε, δ) special case' of Blackwell's theorem. In contrast, our novel proof of Theorem 4 only makes use of the Neyman–Pearson lemma, thereby circumventing the heavy machinery of Blackwell's theorem. This simple proof better illuminates the essence of the composition theorem.

3.2 | Central limit theorems for composition

In this subsection, we identify a central limit theorem type phenomenon of composition in the f-DP framework. Our main results (Theorems 5 and 6), roughly speaking, show that trade-off functions corresponding to small privacy leakage accumulate to G_μ for some μ under composition. Equivalently, the privacy of the composition of many 'very private' mechanisms is best measured by GDP in the limit. This identifies GDP as the focal privacy definition among the family of f-DP privacy guarantees, including (ε, δ)-DP. More precisely, *all* privacy definitions that are based on a hypothesis testing formulation of 'indistinguishability' converge to the guarantees of GDP in the limit of composition. We remark that Sommer et al. (2018) proved a conceptually related central limit theorem for random variables corresponding to the privacy loss. This theorem is used to reason about the non-adaptive composition for (ε, δ)-DP. In contrast, our central limit theorem is concerned with the optimal hypothesis testing trade-off functions for the composition theorem. Moreover, our theorem is applicable in the setting of composition, where each mechanism is informed by prior interactions with the same database.

From a computational viewpoint, these limit theorems yield an efficient method of approximating the composition of general f-DP mechanisms. This is very appealing for analysing the privacy properties of algorithms that are comprised of many building blocks in a sequence. For comparison, the exact computation of privacy guarantees under composition can be computationally hard (Murtagh & Vadhan, 2016) and, thus, tractable approximations are important. Using our central limit theorems, the computation of the exact overall privacy guarantee $f_1 \otimes \cdots \otimes f_n$ in Theorem 4 can be reduced to the evaluation of a single mean parameter μ in a GDP guarantee. We give an exemplary application of this powerful technique in Section 5.

Explicitly, the mean parameter μ in the approximation depends on certain functionals of the trade-off functions:

$$\mathrm{kl}(f) := -\int_0^1 \log|f'(x)|\mathrm{d}x, \quad \kappa_2(f) := \int_0^1 \log^2|f'(x)|\mathrm{d}x$$
$$\kappa_3(f) := \int_0^1 |\log|f'(x)||^3\,\mathrm{d}x, \quad \overline{\kappa}_3(f) := \int_0^1 |\log|f'(x)| + \mathrm{kl}(f)|^3\,\mathrm{d}x.$$

All of these functionals take values in $[0, +\infty]$, and the last is defined for f such that $\mathrm{kl}(f) < \infty$. In essence, these functionals are calculating moments of the log-likelihood ratio of P and Q such that $f = T(P, Q)$. In particular, all of these functionals are 0 if $f(x) = \mathrm{Id}(x) = 1 - x$, which corresponds to zero privacy leakage. As its name suggests, $\mathrm{kl}(f)$ is the Kullback–Leibler (KL) divergence of P and Q and, therefore, $\mathrm{kl}(f) \geq 0$. Detailed elaboration on these functionals is deferred to Appendix D.

In the following theorem, $\mathbf{kl}$ denotes the vector $(\mathrm{kl}(f_1), \ldots, \mathrm{kl}(f_n))$ and $\boldsymbol{\kappa_2}, \boldsymbol{\kappa_3}, \overline{\boldsymbol{\kappa}}_\mathbf{3}$ are defined similarly; in addition, $\|\cdot\|_1$ and $\|\cdot\|_2$ are the ℓ_1 and ℓ_2 norms, respectively. Its proof can be found in Appendix D.

Theorem 5 *Let $f_1, \ldots, f_n$ be symmetric trade-off functions such that $\kappa_3(f_i) < \infty$ for all $1 \leq i \leq n$. Denote*

$$\mu := \frac{2\|\mathbf{kl}\|_1}{\sqrt{\|\boldsymbol{\kappa_2}\|_1 - \|\mathbf{kl}\|_2^2}} \quad \text{and} \quad \gamma := \frac{0.56\,\|\overline{\boldsymbol{\kappa}}_\mathbf{3}\|_1}{(\|\boldsymbol{\kappa_2}\|_1 - \|\mathbf{kl}\|_2^2)^{3/2}}$$

and assume $\gamma < \frac{1}{2}$. Then, for all $\alpha \in [\gamma, 1-\gamma]$, we have

$$G_\mu(\alpha + \gamma) - \gamma \leq f_1 \otimes f_2 \otimes \cdots \otimes f_n(\alpha) \leq G_\mu(\alpha - \gamma) + \gamma. \tag{9}$$

From a technical viewpoint, Theorem 5 can be thought of as a Berry–Esseen type central limit theorem. Loosely speaking, the lower bound in Equation (9) shows that the composition of f_i-DP mechanisms for $i = 1, \ldots, n$ is approximately μ-GDP and, in addition, the upper bound demonstrates that the tightness of this approximation is specified by γ. In the case where all f_i are equal to some $f \neq \mathrm{Id}$, the theorem reveals that the composition becomes blatantly non-private as $n \to \infty$ because $\mu \asymp \sqrt{n} \to \infty$. More interesting applications of the theorem, however, are cases where each f_i is close to the 'perfect privacy' trade-off function Id such that collectively μ is convergent and γ vanishes as $n \to \infty$ (see the example in Section 5). For completeness, the condition $\kappa_3(f_i) < \infty$ (which implies that the other three functionals are also finite) for the use of this theorem excludes the case where $f_i(0) < 1$, in particular, $f_{\varepsilon,\delta}$ in (ε, δ)-DP with $\delta > 0$. We introduce an easy and general technique in Section 3.3 to deal with this issue.

Next, we present an asymptotic version of Theorem 5 for composition of f-DP mechanisms. In analog to classical central limit theorems, below we consider a triangular array of mechanisms $\{M_{n1}, \ldots, M_{nn}\}_{n=1}^\infty$, where M_{ni} is f_{ni}-DP for $1 \leq i \leq n$. As with Theorem 5, the proof of Theorem 6 is relegated to Appendix D.

Theorem 6 *Let $\{f_{ni} : 1 \le i \le n\}_{n=1}^{\infty}$ be a triangular array of symmetric trade-off functions and assume the following limits for some constants $K \ge 0$ and $s > 0$ as $n \to \infty$:*

1. $\sum_{i=1}^{n} \mathrm{kl}(f_{ni}) \to K$;
2. $\max_{1 \le i \le n} \mathrm{kl}(f_{ni}) \to 0$;
3. $\sum_{i=1}^{n} \kappa_2(f_{ni}) \to s^2$;
4. $\sum_{i=1}^{n} \kappa_3(f_{ni}) \to 0$.

Then, we have

$$\lim_{n\to\infty} f_{n1} \otimes f_{n2} \otimes \cdots \otimes f_{nn}(\alpha) = G_{2K/s}(\alpha)$$

uniformly for all $\alpha \in [0, 1]$.

Taken together, this theorem and Theorem 4 amount to saying that the composition $M_{n1} \otimes \ldots \otimes M_{nn}$ is asymptotically $2K/s$-GDP. In fact, this asymptotic version is a consequence of Theorem 5 as one can show $\mu \to 2K/s$ and $\gamma \to 0$ for the triangular array of symmetric trade-off functions. This central limit theorem implies that GDP is the *only* parameterized family of trade-off functions that can faithfully represent the effects of composition. In contrast, neither ε- nor (ε, δ)-DP can losslessly be tracked under composition—the parameterized family of functions $f_{\varepsilon,\delta}$ cannot represent the trade-off function that results from the limit under composition.

The conditions for use of this theorem are reminiscent of Lindeberg's condition in the central limit theorem for independent random variables. The proper scaling of the trade-off functions is that both $\mathrm{kl}(f_{ni})$ and $\kappa_2(f_{ni})$ are of order $O(1/n)$ for most $1 \le i \le n$. As a consequence, the cumulative effects of the moment functionals are bounded. Furthermore, as with Lindeberg's condition, the second condition in Theorem 6 requires that no single mechanism has a significant contribution to the composition in the limit.

In passing, we remark that K and s satisfy the relationship $s = \sqrt{2K}$ in all examples of the application of Theorem 6 in this paper, including Theorems 7 and 11 as well as their corollaries. As such, the composition is asymptotically s-GDP. A proof of this interesting observation or the construction of a counterexample is left for future work.

3.3 | Composition of (ε, δ)-DP: Beating Berry–Esseen

Now, we extend central limit theorems to (ε, δ)-DP. As shown by Proposition 3, (ε, δ)-DP is equivalent to $f_{\varepsilon,\delta}$-DP and, therefore, it suffices to approximate the trade-off function $f_{\varepsilon_1,\delta_1} \otimes \cdots \otimes f_{\varepsilon_n,\delta_n}$ by making use of the composition theorem for f-DP mechanisms. As pointed out in Section 3.2, however, the moment conditions required in the two central limit theorems (Theorems 5 and 6) exclude the case where $\delta_i > 0$.

To overcome the difficulty caused by a nonzero δ, we start by observing the useful fact that

$$f_{\varepsilon,\delta} = f_{\varepsilon,0} \otimes f_{0,\delta}. \tag{10}$$

This decomposition, along with the commutative and associative properties of the tensor product, shows

$$f_{\varepsilon_1,\delta_1} \otimes \cdots \otimes f_{\varepsilon_n,\delta_n} = (f_{\varepsilon_1,0} \otimes \cdots \otimes f_{\varepsilon_n,0}) \otimes (f_{0,\delta_1} \otimes \cdots \otimes f_{0,\delta_n}).$$

This identity allows us to work on the ε part and δ part separately. In short, the ε part $f_{\varepsilon_1,0} \otimes \cdots \otimes f_{\varepsilon_n,0}$ now can be approximated by $G_{\sqrt{\varepsilon_1^2+\cdots+\varepsilon_n^2}}$ by invoking Theorem 6. For the δ part, we can iteratively apply the rule

$$f_{0,\delta_1} \otimes f_{0,\delta_2} = f_{0,1-(1-\delta_1)(1-\delta_2)} \tag{11}$$

to obtain $f_{0,\delta_1} \otimes \cdots \otimes f_{0,\delta_n} = f_{0,1-(1-\delta_1)(1-\delta_2)\cdots(1-\delta_n)}$. This rule is best seen via the interesting fact that $f_{0,\delta}$ is the trade-off function of shifted uniform distributions $T(U[0,1], U[\delta, 1+\delta])$.

Now, a central limit theorem for (ε, δ)-DP is just a stone's throw away. In what follows, the privacy parameters ε and δ are arranged in a triangular array $\{(\varepsilon_{ni}, \delta_{ni}) : 1 \le i \le n\}_{n=1}^{\infty}$.

Theorem 7 *Assume*

$$\sum_{i=1}^{n} \varepsilon_{ni}^2 \to \mu^2, \quad \max_{1\le i\le n} \varepsilon_{ni} \to 0, \quad \sum_{i=1}^{n} \delta_{ni} \to \delta, \quad \max_{1\le i\le n} \delta_{ni} \to 0$$

for some nonnegative constants μ, δ as $n \to \infty$. Then, we have

$$f_{\varepsilon_{n1},\delta_{n1}} \otimes \cdots \otimes f_{\varepsilon_{nn},\delta_{nn}} \to G_\mu \otimes f_{0,1-\mathrm{e}^{-\delta}}$$

uniformly over $[0, 1]$ *as* n $\to \infty$.

The proof of this theorem is provided in Appendix D. The assumptions concerning $\{\delta_{ni}\}$ give rise to $1-(1-\delta_{n1})(1-\delta_{n2})\cdots(1-\delta_{nn}) \to 1-\mathrm{e}^{-\delta}$. In general, tensoring with $f_{0,\delta}$ is equivalent to scaling the graph of the trade-off function f toward the origin by a factor of $1-\delta$. This property is specified by the following formula, and we leave its proof to Appendix D:

$$f \otimes f_{0,\delta}(\alpha) = \begin{cases} (1-\delta)\cdot f(\frac{\alpha}{1-\delta}), & 0 \le \alpha \le 1-\delta \\ 0, & 1-\delta \le \alpha \le 1. \end{cases} \tag{12}$$

In particular, $f \otimes f_{0,\delta}$ is symmetric if f is symmetric. Note that Equations (10) and (11) can be deduced by the formula above.

This theorem interprets the privacy level of the composition using Gaussian and uniform distributions. Explicitly, the theorem demonstrates that, based on the released information of the composed mechanism, distinguishing between any neighbouring data sets is at least as hard as distinguishing between the following two bivariate distributions:

$$\mathcal{N}(0,1) \times U[0,1] \text{ versus } \mathcal{N}(\mu,1) \times U[1-\mathrm{e}^{-\delta}, 2-\mathrm{e}^{-\delta}].$$

We note that for small δ, $\mathrm{e}^{-\delta} \approx 1 - \delta$. So $U[1 - \mathrm{e}^{-\delta}, 2 - \mathrm{e}^{-\delta}] \approx U[\delta, 1 + \delta]$.

This approximation of the tensor product $f_{\varepsilon_{n1}, \delta_{n1}} \otimes \cdots \otimes f_{\varepsilon_{nn}, \delta_{nn}}$ using simple distributions is important from the viewpoint of computational complexity. Murtagh and Vadhan (2016) showed that, given a collection of $\{(\varepsilon_i, \delta_i)\}_{i=1}^n$, finding the smallest ε such that $f_{\varepsilon,\delta} \leq f_{\varepsilon_1, \delta_1} \otimes \cdots \otimes f_{\varepsilon_n, \delta_n}$ is #P-hard for any δ (#P is a complexity class that is 'even harder than' NP; see, e.g. Ch. 9. of Arora & Barak, 2009). From the dual perspective (see Section 2.4), this negative result is equivalent to the #P-hardness of evaluating the convex conjugate $(f_{\varepsilon_1, \delta_1} \otimes \cdots \otimes f_{\varepsilon_n,\delta_n})^*$ at any point. For completeness, we remark that Murtagh and Vadhan (2016) provided an FPTAS (an approximation algorithm is called a fully polynomial-time approximation scheme (FPTAS) if its running time is polynomial in both the input size and the inverse of the relative approximation error; see, e.g., Ch. 8. of Vazirani, 2013) to approximately find the smallest ε in $O(n^3)$ time for a *single* δ. In comparison, Theorem 7 offers a *global* approximation of the tensor product in $O(n)$ time using a closed-form expression, subsequently enabling an analytical approximation of the smallest ε for each δ.

That being said, Theorem 7 remains silent on the approximation error in applications with a moderately large number of (ε, δ)-DP mechanisms. Alternatively, we can apply Theorem 5 to obtain a non-asymptotic normal approximation to $f_{\varepsilon_1,0} \otimes \cdots \otimes f_{\varepsilon_n,0}$ and use γ to specify the approximation error. It can be shown that $\gamma = O(1/\sqrt{n})$ under mild conditions (Corollary D.7). This bound, however, is not sharp enough for tight privacy guarantees if n is not too large (note that $1/\sqrt{n} \approx 0.14$ if $n = 50$, for which exact computation is already challenging, if possible at all). Surprisingly, the following theorem establishes a $O(1/n)$ bound, thereby 'beating' the classical Berry–Esseen bound. The proof is given in Appendix D.

Theorem 8 *Fix $\mu > 0$ and let $\varepsilon = \mu/\sqrt{n}$. There is a constant $c > 0$ that only depends on μ satisfying*

$$G_\mu\left(\alpha + \frac{c}{n}\right) - \frac{c}{n} \leq f_{\varepsilon,0}^{\otimes n}(\alpha) \leq G_\mu\left(\alpha - \frac{c}{n}\right) + \frac{c}{n}$$

for all $n \geqslant 1$ *and* $c/n \leq \alpha \leq 1 - c/n$.

As with Theorem 7, this theorem can be extended to approximate DP ($\delta \neq 0$) by making use of the decomposition (10). Our simulation studies suggest that $c \approx 0.1$ for $\mu = 1$, which is best illustrated in the right panel of Figure 5. Despite a fairly small n=10, the difference between G_1 and its target $f_{\varepsilon,0}^{\otimes n}$ is less than 0.013 in the pointwise sense. For completeness, it is worthwhile mentioning that a better approximation can be obtained by using the Edgeworth expansion in place of the central limit theorem (Zheng et al., 2020). Interestingly, our numerical evidence suggests the same $O(1/n)$ rate under inhomogeneous composition, provided that $\varepsilon_1, \ldots, \varepsilon_n$ are roughly the same size. A formal proof, or even a quantitative statement of this observation, constitutes an interesting problem for future investigation.

In closing this section, we highlight some novelties in the proof of Theorem 8. Denoting $p_\varepsilon = \frac{1}{1+\mathrm{e}^\varepsilon}$ and $q_\varepsilon = \frac{\mathrm{e}^\varepsilon}{1+\mathrm{e}^\varepsilon}$, Kairouz et al. (2017) presented a very useful expression (rephrased in our framework):

$$f_{\varepsilon,0}^{\otimes n} = T(B(n, p_\varepsilon), B(n, q_\varepsilon)),$$

where $B(n, p)$ denotes the binomial distribution with n trials and success probability p. However, directly approximating $f_{\varepsilon,0}^{\otimes n}$ through these two binomial distributions is unlikely to yield an $O(1/n)$

bound because the Berry–Esseen bound is rate-optimal for binomial distributions. Our analysis, instead, rests crucially on a certain smoothing effect that comes for free in testing between the two distributions. It is analogous to the continuity correction for normal approximations to binomial probabilities. See the technical details in Appendix D.

4 | AMPLIFYING PRIVACY BY SUBSAMPLING

Subsampling is often used prior to a private mechanism M as a way to *amplify* privacy guarantees. Specifically, we can construct a smaller data set $\tilde{S}$ by flipping a fair coin for each individual in the original data set S to decide whether the individual is included in $\tilde{S}$. This subsampling scheme roughly shrinks the data set by half and, therefore, we would expect that the induced mechanism applied to $\tilde{S}$ is about twice as private as the original mechanism M. Intuitively speaking, this privacy amplification is due to the fact that every individual enjoys perfect privacy if the individual is not included in the resulting data set $\tilde{S}$, which happens with probability 50%.

The claim above was first formalized in Kasiviswanathan et al. (2011) for (ε, δ)-DP. Such a privacy amplification property is, unfortunately, no longer true for the most natural previous relaxations of differential privacy aimed at recovering precise compositions (like concentrated differential privacy (CDP) (Bun & Steinke, 2016; Dwork & Rothblum, 2016)). Further modifications such as truncated CDP (Bun et al., 2018a) have been introduced primarily to remedy this deficiency of CDP—but at the cost of extra complexity in the definition. Other relaxations like Rényi differential privacy (Mironov, 2017) can be shown to satisfy a form of privacy amplification by subsampling, but both the analysis and the statement are complex (Wang et al., 2018).

In this section, we show that these obstacles can be overcome by our hypothesis testing based relaxation of differential privacy. Explicitly, our main result is a simple, general and easy-to-interpret subsampling theorem for f-DP. Somewhat surprisingly, our theorem significantly improves on the classical subsampling theorem for privacy amplification in the (ε, δ)-DP framework (Ullman, 2017). Note that this classical theorem continues to use (ε, δ)-DP to characterize the subsampled mechanism. However, (ε, δ)-DP is simply not expressive enough to capture the amplification of privacy.

4.1 | A subsampling theorem

Given an integer $1 \le m \le n$ and a data set S of n individuals, let $\mathtt{Sample}_m(S)$ be a subset of S that is chosen uniformly at random among all the m-sized subsets of S. For a mechanism M defined on X^m, we call $M(\mathtt{Sample}_m(S))$ the subsampled mechanism, which takes as input an n-sized data set. Formally, we use $M \circ \mathtt{Sample}_m$ to denote this subsampled mechanism. To clear up any confusion, note that intermediate result $\mathtt{Sample}_m(S)$ is not released and, in particular, this is different from the composition in Section 3.

In brief, our main theorem shows that the privacy bound of the subsampled mechanism in the f-DP framework is given by an operator acting on trade-off functions. To introduce this operator, write the convex combination $f_p := pf + (1-p)\,\mathrm{Id}$ for $0 \le p \le 1$, where $\mathrm{Id}(x) = 1 - x$. Note that the trade-off function f_p is asymmetric in general.

Definition 6 For any $0 \le p \le 1$, define the operator C_p acting on trade-off functions as

$$C_p(f) := \min\{f_p, f_p^{-1}\}^{**}.$$

We call C_p the p-sampling operator.

Above, the inverse f_p^{-1} is defined in Equation (3). The biconjugate $\min\{f_p, f_p^{-1}\}^{**}$ is derived by applying the conjugate as defined in Equation (7) twice to $\min\{f_p, f_p^{-1}\}$. For the moment, take for granted the fact that $C_p(f)$ is a symmetric trade-off function.

Now, we present the main theorem of this section. Section 4.2 is devoted to proving this result.

Theorem 9 *If* M *is* f*-DP on* X^{m}*, then the subsampled mechanism* $\mathrm{M} \circ \mathtt{Sample}_m$ *is* $\mathrm{C_p(f)}$*-DP on* X^{n}*, where the sampling ratio* $p = \frac{m}{n}$.

Appreciating this theorem calls for a better understanding of the operator C_p. In effect, C_p performs a two-step transformation: symmetrization (taking the minimum of f_p and its inverse f_p^{-1}) and convexification (taking the largest convex lower envelope of $\min\{f_p, f_p^{-1}\}$). The convexification step is seen from convex analysis that the biconjugate h^{**} of any function h is the greatest convex lower bound of h. As such, $C_p(f)$ is convex and, with a bit more analysis, Proposition 1 ensures that $C_p(f)$ is indeed a trade-off function. As an aside, $C_p(f) \leq \min\{f_p, f_p^{-1}\} \leq f_p$. See Figure 6 for a graphical illustration.

Next, the following facts concerning the p-sampling operator qualitatively illustrate this privacy amplification phenomenon.

1. If $0 \leq p \leq q \leq 1$ and f is symmetric, we have $f = C_1(f) \leq C_q(f) \leq C_p(f) \leq C_0(f) = \mathrm{Id}$. That is, as the sampling ratio declines from 1 to 0, the privacy guarantee interpolates monotonically between the original f and the perfect privacy guarantee Id. This monotonicity follows from the fact that $g \geq h$ is equivalent to $g^{-1} \geq h^{-1}$ for any trade-off functions g and h.

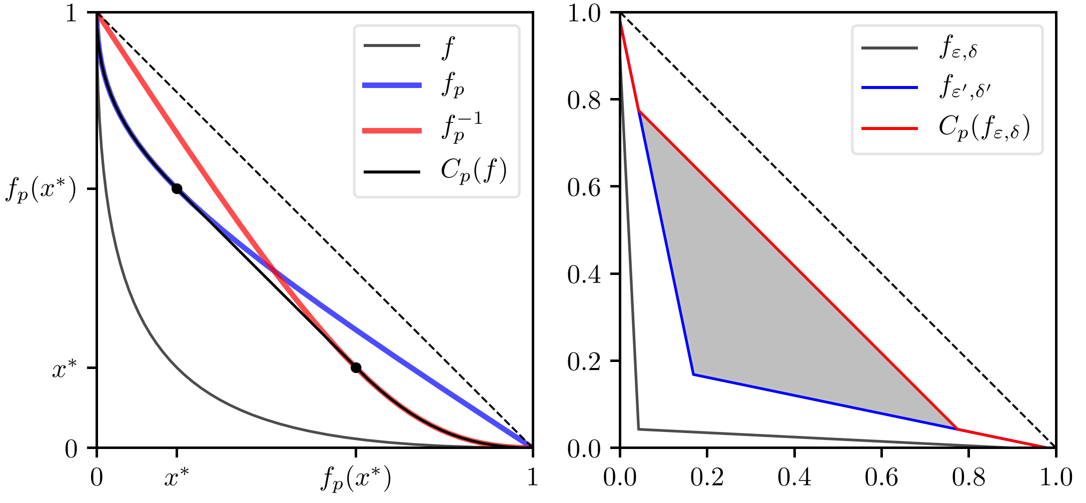

FIGURE 6 The action of C_p. Left panel: $f = G_{1.8}$, $p = 0.35$. Right panel: $\varepsilon = 3$, $\delta = 0.1$, $p = 0.2$. The subsampling Theorem 9 results in a significantly tighter trade-off function compared to the classical theorem for (ε, δ)-DP [Colour figure can be viewed at wileyonlinelibrary.com]

2. If two trade-off functions f and g satisfy $f \geq g$, then $C_p(f) \geq C_p(g)$. This means that if a mechanism is more private than the other, using the same sampling ratio, the subsampled mechanism of the former remains more private than that of the latter, at least in terms of lower bounds.
3. For any $0 \leq p \leq 1$, $C_p(\mathrm{Id}) = \mathrm{Id}$. That is, perfect privacy remains perfect privacy with subsampling.

Explicitly, we provide a formula to calculate $C_p(f)$ for a symmetric trade-off function f. Letting x^* be the unique fixed point of f, that is $f(x^*) = x^*$, we have

$$C_p(f)(x) = \begin{cases} f_p(x), & x \in [0, x^*] \\ x^* + f_p(x^*) - x, & x \in [x^*, f_p(x^*)] \\ f_p^{-1}(x), & x \in [f_p(x^*), 1]. \end{cases} \tag{13}$$

This expression is almost self-evident from the left panel of Figure 6. Nevertheless, a proof of this formula is given in Appendix E. This formula, together with Theorem 9, allows us to get a closed-form characterization of the privacy amplification for (ε, δ)-DP.

Corollary 3 *If M is (ε, δ)-DP on X^m, then the subsampled mechanism $M \circ \mathtt{Sample}_m$ is $C_p(f_{\varepsilon,\delta})$-DP on X^n, where*

$$C_p(f_{\varepsilon,\delta})(\alpha) = \max\left\{ f_{\varepsilon',\delta'}(\alpha), 1 - p\delta - p\frac{\mathrm{e}^{\varepsilon} - 1}{\mathrm{e}^{\varepsilon} + 1} - \alpha \right\}. \tag{14}$$

Above, $\varepsilon' = \log(1 - p + p\mathrm{e}^{\varepsilon})$, $\delta' = p\delta$ and $p = \frac{m}{n}$.

For comparison, we now present the existing bound on the privacy amplification by subsampling for (ε,δ)-DP. To be self-contained, Appendix E gives a proof of this result, which primarily follows Ullman (2017).

Lemma 2 (Ullman, 2017). *If M is (ε, δ)-DP, then $M \circ \mathtt{Sample}_m$ is (ε', δ')-DP with ε' and δ' defined in Corollary 3.*

Using the language of the f-DP framework, Lemma 2 states that $M \circ \mathtt{Sample}_m$ is $f_{\varepsilon',\delta'}$-DP. Corollary 3 improves on Lemma 2 because, as is clear from Equation (14), $C_p(f_{\varepsilon,\delta}) \geq f_{\varepsilon',\delta'}$. The right panel of Figure 6 illustrates Lemma 2 and our Corollary 3 for $\varepsilon = 3$, $\delta = 0.1$ and $p = 0.2$. In effect, the improvement is captured by the shaded triangle enclosed by $C_p(f_{\varepsilon,\delta})$ and $f_{\varepsilon',\delta'}$, revealing that the minimal sum of type I and type II errors in distinguishing two neighbouring data sets with subsampling can be significantly lower than the prediction of Lemma 2. This gain is only made possible by the flexibility of trade-off functions in the sense that $C_p(f_{\varepsilon,\delta})$ *cannot* be expressed within the (ε, δ)-DP framework. The unavoidable loss in the (ε, δ)-DP representation of the subsampled mechanism is compounded when analysing the composition of many private mechanisms.

In the next subsection, we prove Theorem 9 by making use of Lemma 2. Its proof implies that Theorem 9 holds for any subsampling scheme for which Lemma 2 is true. In particular, it holds for the subsampling scheme described at the beginning of this section, that is, independent coin flips for every data item.

4.2 Proof of the subsampling theorem

The proof strategy is as follows. First, we convert the f-DP guarantee into an infinite collection of (ε, δ)-DP guarantees by taking a dual perspective that is enabled by Proposition 6. Next, by applying the classical subsampling theorem (that is, Lemma 2) to these (ε, δ)-DP guarantees, we conclude that the subsampled mechanism satisfies a new infinite collection of (ε, δ)-DP guarantees. Finally, Proposition 5 allows us to convert these new privacy guarantees back into an $\tilde{f}$-DP guarantee, where $\tilde{f}$ can be shown to coincide with $C_p(f)$.

Proof of Theorem 9 Provided that M is f-DP, from Proposition 6 it follows that M is $(\varepsilon, \delta(\varepsilon))$-DP with $\delta(\varepsilon) = 1 + f^*(-e^{\varepsilon})$ for all $\varepsilon \geqslant 0$. Making use of Lemma 2, the subsampled mechanism $M \circ \texttt{Sample}_m$ satisfies the following collection of (ε', δ')-DP guarantees for all $\varepsilon \geqslant 0$:

$$\varepsilon' = \log(1 - p + pe^{\varepsilon}), \quad \delta' = p\left(1 + f^*(-e^{\varepsilon})\right).$$

Eliminating the variable ε from the two parametric equations above, we can relate ε' to δ' using

$$\delta' = 1 + f_p^*(-e^{\varepsilon'}), \tag{15}$$

which is proved in Appendix E. The remainder of the proof is devoted to showing that (ε', δ')-DP guarantees for all $\varepsilon' \geq 0$ is equivalent to the $C_p(f)$-DP guarantee.

At first glance, Equation (15) seems to enable the use of Proposition 6. Unfortunately, that would be invalid because f_p is asymmetric. To this end, we need to extend Proposition 6 to general trade-off functions. To avoid conflicting notation, let g be a generic trade-off function, not necessarily symmetric. Denote by $\bar{x}$ be the smallest point such that $g'(x) = -1$, that is, $\bar{x} = \inf\{x \in [0, 1]: g'(x) = -1\}$. As a special instance of Proposition E.1 in the appendix, the following result serves our purpose.

Proposition 8 *If $g(\bar{x}) \geqslant \bar{x}$ and a mechanism M is $(\varepsilon, 1 + g^*(-e^{\varepsilon}))$-DP for all $\varepsilon \geqslant 0$, then M is $\min\{g, g^{-1}\}^{**}$-DP.*

The proof of the present theorem would be complete if Proposition 8 can be applied to the collection of privacy guarantees in Equation (15) for f_p. To use Proposition 8, it suffices to verify the condition $f_p(\bar{x}) \geqslant \bar{x}$ where $\bar{x}$ is the smallest point such that $f_p'(x) = -1$. Let x^* be the (unique) fixed point of f. To this end, we collect a few simple facts:

- First, $f'(x^*) = -1$. This is because the graph of f is symmetric with respect to the 45° line passing through the origin.
- Second, $\bar{x} \leq x^*$. This is because $f_p'(x^*) = pf'(x^*) + (1 - p)\,\mathrm{Id}'(x^*) = -1$ and, by definition, $\bar{x}$ can only be smaller.

With these facts in place, we get

$$f_p(\overline{x}) \geqslant f_p(x^*) \geqslant f(x^*) = x^* \geqslant \overline{x}$$

by recognizing that f_p is decreasing and $f_p \geq f$. Hence, the proof is complete.

5 | APPLICATION: PRIVACY ANALYSIS OF STOCHASTIC GRADIENT DESCENT

One of the most important algorithms in machine learning and optimization is stochastic gradient descent (SGD). This is an iterative optimization method used to train a wide variety of models, for example, deep neural networks. SGD has also served as an important benchmark in the development of private optimization: as an iterative algorithm, the tightness of its privacy analysis crucially depends on the tightness with which composition can be accounted for. The analysis also crucially requires a privacy amplification by subsampling argument.

The first asymptotically optimal analysis of differentially private SGD was given by Bassily et al. (2014). Because of the inherent limits of (ε, δ)-DP, however, this analysis stops short of giving meaningful privacy bounds for realistically sized data sets. This is in part what motivated the development of divergence based relaxations of differential privacy. Unfortunately, these relaxations cannot be directly applied to the analysis of SGD due to the lack of a privacy amplification by subsampling theorem. In response, Abadi et al. (2016) circumvented this challenge by developing the moments accountant—a numeric technique tailored specifically to repeated application of subsampling, followed by a Gaussian mechanism—to give privacy bounds for SGD that are strong enough to give non-trivial guarantees when training deep neural networks on real data sets. But this analysis is ad-hoc in the sense that it uses a tool designed specifically for the analysis of SGD.

In this section, we use the general tools we have developed so far to give a simple and improved analysis of the privacy of SGD. In particular, the analysis rests crucially on the compositional and subsampling properties of f-DP.

5.1 | Stochastic gradient descent and its privacy analysis

Letting $S = (x_1, \ldots, x_n)$ denote the data set, consider minimizing the empirical risk

$$\frac{1}{n}\sum_{i=1}^{n} L(\theta, x_i)$$

over the parameter θ, where $L(\theta, x_i)$ denotes a loss function. At iteration t, a set I_t of size m is selected uniformly at random from $\{1, 2, \ldots, n\}$. Taking learning rate η_t, SGD seeks to minimize the empirical risk by running

$$\theta_{t+1} = \theta_t - \eta_t \cdot \frac{1}{m}\sum_{i \in I_t} \nabla_\theta L(\theta_t, x_i)$$

from an initial point θ_0.

Algorithm 1 NoisySGD

1: **Input:** Dataset $S = (x_1, \ldots, x_n)$, loss function $L(\theta, x)$.
Parameters: initial state θ_0, learning rate η_t, batch size m, time horizon T, noise scale σ, gradient norm bound C.
2: **for** $t = 1, \ldots, T$ **do**
3: **Subsampling:**
Take a uniformly random subsample $I_t \subseteq \{1, \ldots, n\}$ with batch size m ▷ $\texttt{Sample}_m$ in Section 4
4: **for** $i \in I_t$ **do**
5: **Compute gradient:**
$v_t^{(i)} \leftarrow \nabla_\theta L(\theta_t, x_i)$
6: **Clip gradient:**
$\bar{v}_t^{(i)} \leftarrow v_t^{(i)} / \max\left\{1, \|v_t^{(i)}\|_2 / C\right\}$
7: **Average, perturb, and descend:**
$\theta_{t+1} \leftarrow \theta_t - \eta_t \left(\frac{1}{m} \sum_i \bar{v}_t^{(i)} + \mathcal{N}(0, \frac{4\sigma^2 C^2}{m^2} I) \right)$ ▷ I is an identity matrix
8: **Output** θ_T

A private variant of this optimization algorithm is described in Algorithm 1. We refer to this private algorithm as NoisySGD, which can be viewed as a repeated composition of Gaussian mechanisms operating on subsampled data sets. To analyse the privacy of NoisySGD, we start by building up the privacy properties from the inner loop. Let V be the vector space where parameter θ lives in and $M : X^m \times V \to V$ be the mechanism that executes lines 4–7 in Algorithm 1. Here m denotes the batch size. In effect, what M does in iteration t can be expressed as

$$M(S_{I_t}, \theta_t) = \theta_{t+1},$$

where S_{I_t} is the subset of the data set S indexed by I_t. Next, we turn to the analysis of the subsampling step (line 3) and use $\widetilde{M}$ to denote its composition with M, that is, $\widetilde{M} = M \circ \texttt{Sample}_m$. Taken together, $\widetilde{M}$ executes lines 3–7 and maps from $X^n \times V$ to V.

The mechanism we are ultimately interested in

$$\begin{aligned} \texttt{NoisySGD}: X^n &\to V \times V \times \cdots \times V \\ S &\mapsto (\theta_1, \theta_2, \ldots, \theta_T) \end{aligned}$$

is simply the composition of T copies of $\widetilde{M}$. To see this fact, note that the trajectory $(\theta_1, \theta_2, ..., \theta_T)$ is obtained by iteratively running

$$\theta_{j+1} = \widetilde{M}(S, \theta_j)$$

for $j = 0, ..., T - 1$. Let M be f-DP. Straightforwardly, $\widetilde{M}$ is $C_{m/n}(f)$-DP by Theorem 9. Then, from the composition theorem (Theorem 4), we can readily prove that NoisySGD is $C_{m/n}(f)^{\otimes T}$-DP.
Hence, it suffices to give a bound on the privacy of M. For simplicity, we now focus on a single step and drop the subscript t. Recognizing that changing one of the m data points only affects one $v^{(i)}$, the sensitivity of $\frac{1}{m} \sum_i \bar{v}_t^{(i)}$ is at most $\frac{2C}{m}$ due to the clipping operation. Making use of Theorem 1, adding Gaussian noise $N(0, \sigma^2 \cdot \frac{4C^2}{m^2} I)$ to the average gradient renders this step $\frac{1}{\sigma}$-GDP. Since that the gradient update following the gradient averaging step is deterministic, we conclude that M satisfies $\frac{1}{\sigma}$-GDP.

In summary, the discussion above has proved the following theorem:

Theorem 10 *Algorithm 1 is $C_{m/n}(G_{\sigma^{-1}})^{\otimes T}$-DP.*

To clear up any confusion, we remark that this $C_{m/n}(G_{\sigma^{-1}})^{\otimes T}$-DP mechanism does not release the subsampled indices.

The use of Theorem 10 relies on an efficient evaluation of $C_{m/n}(G_{\sigma^{-1}})^{\otimes T}$. Our central limit theorems provide an analytical approach to approximating this tensor product and the approximation is accurate for large T. The next two subsections present two such results, corresponding to our two central limit theorems (Theorems 5 and 6), respectively. An asymptotic privacy analysis of `NoisySGD` is given in Section 5.2 by developing a general limit theorem for composition of subsampled mechanisms, and an illustration of this result is shown in Figure 7. A Berry–Esseen type analysis is developed in Section 5.3. The implementation of our privacy analysis of `NoisySGD` is available in the `TensorFlow privacy` package (https://github.com/tensorflow/privacy); see details in https://github.com/tensorflow/privacy/blob/master/tensorflow_privacy/privacy/analysis/gdp_accountant.py.

5.2 | Asymptotic privacy analysis

In this subsection, we first consider the limit of $C_p(f)^{\otimes T}$ for a general trade-off function f, then plug in $f = G_{\sigma^{-1}}$ for the analysis of `NoisySGD`. The more general approach is useful for analysing other iterative algorithms.

Recall from Section 4 that a p-subsampled f-DP mechanism is $C_p(f)$-DP, where $C_p(f)$ is defined as

$$C_p(f)(x) = \begin{cases} f_p(x), & x \in [0, x^*] \\ x^* + f_p(x^*) - x, & x \in [x^*, f_p(x^*)] \\ f_p^{-1}(x), & x \in [f_p(x^*), 1], \end{cases}$$

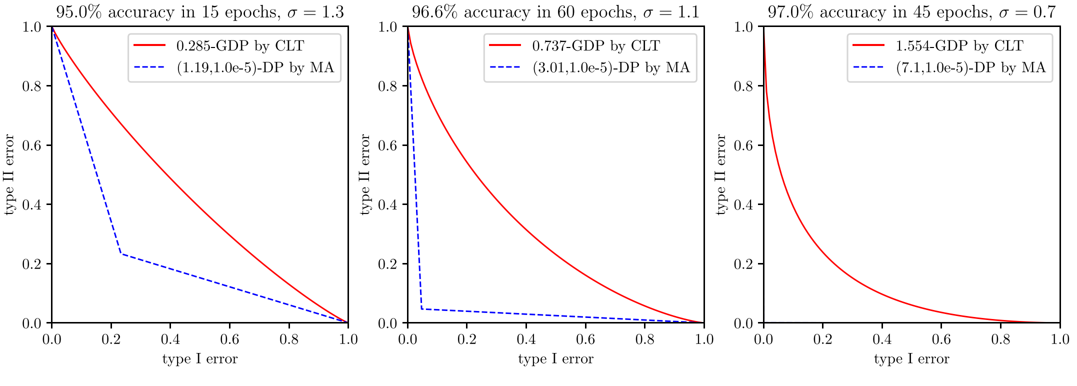

FIGURE 7 Comparison of the Gaussian differential privacy bounds derived from our method, and the (ε, δ)-DP bounds derived using the moments accountant (Abadi et al., 2016), which is essentially based on Rényi differential privacy (Mironov, 2017). All three experiments run Algorithm 1 on the entire MNIST data set with $n = 60{,}000$ data points, batch size $m = 256$, learning rates η_t set to 0.25, 0.15 and 0.25, respectively, and clipping thresholds C set to 1.5, 1.0 and 1.5, respectively. The red lines are obtained via Corollary 4, while the blue dashed lines are produced by the tensorflow/privacy library. See https://github.com/tensorflow/privacy for the details of the setting and more experiments in follow-up work (Bu et al., 2019) [Colour figure can be viewed at wileyonlinelibrary.com]

where x^* is the unique fixed point of f. We will let the sampling fraction p tend to 0 as T approaches infinity. In the following theorem, a_+^2 is a short-hand for $(\max\{a, 0\})^2$.

Theorem 11 *Suppose f is a symmetric trade-off function such that $f(0) = 1$ and $\int_0^1 (f'(x)+1)^4 \, dx < +\infty$. Furthermore, assume $p\sqrt{T} \to p_0$ as $T \to \infty$ for some constant $p_0 > 0$. Then we have the uniform convergence*

$$C_p(f)^{\otimes T} \to G_{p_0\sqrt{2\chi_+^2(f)}}$$

as $T \to \infty$, *where*

$$\chi_+^2(f) = \int_0^1 (|f'(x)| - 1)_+^2 \, dx.$$

The proof is deferred to Appendix F. This theorem has implications for the design of iterative private mechanisms involving subsampling as a subroutine. One way to bound the privacy of such a mechanism is to let the sampling ratio p go to zero as the total number of iterations T goes to infinity. The theorem says that the correct scaling between the two values is $p \sim 1/\sqrt{T}$ and, furthermore, gives an explicit form of the limit.

In order to analyse `NoisySGD`, we need to compute the quantity $\chi_+^2(G_\mu)$. This can be done by directly working with its definition. In Appendix F, we provide a different approach by relating $\chi_+^2(f)$ to χ^2-divergence.

Lemma 3 *We have*

$$\chi_+^2(G_\mu) = e^{\mu^2} \cdot \Phi(3\mu/2) + 3\Phi(-\mu/2) - 2.$$

When using SGD to train large models, we typically perform a very large number of iterations, so it is reasonable to consider the parameter regime in which $n \to \infty$, $T \to \infty$. The batch size can also vary with these quantities. The following result is a direct consequence of Theorems 10 and 11 and Lemma 3.

Corollary 4 *If $m\sqrt{T}/n \to c$ for a constant $c > 0$, then* `NoisySGD` *is asymptotically μ-GDP with*

$$\mu = \sqrt{2}c \cdot \sqrt{e^{\sigma^{-2}} \cdot \Phi(1.5\sigma^{-1}) + 3\Phi(-0.5\sigma^{-1}) - 2}.$$

The condition required in this theorem is more general than that required in the analysis of private SGD by Bassily et al. (2014), which assumes $m = 1$ and $T = O(n^2)$. Moreover, we note that $\frac{m}{n} \cdot \sqrt{T}$ in deep learning research is generally quite small. The convention in this literature is to reparameterize the number of gradient steps T by the number of 'epochs' E, which is the number of sweeps of the entire data set. The relationship between these parameters is that $E = Tm/n$. In this reparameterization, our assumption is that $Em/n \to c^2$. Concretely, the AlexNet (Krizhevsky et al., 2012) sets the parameters as $m = 128$, $E \approx 90$ on the ILSVRC-2010 data set with $n \approx 1.2 \times 10^6$, leading to $Em/n < 0.01$. Many other prominent implementations also lead to a small value of Em/n. See the webpage of the Gluon CV Toolkit

(He et al., 2018; Zhang et al., 2019) for a collection of such hyperparameters in computer vision tasks.

5.3 | A Berry–Esseen privacy bound

Now, we apply the Berry–Esseen style central limit theorem (Theorem 5) to the privacy analysis of `NoisySGD`, highlighting the advantage of giving sharp privacy guarantees. However, the shortcoming is that the expressions that it yields are more unwieldy: they are computer evaluable, so usable in implementations, but do not admit simple closed forms.

The individual components in Theorem 5 have have the form $C_p(G_\mu)$ with $p = m/n$, $\mu = \sigma^{-1}$. It suffices to evaluate the moment functionals on $C_p(G_\mu)$. This is done in the following lemma, with its proof given in Appendix F.

Lemma 4 *Let* $Z(x) = \log(p \cdot \mathrm{e}^{\mu x - \mu^2/2} + 1 - p)$ *and* $\varphi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$ *be the density of the standard normal distribution. Then*

$$\begin{aligned}
\mathrm{kl}\left(C_p(G_\mu)\right) &= p\int_{\mu/2}^{+\infty} Z(x)\cdot\left(\varphi(x-\mu)-\varphi(x)\right)\,\mathrm{d}x\\
\kappa_2\left(C_p(G_\mu)\right) &= \int_{\mu/2}^{+\infty} Z^2(x)\cdot\left(p\varphi(x-\mu)+(2-p)\varphi(x)\right)\,\mathrm{d}x\\
\overline{\kappa}_3\left(C_p(G_\mu)\right) &= \int_{\mu/2}^{+\infty}\left|Z(x)-\mathrm{kl}\left(C_p(G_\mu)\right)\right|^3\cdot(p\varphi(x-\mu)+(1-p)\varphi(x))\,\mathrm{d}x\\
&\quad+\int_{\mu/2}^{+\infty}\left|Z(x)+\mathrm{kl}\left(C_p(G_\mu)\right)\right|^3\cdot\varphi(x)\,\mathrm{d}x.
\end{aligned}$$

By plugging these expressions into Theorem 5, we get

Corollary 5 *Let* $p = m/n$, $\mu = \sigma^{-1}$ *and*

$$\tilde{\mu} = \frac{2\sqrt{T}\cdot\mathrm{kl}\left(C_p(G_\mu)\right)}{\sqrt{\kappa_2\left(C_p(G_\mu)\right) - \mathrm{kl}^2\left(C_p(G_\mu)\right)}}, \quad \gamma = \frac{0.56}{\sqrt{T}}\cdot\frac{\overline{\kappa}_3\left(C_p(G_\mu)\right)}{\left(\kappa_2\left(C_p(G_\mu)\right) - \mathrm{kl}^2\left(C_p(G_\mu)\right)\right)^{\frac{3}{2}}}.$$

Then, `NoisySGD` *is f-DP with* $f(\alpha) = \max\{G_{\tilde{\mu}}(\alpha+\gamma)-\gamma, 0\}$.

We remark that $G_{\tilde{\mu}}$ can be set to 0 in $(1, +\infty)$ so that f is well defined for $\alpha > 1-\gamma$.

6 | DISCUSSION

In this paper, we have introduced a new framework for private data analysis that we refer to as f-differential privacy, which generalizes (ε, δ)-DP and has a number of attractive properties that escape the difficulties of prior work. This new privacy definition uses trade-off functions of hypothesis testing as a measure of indistinguishability of two neighbouring data sets rather than a few parameters as in prior differential privacy relaxations. Our *f*-DP retains an interpretable hypothesis testing semantics and is expressive enough to losslessly reason about composition, post-processing and group privacy by virtue of the informativeness of trade-off functions. Moreover, *f*-DP admits a central limit theorem that identifies a simple and single-parameter family of privacy definitions as focal: Gaussian differential privacy. Precisely, all hypothesis testing based definitions of privacy converge to Gaussian differential privacy in the limit under composition, which implies that Gaussian differential privacy is the unique such definition that can tightly handle composition. The central limit theorem and its Berry–Esseen variant give a tractable analytical approach to tightly analysing the privacy cost of iterative methods such as SGD. Notably, *f*-DP is *dual* to (ε, δ)-DP in a constructive sense, which gives the ability to import results proven for (ε, δ)-DP. This powerful perspective allows us to obtain an easy-to-use privacy amplification by subsampling theorem for *f*-DP, which in particular significantly improves on the state-of-the-art counterpart in the (ε, δ)-DP setting.

We see several promising directions for future work using and extending the *f*-DP framework. First, Theorem 8 can possibly be extended to the inhomogeneous case where trade-off functions are different from each other in the composition. Such an extension would allow us to apply the central limit theorem for privacy approximation with strong finite-sample guarantees to a broader range of problems. Second, it would be of interest to investigate whether the privacy guarantee of the subsampled mechanism in Theorem 9 can be improved for some trade-off functions. Notably, we have shown in Appendix E that this bound is tight if the trade-off function $f = 0$, that is, the original mechanism is blatantly non-private. Third, the notion of *f*-DP naturally has a *local* realization where the obfuscation of the sensitive information is applied at the individual record level. In this setting, what are the fundamental limits of estimation with local *f*-DP guarantees (Duchi et al., 2018)? In light of Duchi and Ruan (2018), what is the correct complexity measure in local *f*-DP estimation? If it is not the Fisher information, can we identify an alternative to the Fisher information for some class of trade-off functions? Moreover, we recognize that an adversary in differentially private learning may set different pairs of target type I and type II errors. For example, an adversary that attempts to control type I and II errors at 10% and 10%, respectively, can behave very differently from one who aims to control the two errors at 0.1% and 99%, respectively. An important question is to address the trade-offs between resources such as privacy and statistical efficiency and target type I and type II errors in the framework of *f*-DP.

Finally, we wish to remark that *f*-DP can possibly offer a mathematically tractable and flexible framework for minimax estimation under privacy constraints (see, for example, Bun et al., 2018b; Cai et al., 2019; Dwork et al., 2015). Concretely, given a candidate estimator satisfying (ε, δ)-DP appearing in the upper bound and a possibly loose lower bound under the (ε, δ)-DP constraint, we can replace the (ε, δ)-DP constraint by the *f*-DP constraint where f is the tightest trade-off function characterizing the estimation procedure. As is clear, the *f*-DP constraint is more stringent than the (ε, δ)-DP constraint by recognizing the primal-dual conversion (see Proposition 6). While the upper bound remains the same as the estimator continues to satisfy the new privacy constraint, the lower bound can be possibly improved due to a more stringent constraint. It would be of great interest to investigate to what extent this *f*-DP based approach can reduce the gap between upper and lower bounds minimax estimation under privacy constraints.

Ultimately, the test of a privacy definition lies not just in its power and semantics, but also in its ability to usefully analyse diverse algorithms. In this paper, we have given convincing evidence that *f*-DP is up to the task. We leave the practical evaluation of this new privacy definition to future work.

7 | SUPPLEMENTAL MATERIALS

Due to space constraints, we have relegated proofs of theorems and other technical details to the on-line appendices Appendix A–Appendix F in the Supplement to 'Gaussian Differential Privacy'. Python code for analysing the privacy loss of SGD in the *f*-DP framework is available at https://github.com/tensorflow/privacy/blob/master/tensorflow_privacy/privacy/analysis/gdp_accountant.py.

ACKNOWLEDGEMENTS

We thank the associate editor and the referees for constructive comments that improved the presentation of the paper. This work was supported in part by NSF grants CAREER DMS-1847415, CCF-1763314 and CNS-1253345, by the Sloan Foundation, and by a sub-contract for the DARPA Brandeis program.

ORCID

Weijie J. Su http://orcid.org/0000-0003-1787-1219

REFERENCES

Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K. et al. (2016) Deep learning with differential privacy. In: *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. ACM, pp. 308–318.

Abowd, J.M. (2018) The US Census Bureau adopts differential privacy. In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. ACM, pp. 2867.

Apple, D.P.T. (2017) Learning with privacy at scale. Technical report, Apple.

Arora, S. & Barak, B. (2009) *Computational complexity: A modern approach*. Cambridge: Cambridge University Press.

Balle, B. & Wang, Y.-X. (2018) Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. *arXiv preprint* arXiv:1805.06530.

Balle, B., Barthe, G. & Gaboardi, M. (2018) Privacy amplification by subsampling: Tight analyses via couplings and divergences. In: *Advances in neural information processing systems*, pp. 6280–6290.

Barbaro, M. & Zeller, T. (2006) A face is exposed for AOL searcher no. 4417749. *The New York times*, August Available from: http://select.nytimes.com/gst/abstract.html?res=F10612FC345B0C7A8CDDA10894DE404482

Bassily, R., Smith, A. & Thakurta, A. (2014) Private empirical risk minimization: Efficient algorithms and tight error bounds. In: *2014 IEEE 55th annual symposium on foundations of computer science*. IEEE, pp. 464–473.

Blackwell, D. (1950) Comparison of experiments. Technical report, Howard University Washington United States.

Bu, Z., Dong, J., Long, Q. & Su, W.J. (2019) Deep learning with Gaussian differential privacy. *arXiv preprint* arXiv:1911.11607.

Bun, M. & Steinke, T. (2016) Concentrated differential privacy: Simplifications, extensions, and lower bounds. In: *Theory of cryptography conference*. Springer, pp. 635–658.

Bun, M., Dwork, C., Rothblum, G.N. & Steinke, T. (2018a) Composable and versatile privacy via truncated CDP. In: *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. ACM, pp. 74–86.

Bun, M., Ullman, J. & Vadhan, S. (2018b) Fingerprinting codes and the price of approximate differential privacy. *SIAM Journal on Computing*, 47(5), 1888–1938.

Cai, T.T., Wang, Y. & Zhang, L. (2019) The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *arXiv preprint* arXiv:1902.04495.

Ding, B., Kulkarni, J. & Yekhanin, S. (2017) Collecting telemetry data privately. In: *Proceedings of advances in neural information processing systems 30 (NIPS 2017)*.

Duchi, J.C. & Ruan, F. (2018) The right complexity measure in locally private estimation: It is not the fisher information. *arXiv preprint* arXiv:1806.05756.

Duchi, J.C., Jordan, M.I. & Wainwright, M.J. (2018) Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521), 182–201.

Dwork, C. & Rothblum, G.N. (2016) Concentrated differential privacy. *arXiv preprint* arXiv:1603.01887.

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I. & Naor, M. (2006a) Our data, ourselves: Privacy via distributed noise generation. In: *Annual international conference on the theory and applications of cryptographic techniques*. Springer, pp. 486–503.

Dwork, C., McSherry, F., Nissim, K. & Smith, A. (2006b) Calibrating noise to sensitivity in private data analysis. In: *Theory of cryptography conference*. Springer, pp. 265–284.

Dwork, C., Rothblum, G.N. & Vadhan, S. (2010) Boosting and differential privacy. In: *Foundations of computer science (FOCS), 2010 51st annual IEEE symposium on*. IEEE, pp. 51–60.

Dwork, C., Smith, A., Steinke, T., Ullman, J. & Vadhan, S. (2015) Robust traceability from trace amounts. In: *2015 IEEE 56th annual symposium on foundations of computer science*. IEEE, pp. 650–669.

Erlingsson, Ú. Pihur, V. & Korolova, A. (2014) Rappor: Randomized aggregatable privacy-preserving ordinal response. In: *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. ACM, pp. 1054–1067.

He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J. & Li, M. (2018). Bag of tricks for image classification with convolutional neural networks. *arXiv preprint* arXiv:1812.01187.

Holmes, S. (2019) Challenges in the analyses of multidomain longitudinal data: Layered solutions. 3rd Workshop on statistical and algorithmic challenges in microbiome data analysis.

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J. et al. (2008) Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*, 4(8), e1000167.

Kairouz, P., Oh, S. & Viswanath, P. (2017) The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6), 4037–4049.

Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Raskhodnikova, S. & Smith, A. (2011) What can we learn privately? *SIAM Journal on Computing*, 40(3), 793–826.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105.

LeCun, Y. & Cortes, C. (2010) MNIST handwritten digit database. Available from: http://yann.lecun.com/exdb/mnist/

Mironov, I. (2017) Rényi differential privacy. In: *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, pp. 263–275.

Murtagh, J. & Vadhan, S. (2016) The complexity of computing the optimal composition of differential privacy. In: *Theory of cryptography conference*. Springer, pp. 157–175.

Narayanan, A. & Shmatikov, V. (2008) Robust de-anonymization of large sparse datasets. In: *2008 IEEE symposium on security and privacy*. IEEE, pp. 111–125.

Raginsky, M. (2011) Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments. In: *2011 IEEE international symposium on information theory proceedings*. IEEE, pp. 1220–1224.

Sommer, D., Meiser, S. & Mohammadi, E. (2018) Privacy loss classes: The central limit theorem in differential privacy.

Ullman, J. (2017) Cs7880: Rigorous approaches to data privacy, Spring 2017. Available from: http://www.ccs.neu.edu/home/jullman/PrivacyS17/HW1sol.pdf

Vazirani, V.V. (2013) *Approximation algorithms*. Berlin/Heidelberg: Springer Science & Business Media.

Wang, Y.-X., Balle, B. & Kasiviswanathan, S. (2018) Subsampled Rényi differential privacy and analytical moments accountant. *arXiv preprint* arXiv:1808.00087.

Wasserman, L. & Zhou, S. (2010) A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489), 375–389.

Zhang, Z., He, T., Zhang, H., Zhang, Z., Xie, J. & Li, M. (2019) Bag of freebies for training object detection neural networks. *arXiv preprint* arXiv:1902.04103.

Zheng, Q., Dong, J., Long, Q. & Su, W.J. (2020) Sharp composition bounds for Gaussian differential privacy via Edgeworth expansion. *arXiv preprint* arXiv:2003.04493.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Dong J, Roth A, Su WJ. Gaussian differential privacy. *J R Stat Soc Series B*. 2022;84:3–54. https://doi.org/10.1111/rssb.12454

DOI: 10.1111/rssb.12455

DISCUSSION CONTRIBUTIONS

Proposer of the vote of thanks to Dong *et al.* and contribution to the Discussion of 'Gaussian Differential Privacy'

Borja Balle

DeepMind, London, UK

Correspondence
Borja Balle, DeepMind, London, UK
Email: borja.balle@gmail.com

I wholeheartedly welcome the proposal of Dong, Roth and Su to revisit the foundations of differential privacy from the hypothesis testing point of view. Designing formal privacy definitions that can be interpreted by relevant stakeholders and decision makers is a necessary condition for adoption outside the technical literature, and basing such formulations on the trade-off between Type I and Type II errors in a classical hypothesis testing problem is (probably) as close as one can get to this goal. It was already well known that hypothesis testing interpretations can be derived from differential privacy and some of its variants, but the present work makes an interesting twist: it introduces new privacy definitions—namely, f-DP and its focal instance Gaussian DP—where hypothesis testing is front and centre, and shows this has many interesting consequences.

The first remarkable observation is that these new definitions capture all the desirable properties of prior differential privacy definitions (e.g. composition, amplification by sampling, Gaussian mechanism) in a *tight* analytical way. This should be contrasted with the limitations of other well-established definitions; for example, ε-DP cannot accommodate the Gaussian mechanism, (ε, δ)-DP leads to cumbersome composition formulas, and amplification by sampling in

Rényi DP is plagued with technical difficulties. On a qualitative level, we care about these properties because they allow us to analyse complex algorithms in terms of their building blocks. More importantly, having a definition that provides quantitatively tighter privacy guarantees is crucial when tuning hyperparameters to obtain the best possible trade-off between privacy and utility.

The limit theorems for composition and group privacy proved in the paper are quite interesting, most notably because they improve our understanding of these phenomena at an intuitive level and bring to light some underlying regularities in operations that are routinely encountered in differential privacy. In particular, the focality of Gaussian DP under the composition CLT enables more precise back-of-the-envelope computations in the course of algorithm design, and the error bounds in its Berry–Esseen version can serve as the building block for novel, more efficient so-called privacy accounting algorithms.

On a more technical note, I was surprised by the significant gap between the privacy amplification properties of standard DP and Gaussian DP. This gap suggests we might not yet fully understand the quantitative properties of this important primitive, and point to interesting directions for future research. In more practical terms, understanding privacy amplification by sampling is key to provide tight guarantees for differentially private stochastic gradient descent (DP-SGD) algorithms. The analysis of this algorithm based on Gaussian DP is very illuminating in terms of the gap with respect to the moments accountant technique. And although this gap can be reduced by considering the whole collection of $(\varepsilon, \delta(\varepsilon))$-DP guarantees that moments accountant provide—as opposed to using a single value of δ as is done in the paper—the remaining gap is still significant. I would encourage practitioners to take note of this and start using Gaussian DP accounting in their DP-SGD implementations.

Overall, I think it is fair to say that the work of Dong, Roth and Su represents a significant leap in our understanding of the hypothesis testing viewpoint on differential privacy. More importantly, it indicates that this point of view has many other practical benefits besides the obvious interpretability gains. I look forward to seeing more Guassian DP analyses of complex privacy-preserving algorithms in the coming years.

How to cite this article: Balle, B. Proposer of the vote of thanks to Dong *et al.* and contribution to the Discussion of 'Gaussian Differential Privacy'. *J R Stat Soc Series B*. 2022; 84, 3–54. https://doi.org/10.1111/rssb.12455

DOI: 10.1111/rssb.12456

Seconder of the vote of thanks to Dong *et al.* and contribution to the Discussion of 'Gaussian Differential Privacy'

Marco Avella-Medina

Department of Statistics, Columbia University, New York, NY, USA

Correspondence
Marco Avella-Medina, Columbia University, Department of Statistics, New York, NY, USA.
Email: marco.avella@columbia.edu

I congratulate the authors for a remarkable foundational paper that introduces an appealing new variant of differential privacy. It elegantly frames the problem of private data releases as a hypothesis testing problem. The impressive set of results established in this work sheds new light into the fundamental problem of composition. It demonstrates how the *f*-differential privacy framework successfully overcomes inevitable drawbacks of existing alternatives and establishes Gaussian differential privacy (GDP) at the core of this theory. One can expect the latter to become a dominant approach in this literature given its appealing intuitive hypothesis testing interpretation, exact composition property, central limit role for composition and computational tractability for approximating privacy losses.

Arguably one of the main practical benefits of the refined analysis of composition developed in this work is its direct application in numerous machine learning tasks via a differentially private algorithms in the of spirit stochastic gradient descent (SGD). See Bu et al. (2020) for interesting applications in deep learning. I shall focus my discussion on three important questions than one may ask when considering such a noisy SGD algorithm.

1. The suggested algorithm clips the gradient at some prespecified level. How do we choose this clipping constant in practice?
2. What can we say about the convergence of the noisy algorithm?
3. What are the statistical properties of the resulting estimators?

I attempt to provide some partial answers by considering a parametric M-estimation framework. Let $\hat{\theta}$ be defined as

$$\hat{\theta} = \underset{\theta\in\Theta}{\text{argmin}}\mathcal{L}_n(\theta) = \underset{\theta\in\Theta}{\text{argmin}}\frac{1}{n}\sum_{i=1}^{n}\rho(x_i,\theta) = \underset{\theta\in\Theta}{\text{argmin}}\mathbb{E}_{F_n}[\rho(X,\theta)],$$

where $x_1, \cdots, x_n \in \mathcal{X} \subset \mathbb{R}^m$ are i.i.d. according to F_{θ_0} and F_n denotes the empirical distribution function. Note that convexity of ρ typically guarantees the uniqueness of $\hat{\theta}$ and that if ρ is differentiable, $\hat{\theta}$ is also implicitly defined as the solution of the equation

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(x_i,\hat{\theta})=0, \tag{1}$$

where $\Psi(x,\ \theta)=\frac{\partial}{\partial\theta}\rho(x,\theta)$. This class of estimators is a strict generalization of the class of regular maximum likelihood estimators which are recovered when $\mathcal{L}_n(\theta)$ is the log-likelihood, that is when we take $\rho(x,\ \theta)=\ -\log f_\theta(x)$. In robust statistics, M-estimators defined through a function Ψ that is bounded in $x\in\mathcal{X}$ are particularly appealing. Indeed, a bounded Ψ guarantees that the M-estimator has a bounded influence function and therefore ensures that it is robust to the presence of a small fraction of outliers in the data (Hampel et al., 1986; Huber & Ronchetti, 2009). In the context of noisy SGD it is also critical to have $\sup_{x,\theta}\|\Psi(x,\ \theta)\|_2\ \le\ B\ <\ \infty$ since the bound B is used in the calibration of the privacy inducing noise. See lines 6–7 of `noisySGD`.

With the above setting in mind, let us try to answer the first question. Note that if we use noisy SGD to compute a maximum likelihood estimator we will in fact be computing a differentially private counterpart of the clipped likelihood estimator

$$\tilde{\theta}\colon\frac{1}{n}\sum_{i=1}^{n}h_c\left(\nabla\log f(x_i,\tilde{\theta})\right)\ =0,$$

where $h_c(z)=z\min\{1,\ \frac{c}{\|z\|_2}\}$ is the multivariate Huber function (Hampel et al., 1986, p.239). While clipping guarantees robustness via a bounded influence function, the resulting estimators are in general not consistent since the estimating equations are in general not unbiased i.e. $\mathbb{E}_{F_{\theta_0}}[h_c\left(\nabla\log f(x_i,\theta_0)\right)]\ \neq\ 0$. Hence, even though gradient clipping is a common suggestion in the differential privacy literature, it is not the most appealing from a statistical viewpoint. A possible solution is to consider a diverging clipping constant c, but a natural simple alternative is to use instead a consistent bounded influence M-estimator. In the context of normal linear regression, we observe $\{y_i,\ x_i\}_{i=1}^{n}$ and obtain the clipped least squares estimator

$$\tilde{\theta}\colon\frac{1}{n}\sum_{i=1}^{n}h_c\left((y_i-x_i^\top\tilde{\theta})x_i\right)\ =0.$$

In this particular case the clipped estimator happens to be a consistent estimator because of the symmetric errors assumed by the model. One could also consider a differentially private analogue of a Mallows' type estimator

$$\hat{\theta}\colon\frac{1}{n}\sum_{i=1}^{n}\psi_c(y_i-x_i^\top\hat{\theta})\frac{x_i}{\|x_i\|_2}=0,$$

where $\psi_c(r)=\max\{\ -c,\ \ \min(r,\ c)\}$ is the Huber function.

In order to give some insights into the next two questions, I will first consider an alternative noisy gradient descent (GD) algorithm defined by the iterates

$$\theta^{(k+1)}=\theta^{(k)}-\eta\left(\frac{1}{n}\sum_{i=1}^{n}\Psi(x_i,\theta^{(k)})+\frac{2\sup\|\Psi\|_2\cdot\sqrt{K}}{n\mu}Z_k\right),\quad\{Z_k\}\overset{iid}{\sim}N(0,\mathbb{I}_p).$$

One can give a clear answer to questions 2 and 3 for noisy GD based on recent results of a collaborative work (Avella-Medina et al., 2021). They also give an idea of what might be expected for noisy SGD. The following informal statement relates the properties of the Kth iterate of noisy GD $\theta^{(K)}$ to those of the M-estimator $\hat{\theta}$ defined in Equation (1).

Theorem 1 Assuming local strong convexity, after $K \geq C \log n$ iterations of NGD we have that $\theta^{(K)}$ is μ-GDP and $\theta^{(K)} - \theta_0 = \hat{\theta} - \theta_0 + O_p\left(\frac{\sqrt{K}p}{\mu n}\right)$.We can draw a few important conclusions from this theorem. We see that $O(\log \ n)$ steps suffice in order to guarantee that the iterates $\theta^{(K)}$ approaches $\hat{\theta}$ up to an error that is proportional to the privacy inducing noise added to the usual GD step. An intuitive interpretation of this result from standard optimization theory is that, under local strong convexity, GD requires K to be of the order $O(\log \ (1/\Delta))$ if we want to guarantee the optimization error to be $\| \theta^{(K)} - \hat{\theta} \|_2 = O(\Delta)$. This means that $O(\log \ n)$ steps suffice if we want to make the optimization error be of the same order as the privacy inducing noise. The theorem proves that this is also the case for noisy GD. Importantly, as long as $\frac{\sqrt{K}p}{\mu\sqrt{n}} \to 0$ the added statistical cost of μ-GDP is negligible in the sense that $\theta^{(K)} = \hat{\theta} + o_p(1/\sqrt{n})$ and hence under standard regularity conditions $\theta^{(K)}$ is also asymptotically normally distributed. In fact, if we translate the μ-GDP guarantee into a (ε, δ)-DP guarantee, we see that the rates of convergence of noisy GD match the minimax lower bound rates of Cai et al. (2021) up to $\sqrt{\log n}$ factor as long as we take $K = C \log n$ iterations. Thus noisy GD achieves optimal rates of convergence among the class of (ε, δ)-DP estimators.

Let me now return to noisy SGD and conclude by pointing out some challenges. Indeed, the theory of noisy GD combined with standard SGD suggest a couple of possible statistical issues. A first potential problem arises from the well-known fact that the standard SGD converges at a slower rate than GD (Bubeck, 2015). More precisely, under strong convexity, $O(\log n)$ steps of GD give the same accuracy as $O(\sqrt{n})$ steps of GD. This is not an issue in classical settings, but the theory of noisy GD suggests it might be a problem for noisy SGD since the number of iterations has a direct impact on the magnitude of the noise term. A second important problem is that a fixed mini-batch size m also entails that we have a non-vanishing noise term in line 7 of `noisySGD`. Here again the theory of noisy GD suggests that we might not have consistent noisy SGD estimators unless $m \to \infty$ and the cost of privacy might not negligible unless we also have that $\frac{m^2}{n} \to \infty$. I think that these problems deserve further attention in future research.

REFERENCES

Avella-Medina, M., Bradshaw, C. & Loh, P.L. (2021) Differentially private inference via noisy optimization. ArXiv:2103.11003.

Bu, Z., Dong, J., Long, Q. & Su, W.J. (2020) Deep learning with Gaussian differential privacy. *Harvard Data Science Review*, 2, 3.

Bubeck, S. (2015) Convex optimization: algorithms and complexity. *Foundations and Trends ® in Machine Learning*, 8(3–4), 231–357.

Cai, T.T., Wang, Y. & Zhang, L. (2021) The cost of privacy: optimal rates of convergence for parameter estimation with differential privacy. Annals of Statisttics (to appear).

Hampel, F., Ronchetti, E., Rousseeuw, P. & Stahel, W. (1986) *Robust statistics: the approach based on influence functions*. New York: Wiley.

Huber, P. & Ronchetti, E. (2009) *Robust statistics*, 2nd edn. New York: Wiley.

How to cite this article: Avella-Medina M. Seconder of the vote of thanks to Dong *et al.* and contribution to the Discussion of 'Gaussian Differential Privacy'. *J R Stat Soc Series B*. 2022;84:3–54. https://doi.org/10.1111/rssb.12456

The vote of thanks was passed by acclamation.

DOI: 10.1111/rssb.12457

Peter Krusche and Frank Bretz's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

Peter Krusche | **Frank Bretz**

Novartis Pharma, Advanced Methodology and Data Science, Basel, Switzerland

Correspondence
Peter Krusche, Novartis Pharma, Advanced Methodology and Data Science, Basel, Switzerland.
Email: peter.krusche@novartis.com

We congratulate Drs. Dong, Roth and Su for advancing our understanding of differential privacy from a hypothesis-testing perspective. As more personal data are collected for research, we need to mathematically understand when an adversary may be tempted to identify individuals and obtain sensitive information about them. This is especially true when working with patient data in healthcare, as respecting patient consent and privacy is imperative and subject to strong legal and regulatory constraints. In addition, not all data modalities collected in clinical trials are uniformly suitable for anonymization or de-identification. For example, only a few genetic markers or peripheral information present in medical images may be sufficient to uniquely re-identify individuals. As a result, patient privacy requirements can severely limit our ability to link data across clinical studies and build complex data sets to improve our understanding of disease and response to treatments.

Despite the progress made so far, there is limited work on privacy-preserving techniques with applications to clinical research. When computing differentially private data summaries or model parameters on clinical data sets, we trade data utility off for privacy guarantees through randomized algorithms that are parameterized by a 'privacy budget'. Data from clinical trials present both unique challenges and opportunities in this context: for one, knowledge of randomization by study design may allow us to suppress data with little impact on the summary being computed. For another, clinical trial data sets with only hundreds or thousands of participants are much smaller than other types of data typically considered in the context of differential privacy (insurance data, electronic health records, etc.), making it challenging to retain data utility as well as privacy.

These specific constraints raise interesting questions as to how the proposed methods can be applied to clinical research, for example around the management of the differential privacy budget. As such, we believe it would be highly useful to study the practical implications of μ/f-DP and ε-DP for small clinical data sets. For example, could we address the trade-offs between data utility and preserving privacy in the framework of f-DP while acknowledging the customary hypothesis-testing approaches in clinical research where power is maximized while controlling the Type I error rate at a specified threshold. Moreover, we believe it would be interesting to study if there are non-trivial ways to utilize the rules of composition in f-DP for optimizing complex sets of private data summaries, or the training process of private generative models.

How to cite this article: Krusche, P. & Bretz, F. Peter Krusche and Frank Bretz's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al. J R Stat Soc Series B*. 2022; 84, 3–54. https://doi.org/10.1111/rssb.12457

The following contributions were received in writing after the meeting.

DOI: 10.1111/rssb.12458

Christine P. Chai's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

Christine P. Chai[1]

Microsoft Corporation, Redmond, Washington, USA

Correspondence
Christine P. Chai, Microsoft Corporation, Redmond, WA, USA.
Email: cpchai21@gmail.com

Due to the emerging concerns of individual confidentiality, differential privacy has been a hot topic in recent years. Most papers talk about the applications, and this article stands out by focusing on the mathematical analysis and proving the asymptotic bounds. For the readers who are unfamiliar with differential privacy, I recommend a shorter article (Snoke & Bowen, 2020) as a starting point to learn about the context.

Since the authors used subsampling to amplify privacy guarantees, I wonder if they have considered moving to fully synthetic data to preserve perfect privacy for all individuals. In this way, individuals cannot be identified from the synthetic data because the data do not contain real people (Howe et al., 2017; Jarmin et al., 2014). Can the hypothesis testing framework be applied on proving that the synthetic data have the same key statistical properties as the original data?

Another question I have is how Gaussian differential privacy preserves data usability. The paper discussed the trade-off in terms of Type I and Type II errors from the attacker's perspective, so I am curious about the trade-off between privacy guarantees and data usability, that is

[1]Disclaimer: The opinions and views expressed here are those of the author and do not necessarily state or reflect those of Microsoft.

the privacy budget. Researchers have expressed concerns about the accuracy of Census public release data, due to the implementation of differential privacy and disclosure avoidance methods (Hauer & Santos-Lozada, 2021; Ruggles et al., 2019).

Last but not least, can the proposed Gaussian Differential Privacy framework be applied to COVID-19 contact tracing data (Cho et al., 2020) in the future? Do the authors anticipate any major challenges in the implementation?

REFERENCES

Cho, H., Ippolito, D. & Yu, Y.W. (2020) Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv preprint arXiv:2003.11511.

Hauer, M.E. & Santos-Lozada, A.R. (2021) Differential Privacy in the 2020 Census Will Distort COVID-19 Rates. *Socius: Sociological Research for a Dynamic World*, 7 http://dx.doi.org/10.1177/2378023121994014.

Howe, B., Stoyanovich, J., Ping, H., Herman, B. & Gee, M. (2017) Synthetic data for social good. arXiv preprint arXiv:1710.08874.

Jarmin, R.S., Louis, T.A. & Miranda, J. (2014) Expanding the role of synthetic data at the U.S. Census Bureau. *Statistical Journal of the IAOS*, 30(2), 117–121.

Ruggles, S., Fitch, C., Magnuson, D. & Schroeder, J. (2019) Differential privacy and census data: Implications for social and economic research. *AEA Papers and Proceedings*, 109, 403–408.

Snoke, J. & Bowen, C.M. (2020) How statisticians should grapple with privacy in a changing data landscape. *CHANCE*, 33(4), 6–13.

How to cite this article: Chai CP. Christine P. Chai's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al*. *J R Stat Soc Series B*. 2022; 84, 3–54. https://doi.org/10.1111/rssb.12458

DOI: 10.1111/rssb.12459

Sebastian Dietz's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

Sebastian Dietz

Munich, Germany

Correspondence
Sebastian Dietz, Munich, Germany
Email: sjd@hotmail.de

The remarkable new contribution is the usage of trade-off functions between error types I and II, which leads to the generalization of differential privacy. It is intuitively well visualized by a kind of inverse receiver operating characteristic. Via the central limit theorem (Theorems 4), it is shown that a tensor product of symmetric trade-off functions with finite first to third moments is bounded by a Gaussian trade-off function. The proofs in section D incorporate the classical Berry–Esseen theorem as theorem D.4.

Shortly in section 3.2, the topic of requirement of a finite third moment is touched, which is crucial for the Berry–Esseen central limit theorems. For transferring the results of the DP central limit theorem (Theorems 4 and 5) to (ε, δ)-DP, there the question is taken up again to prove convergence to a Gaussian trade-off function in section 3.3.

In literature, some generalizations of Berry–Esseen exist, for example by Petrov (1975), where Theorem 5 states in a simplified form the following:

Let $X_i, \ldots, X_n$ be random variables with mean zero and variance 1. Let $F_n(x) = P(n^{-\frac{1}{2}} \sum_{i=1}^{n} < x)$. Let g() be a non-negative, non-decreasing and even function in the interval $x > 0$ such that also $\frac{x}{g(x)}$ is non-decreasing in the interval $x > 0$. For $E[X_1^2 g(X_1)] < \infty$ it holds that

$$sup|F_n(x) - \phi(x)| \leq \frac{A}{g(\sqrt{n})} E[X_1^2 g(X_1)]$$

for some absolute positive constant A. Results might also be extended to different variances as shown in Petrov (1975) and also non-zero means as shown in DasGupta (2008).

Using such a third-moment-free central limit theorem, the (ε, δ)-DP trade-off function could also be covered by the central limit theorem as well as a potentially larger scope of trade-off functions. The proofs do not require a specific value for the constant, such that a consideration of this Berry–Esseen extension might be an option, whereas handling of g() might be a limiting issue.

REFERENCES

Petrov, V.V. (1975) *Sum of independent random variables*. Berlin Heidelberg, NY: Springer.
DasGupta, A. (2008) *Asymptotic theory of statistics and probability*. Heidelberg, NY: Springer.

How to cite this article: Dietz S. Sebastian Dietz's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al*. *J R Stat Soc Series B*. 2022;84:3–54. https://doi.org/10.1111/rssb.12459

DOI: 10.1111/rssb.12460

J. Goseling and M.N.M. van Lieshout's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

J. Goseling[1] | **M.N.M. van Lieshout**[1,2]

[1]University of Twente, Enschede, Netherlands

[2]CWI, Amsterdam, Netherlands

Correspondence: M.N.M. van Lieshout, CWI, Amsterdam, Netherlands. Email: marie-colette.van.lieshout@cwi.nl

We congratulate Professors Dong, Roth and Su on their compelling work on *Gaussian Differential Privacy*.

In official statistics, the *p*%-rule (Hundepoel et al., 2012) is widely used to protect tabular data. In recent work (Hut et al., 2020) we adapted this concept to thematic maps, for example, of energy consumption per company. Usually such maps are drawn directly from an underlying table that is protected from disclosure. The resulting colour-coded map, however, is, by construction, discretised in regions defined by the cells in the table. These geographic regions are usually large, corresponding, for instance, to municipalities. The resulting protection is very conservative, leading to a map with reduced utility. Therefore, there is a need for smooth thematic maps.

One might use the Nadaraya–Watson kernel weighted average. This procedure, however, is not necessarily safe. Indeed, suppose that an attacker is able to read off the plotted, smoothed, values of the variables of interest at all measurement locations. Then their original values satisfy a linear system which in many cases (including that of a Gaussian kernel) can be solved exactly if the measurement locations are distinct.

To protect sensitive information we propose to add correlated Gaussian noise E with variance τ and map

$$\frac{\sum_{i=1,\cdots,N} g_i \kappa((r-r_i)/h) + E(r)}{\sum_{i=1,\cdots,N} \kappa((r-r_i)/h)}, \quad r \in D.$$

Here the $g_i > 0$ are the values of the variable at distinct locations r_i in a planar region D, κ is the Gaussian kernel and $h > 0$ the bandwidth that determines the amount of smoothing.

The counterpart of the *p*%-rule is as follows. Let $0 \le \alpha < 1$. Then a map is *unsafe* if

$$\max_{1=1,\cdots,N} P\left(\left|\frac{\hat{g}_i - g_i}{g_i}\right| < \frac{p}{100}\right) > \alpha.$$

In words, a map is safe when small relative errors happen with small probability. We proved that if

$$\sqrt{\tau} \geq \frac{p}{100\Phi^{-1}((1+\alpha)/2)} \max_{i=1,\cdots,N} \left\{ \frac{g_i}{\sqrt{(K_h^{-1})_{ii}}} \right\},$$

where $K_h = (\kappa((r_i - r_j)/h))_{i,j=1,\cdots N}$, the resulting thematic map is safe.

REFERENCES

Hundepoel, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K. et al. (2012) *Statistical disclosure control*. Hoboken: Wiley.

Hut, D., Goseling, J., van Lieshout, M.C., de Wolf, P.P. & de Jonge, E. (2020) Statistical disclosure control when publishing on thematic maps. *LNCS*, 12276, 195–205.

How to cite this article: Goseling J, van Lieshout M.N.M. J. Goseling and M.N.M. van Lieshout's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.* *J R Stat Soc Series B*. 2022;84:3–54. https://doi.org/10.1111/rssb.12460

DOI: 10.1111/rssb.12461

Jorge Mateu's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

Jorge Mateu

Department of Mathematics, University Jaume I, Castellón, Spain

Correspondence
Jorge Mateu, Department of Mathematics, University Jaume I, E-12071, Castellón, Spain.
Email: mateu@uji.es

The authors are to be congratulated on a valuable and thought-provoking contribution motivating this new framework for private data analysis, the f-differential privacy. A key aspect is the use of trade-off functions of hypothesis testing as a measure of indistinguishability of two or group neighbouring data sets.

I would like to frame this contribution within the context of big data together with data mining and data science. As the authors point out, an increasing and unprecedented wealth of methods concern individual data recorded from personal devices or even private or public resources. In this large-scale and big data context, privacy in terms of anonymous personal information is key for any legal and serious data analysis. Modern data collection techniques allow tracking objects (persons) continuously. This means that we do not only know the current location of a moving object, but we also track the objects over time. A set of some tracks from different moving objects may be considered a trajectory pattern. Indeed, studying the behaviour of moving objects over time and their interaction, either between objects or with environment, plays a crucial role in understanding how they use space and more importantly how they interact with each other. In this context, a snapshot of a trajectory pattern might be seen as a spatial point pattern.

I pose the following two cases. One is a data set where the individual events are themselves trajectories (functions) moving within a city. We need algorithms for privacy guarantee and to get groups of (trajectory) data anonymised. In a related context, assume the events are exact space–time coordinates of infected people from an infectious disease. So we have a spatiotemporal point pattern and we need to test if the ratio of the first-order intensity of the infected group against that of the control group behaves in a particular way. In other words, we need to test if groups of space–time events are distinguishable from other events. These two problems deal with data in space–time, and pose problems on privacy over space–time locations. I wonder if GDP applies over this spatial context. Also, a natural question is how f-DP or GDP can be used in contexts where type I and II errors are only approximated by simulations -because the probability distribution under the null or alternative hypothesis is usually unknown, as often happens in the spatiotemporal context.

How to cite this article: Mateu, J. Jorge Mateu's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al*. *J R Stat Soc Series B*. 2022; 84, 3–54. https://doi.org/10.1111/rssb.12461

DOI: 10.1111/rssb.12462

Priyantha Wijayatunga's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

Priyantha Wijayatunga

Department of Statistics, Umeå University, Umeå, Sweden

Correspondence
Priyantha Wijayatunga, Department of Statistics, Umeå University, Umeå, Sweden.
Email: priyantha.wijayatunga@umu.se.

In privacy preserving, for example using a differential privacy framework, when a statistical operation is done on a database, then it is done so that its results are not overly dependent on any data record in the database. In this sense, one may think that this is all about avoiding, for example outliers, when the desired statistical results are generated. But on the other hand, utility of such a result is questionable. When an average of a certain measure is requested, giving the median or an average calculated from implementation of a suitable resampling procedure may preserve the privacy of individual data. Such methods may be simple to implement and may have an appreciable utility. Recently in Santos-Lozada et al. (2020) using US census data, the authors show that implementation of the differential privacy will produce dramatic changes in population counts for racial/ethnic minorities in small areas and less urban settings, significantly altering knowledge about health disparities in mortality. It is also important to note that, according to the *Fundamental Law of Information Recovery*; 'overly accurate answers to too many questions (on statistics) will destroy privacy (of individual data) in a spectacular way'.

The paper tries to use frequentist statistical hypothesis testing framework for defining their differential privacy framework. However, as many of us are aware the hypothesis testing is undergoing immense criticism, especially within the applied statistical community, for example so-called p-value problems. It may be that such problems and oppositions may appear in any privacy framework that is based on the frequentist statistical hypothesis testing methodology. Therefore, ideally the authors should touch upon such problems, especially to attract applied researchers (in the sense of the discipline of statistics) such as computer scientists, social scientists, etc., to their approach. In fact, the authors emphasize the importance of the use of the Neyman–Pearson hypothesis testing framework for interpreting differential privacy over other methods. According to original thesis of R. A. Fisher (1890–1962), the meaning of, for example, p-value < 0.05 is that the respective experiment should be repeated a few times. Such a practice should be handled by authors' composition results. Apart from above issues, it seems that some mathematical expressions should accompany with some verbal expressions too, for example the inequality in the *Definition 1* is valid for a given (fixed) (S, S') pair. Therefore, it is helpful to write $\mathbb{P}\{M(S) \in E \mid S\}$ rather than writing $\mathbb{P}\{M(S) \in E\}$ and indicate how strong the conditional (in)dependence of $M(S)$ on S' given S. Any other confusions should be eliminated.

REFERENCE

Santos-Lozada, A.R., Howard, J.T. & Verderyc, A.M. (2020) How differential privacy will affect our understanding of health disparities in the United States. *Proceedings of the National Academy of Sciences*, 117(24), 13405–13412. https://doi.org/10.1073/pnas.2003714117

How to cite this article: Wijayatunga P. Priyantha Wijayatunga's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al. J R Stat Soc Series B.* 2022; 84, 3–54. https://doi.org/10.1111/rssb.12462

The author replied later, in writing, as follows.

DOI: 10.1111/rssb.12463

Authors' reply to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

Jinshuo Dong | **Aaron Roth** | **Weijie J. Su**

University of Pennsylvania, USA

Correspondence
Weijie Su, Department of Statistics,
Wharton School, University of Pennsylvania, USA.
Email: suw@wharton.upenn.edu

We warmly thank Editor Paul Smith for selecting our paper for discussion and are extremely grateful to all the discussants for taking their valuable time to provide engaging and stimulating feedback on our work. These insights situate our work in context and provide promising directions for future research. We are excited to see that thoughts about theoretical complements and new applications are already emerging.

A general view, shared by all discussants, is that privacy is a first-order concern in many data science problems. We are very pleased to learn that our statistics community welcomes new foundational development and methodological contributions that allow for privacy protections in statistical data analysis.

In this rejoinder, we aim to address two broad issues that cover most comments made in the discussion. First, we discuss some theoretical aspects of our work and comment on how this work might impact the theoretical foundation of privacy-preserving data analysis. Taking a practical viewpoint, we next discuss how *f*-differential privacy (*f*-DP) and Gaussian differential privacy (GDP) can make a difference in a range of applications.

1 | THEORETICAL ASPECTS OF *f*-DP

As echoed by many discussants, the formalization of *f*-DP in our paper (Dong et al., 2021a) starts from a decision-theoretic interpretation of a 'differential' privacy attack, which originates from the work of Wasserman and Zhou (2010). The binary nature of the decision-theoretic problems renders the classical theory of hypothesis testing a basic tool. Specifically, we use the trade-off between type I and type II errors of the hypothesis testing problem as our privacy measure. In response to a question raised by Dr. Jorge Mateu, we remark that this trade-off is concerned with a thought experiment in which someone is trying to determine if an individual's data is in the data set or not, rather than a hypothesis test that is actually conducted. It can therefore always be reasoned about analytically/formally, without needing simulation, even if the algorithms themselves are complex or simulation based.

This treatment of privacy cost in *f*-DP comes with several technical properties that can facilitate the development of better differentially private algorithms. As highlighted by Dr. Borja Balle, *f*-DP gives tight and analytically tractable formulas for composition. This appealing feature arises from applying the central limit theorem to the privacy loss random variables, thereby making GDP a canonical single-parameter family of privacy definitions within the *f*-DP class. While we did not attempt to push hard on weakening assumptions for the privacy central limit theorems, there are several possible extensions. For example, one may be able to identify a necessary and sufficient condition for the privacy central limit theorem to hold, just like the Lindeberg–Feller condition for the usual central limit theorems. Another possibility is to sharpen the central limit theorem by leveraging a refined analysis of the privacy loss random variables (Zheng et al., 2020). More specifically, Dr. Sebastian Dietz suggested a very interesting direction for improving the composition formulas by making use of a third-moment-free central limit theorem (see, for example, DasGupta (2008)). A successful investigation in this direction might extend the applicability of the composition formulas to (ε, δ)-DP and others. More broadly, it would also be interesting to explore central limit theorem phenomena of privacy beyond composition. For example, Dong et al. (2021b) recently showed that a related central limit theorem occurs in high-dimensional query answering and yet privacy cost is best described in the framework of *f*-DP. We see all these as interesting future directions for broadening the scope of the hypothesis testing viewpoint on differential privacy.

In addition to composition, subsampling is another important primitive that is involved in many algorithm designs. As pointed out by Dr. Borja Balle, while divergence-based privacy definitions face technical difficulties in describing privacy amplification by subsampling, *f*-DP gives a relatively concise and coherent expression for understanding how privacy is amplified using this primitive. This also gives a sharper privacy analysis of subsampling than can be obtained by directly using (ε, δ)-DP. An interesting observation made by Dr. Borja Balle is that the significant gap between the two frameworks seems surprising, and warrants further investigation. It is also worth developing similar privacy analyses for the various flavors of subsampling schemes used in training deep learning models (not all of which involve *independent* sampling across rounds).

2 | APPLICATIONS OF *f*-DP

Our main hope for *f*-DP is to see as many applications as possible to improving the privacy analyses of the diverse algorithms used in a variety of data science problems. Encouragingly, we found many such possibilities in the discussants' contributions that either tackle important problems or show great promise.

Understanding the trade-off between privacy and utility for various statistical and computational tasks is *the* central object of study in the differentially privacy literature. The main point of *f*-DP and GDP is to make it possible to capture this fundamental trade-off more precisely. As a result, the *f*-DP framework allows us to obtain better trade-offs between privacy guarantees and data usability. This trade-off is different in different applications, and requires analyses on a case by case basis. We emphasize, as pointed out by several of the discussants that privacy protection does inevitably come with utility loss. Indeed, this is a consequence of the 'fundamental law of information recovery', which applies not just to differential privacy but to any method of releasing data. So while it is true that differential privacy can harm utility (especially for small data sets), this is not an artefact of differential privacy, but an actual, fundamental trade-off that we have to grapple with as a society. We can choose to get exact statistics about our data, but we should understand that this means giving up on privacy. Differential privacy takes no stand on how we should mediate this fundamental trade-off: rather it provides a precise language in which to talk about it.

2.1 | *f*-DP for stochastic optimization

To appreciate how sharply this trade-off can be characterized using a given privacy definition, perhaps the best benchmark is stochastic gradient descent (SGD), the basic foundation for many machine learning algorithms. Owing to its effectiveness in handling composition and subsampling, *f*-DP gives a tighter privacy analysis of SGD than the moments accountant technique (Abadi et al., 2016), which further feeds back into improved test accuracy of trained deep learning models at fixed privacy guarantees (Bu et al., 2020). We are delighted that Dr. Borja Balle wrote 'I would encourage practitioners to take note of this and start using Gaussian DP accounting in their DP-SGD implementations'.

Moving forward, Dr. Marco Avella-Medina raised several interesting and important questions regarding private SGD with *f*-DP guarantees. Although gradient clipping is a necessary step in private SGD that ensures bounded sensitivity to any single data point, this step can lead to inconsistency for some problems. To go around this difficulty, Avella-Medina suggested using a consistent bounded influence *M*-estimator from robust statistics, which we believe is a promising approach worthy of future research effort. Moreover, we are glad to see that Avella-Medina et al. (2021) introduced a kind of noisy gradient descent and analysed its Gaussian differential privacy properties. This opens an exciting research avenue to understand when noisy gradient descent outperforms SGD.

2.2 | Other applications

Differential privacy has applications beyond machine learning. A promising application area—due to strict privacy regulation—is in the analysis and sharing of medical data. A challenge in medical applications is that the size of the relevant data sets is often relatively small. The improved trade-off between privacy and utility is especially important in the challenging small data regime. As noted by Drs. Peter Krusche and Frank Bretz, there are obstacles to combining data across hospitals—for which we think differential privacy might be able to help. Noisy access to a large data set might be better—even from the perspective of utility—than exact access to only a small local data set. When applying privacy protections to small data, it is especially important not be as tight as possible in accounting for privacy loss, which is one of the main benefits of the f-DP framework.

Dr. Jorge Mateu brought up privacy issues that arise when analysing trajectory data. In principle, f-DP and GDP can be applied to any kind of data, such as trajectory data. It would make sense, for example to think about the question of releasing statistics about trajectory data or a synthetic data set consisting of trajectories that maintain consistency with the real data with respect to various statistics of interest, so long as those statistics have low sensitivity and vary only mildly with the data of individuals. These types of problems deserve further study. Of course, providing useful analyses of a single *individual*'s trajectory is by design prevented by technologies that aim to preserve individual privacy. A related question, asked by Dr. Christine Chai, was whether the f-DP framework can be applied to COVID-19 contact tracing data. Differential privacy ((ε, δ)-DP, f-DP, or any related variant) is not directly applicable to what is most commonly known as contact tracing—letting contacts know that someone with COVID-19 has been near them—since by design, this is highly sensitive to a single data point. However, GDP (as well as other differential privacy variants) can be used to improve population level statistics related to contact tracing, such as how crowded grocery stores are by time and mobility data, or even what fraction of visitors to a grocery store in a given day have had potential COVID-19 exposure. More generally, we believe that f-DP has many more connections to various aspects of data science.

Finally, we remark that there are many heuristic approaches to privacy that do not come with the guarantees of differential privacy. There is a vast literature of pros and cons among these approaches, which is beyond the scope of this paper—but in general, ‘syntactic’ approaches do not stand up to attack by a determined adversary. In particular, synthetic data is known to be neither necessary nor sufficient for privacy—but also not incompatible with differential privacy. For example, there is a large literature on generating differentially private synthetic data (see, e.g., Blum et al. (2013); Gaboardi et al. (2014); Vietri et al. (2020); Aydore et al. (2021); Jordon et al. (2018); Beaulieu-Jones et al. (2019)), most of which we believe can be improved by f-DP style analyses.

REFERENCES

Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K. et al. (2016) Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318.

Avella-Medina, M., Bradshaw, C. & Loh, P.-L. (2021) Differentially private inference via noisy optimization. *arXiv preprint arXiv:2103.11003*.

Aydore, S., Brown, W., Kearns, M., Kenthapadi, K., Melis, L., Roth, A. & Siva, A. (2021) Differentially private query release through adaptive projection. *arXiv preprint arXiv:2103.06641*.

Beaulieu-Jones, B.K., Wu, Z.S., Williams, C., Lee, R., Bhavnani, S.P., Byrd, J.B. et al. (2019) Privacy-preserving generative deep neural networks support clinical data sharing. *Circulation: Cardiovascular Quality and Outcomes*, 12(7):e005122.

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

Bu, Z., Dong, J., Long, Q. & Weijie, S. (2020) Deep learning with Gaussian differential privacy. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.cfc5dd25

DasGupta, A. (2008) *Asymptotic theory of statistics and probability*. Springer Science & Business Media.

Dong, J., Roth, A. & Su, W.J. (2021a) Gaussian differential privacy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1–35. https://doi.org/10.1111/rssb.12454

Dong, J., Su, W.J. & Zhang, L. (2021b) A central limit theorem for differentially private query answering. *arXiv preprint arXiv:2103.xxxxx*.

Gaboardi, M., Arias, E.J.G., Hsu, J., Roth, A. & Wu, Z.S. (2014) Dual query: Practical private query release for high dimensional data. In *International Conference on Machine Learning*, pages 1170–1178. PMLR.

Jordon, J., Yoon, J. & Van Der Schaar, M. (2018) Pate-gan: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations*.

Vietri, G., Tian, G., Bun, M., Steinke, T. & Wu, S. (2020) New oracle-efficient algorithms for private synthetic data release. In *International Conference on Machine Learning*, pages 9765–9774. PMLR.

Wasserman, L. & Zhou, S. (2010) A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489), 375–389.

Zheng, Q., Dong, J., Long, Q. & Su, W.J. (2020) Sharp composition bounds for Gaussian differential privacy via Edgeworth expansion. In *International Conference on Machine Learning*, pages 11420–11435.

How to cite this article: Dong J, Roth A, Su WJ. Authors' reply to the Discussion of 'Gaussian Differential Privacy' by Dong *et al*. *J R Stat Soc Series B*. 2022; 84: 3–54. https://doi.org/10.1111/rssb.12463