Subsampled Rényi Differential Privacy and Analytical Moments Accountant

Yu-Xiang Wang
UCSB

Borja Balle
Amazon Research

Shiva Prasad Kasiviswanathan
AWS AI

Abstract

We study the problem of subsampling in differential privacy (DP), a question that is the centerpiece behind many successful differentially private machine learning algorithms. Specifically, we provide a tight upper bound on the Rényi Differential Privacy (RDP) (Mironov, 2017) parameters for algorithms that: (1) subsample the dataset, and then (2) apply a randomized mechanism $\mathcal{M}$ to the subsample, in terms of the RDP parameters of $\mathcal{M}$ and the subsampling probability parameter. Our results generalize the moments accounting technique, developed by Abadi et al. (2016) for the Gaussian mechanism, to any subsampled RDP mechanism.

1 Introduction

Differential privacy (DP) is a mathematical definition of privacy proposed by Dwork et al. (2006b). Ever since its introduction, DP has been widely adopted and as of today, it has become the *de facto* standard of privacy definition in the academic world with also wide adoption in the industry (Erlingsson et al., 2014; Apple, 2017; Uber Security, 2017). DP provides provable protection against adversaries with arbitrary side information and computational power, allows clear quantification of privacy losses, and satisfies graceful composition over multiple access to the same data. Over the past decade, a large body of work has been developed to design basic algorithms and tools for achieving differential privacy, understanding the privacy-utility trade-offs in different data access setups, and on integrating differential privacy with machine learning and statistical inference. We refer the reader to (Dwork and Roth, 2013) for a more comprehensive overview.

Rényi Differential Privacy (RDP, see Definition 4) (Mironov, 2017) is a recent refinement of differential privacy (Dwork et al., 2006b). It offers a unified view of the ϵ-differential privacy (pure DP), (ϵ, δ)-differential privacy (approximate DP), and the related notion of *Concentrated Differential Privacy* (Dwork and Rothblum, 2016; Bun and Steinke, 2016). The RDP point of view is particularly useful when the dataset is accessed by a sequence of randomized mechanisms, as in this case a *moments accountant* technique can be used to effectively keep track of the usual (ϵ, δ) DP parameters across the entire range $\{(\epsilon(\delta), \delta) | \forall \delta \in [0, 1]\}$ (Abadi et al., 2016).

A prime use case for the moments accountant technique is the *NoisySGD* algorithm (Song et al., 2013; Bassily et al., 2014) for differentially private learning, which iteratively executes:

$$\theta_{t+1} \leftarrow \theta_t + \eta_t \left(\frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \nabla f_i(\theta_t) + Z_t \right) \qquad (1)$$

where θ_t is the model parameter at tth step, η_t is the learning rate, f_i is the loss function of data point i, ∇ is the standard gradient operator, $\mathcal{I}$ is an index set of size m drawn uniformly from $\{1, ..., n\}$, and $Z_t \sim \mathcal{N}(0, \sigma^2 I)$. Adding Gaussian noise (also known as the *Gaussian mechanism*) is a standard way of achieving (ϵ, δ)-differential privacy (Dwork et al., 2006a; Dwork and Roth, 2013; Balle and Wang, 2018). Since in the NoisySGD case the randomized algorithm subsamples the mini-batch $\mathcal{I}$ randomly before adding the Gaussian noise, the overall scheme could be viewed as a *subsampled Gaussian mechanism*. Therefore, with the right setting of σ, each iteration of NoisySGD can be thought of as a private release of a stochastic gradient.

More generally, a *subsampled mechanism* first takes a subsample of the dataset generated through some subsampling procedure[1], and then applies a known randomized mechanism $\mathcal{M}$ on the subsampled data points. It is important to exploit the randomness in subsampling because if $\mathcal{M}$ is (ϵ, δ)-DP, then (informally) a

Proceedings of the 22$^{\text{nd}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89. Copyright 2019 by the author(s).

[1] There are different subsampling methods, such as Poisson subsampling, sampling without replacement, sampling with replacement, etc.

subsampled mechanism obeys $(O(\gamma\epsilon), \gamma\delta)$-DP for some $\gamma < 1$ related to the sampling procedure. This is often referred to as the "privacy amplification" lemma[2] — a key property that enables NoisySGD and variants to achieve optimal rates in convex problems (Bassily et al., 2014), and to work competitively in Bayesian learning (Wang et al., 2015) and deep learning (Abadi et al., 2016) settings. We also note that privacy amplification is the key technical tool for characterizing learnability in statistical learning (Wang et al., 2016) and achieving tight sample complexity bounds for simple function classes (Beimel et al., 2013; Bun et al., 2015).

While privacy amplification by subsampling is a very important tool for designing good private algorithms, computing the RDP parameters for a subsampled mechanism is a non-trivial task. A natural question, with wide ranging implications for designing successful differentially private algorithms is the following: Can we obtain good bounds for privacy parameters of a subsampled mechanism in terms of privacy parameters of the original mechanism? With the exception of the special case of the Gaussian mechanism under Poisson subsampling analyzed in (Abadi et al., 2016), there is no analytical formula available to generically convert the RDP parameters of a mechanism $\mathcal{M}$ to the RDP parameters of the subsampled mechanism.

In this paper, we tackle this central problem in private data analysis and provide the first general result in this area. Specifically, we analyze RDP amplification under a *sampling without replacement* procedure: $\mathsf{subsample}$, which takes a data set of n points and outputs a sample from the uniform distribution over all subsets of size $m \leq n$. Our contributions can be summarized as follows:

(i) We provide a tight bound (Theorem 9) on the RDP parameter $(\epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha))$ of a subsampled mechanism $(\mathcal{M} \circ \mathsf{subsample})$ in terms of the RDP parameter $(\epsilon_{\mathcal{M}}(\alpha))$ of the original mechanism $\mathcal{M}$ and the subsampling ratio $\gamma := m/n$. Here, α is the order of the Rényi divergence in the RDP definition (see Definition 4 and the following discussion). This is the first general result in this area that can be applied to any RDP mechanism. For example, in addition to providing RDP parameter bounds for the subsampled Gaussian mechanism case, our result enables analytic calculation of similar bounds for other commonly used privacy mechanisms including subsampled Laplace mechanisms, subsampled randomized response mechanisms, subsampled "posterior sampling" algorithms under exponential family models (Geumlek et al., 2017), etc. Even for the subsampled Gaussian mechanism our bounds are tighter than those provided by Abadi et al. (2016) (albeit the subsampling procedure and the dataset neighboring relation they use are slightly different from ours).

(ii) Interestingly, our bound on the RDP parameter of a subsampled mechanism indicates that as the order of RDP α increases, there is a *phase transition* point α^* satisfying $\gamma\alpha^* e^{\epsilon_{\mathcal{M}}(\alpha^*)} \approx 1$, where $\epsilon_{\mathcal{M}}(\alpha)$ is the RDP parameter of the original mechanism $\mathcal{M}$. For $\alpha < \alpha^*$, the subsampled mechanism has an RDP parameter $\epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha) = O(\alpha\gamma^2(e^{\epsilon_{\mathcal{M}}(2)-1}))$, while for $\alpha > \alpha^*$, the RDP parameter $\epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha)$ either quickly converges to $\epsilon_{\mathcal{M}}(\alpha)$ which does not depend on γ, or tapers off at $O(\gamma\epsilon_{\mathcal{M}}(\infty))$ which happens when $e^{\epsilon_{\mathcal{M}}(\infty)} - 1 \ll 1/\gamma$. The subsampled Gaussian mechanism falls into the first category, while the subsampled Laplace mechanism falls into the second.

(iii) Our analysis reveals a new theoretical quantity of interest that has not been investigated before — a *ternary* version of the Pearson-Vajda divergence (formally defined in Appendix C). A privacy definition in terms of this divergence seems naturally coupled with understanding the effects of subsampling, just like how Rényi differential privacy (RDP) (Mironov, 2017) seems naturally coupled with understanding the effects of composition.

(iv) From a computational efficiency perspective, we propose an efficient data structure to keep track of the Rényi differential privacy parameters in its symbolic form, and output the corresponding (ϵ, δ)-differential privacy as needed using efficient numerical methods. This avoids the need to specify a discrete list of moments ahead of time as required in the *moments accountant* method of Abadi et al. (2016) (see the discussion in Section 3.2). Finally, our experiments confirm the improvements in privacy parameters that can be obtained by applying our bounds.

We end this introduction with a methodological remark. The main result of this paper is the bound in Theorem 9, which at first glance looks cumbersome. The remarks following the theorem's statement in Section 3.1 discuss some of the asymptotic implications of this bound, as well as its meaning in several special cases. These provide intuitive explanations justifying the tightness of the bound. In practice, however, asymptotic bounds are of limited interest: concrete bounds with explicit, tight constants that can be efficiently computed are needed to provide the best possible privacy-utility trade-off in practical applications of differential privacy. Thus, our results should be interpreted under this point of

[2] Informally, this lemma states that, if a private algorithm is run on a random subset of a larger dataset (and the identity of that subset remains hidden), then this new algorithm provides better privacy protection (reflected through improved privacy parameters) to the entire dataset as a whole than the original algorithm did.

view, which is summarized by the leitmotif *"in differential privacy, constants matter"*. Also, due to space constraints we have deferred many technical details to the appendices.

2 Background and Related Work

In this section, we review some background about differential privacy, some related privacy notions, and the technique of moments accountant.

Differential Privacy and Privacy Loss Random Variable. We start with the definition of (ϵ, δ)-differential privacy. We assume $\mathcal{X}$ is the domain where datapoints are drawn from. We call two datasets X and X' *neighboring* (adjacent) if they differ in at most one data point, meaning that we can obtain X' by replacing one data point from X by another arbitrary data point. We represent this as $d(X, X') \leq 1$.

Definition 1 (Differential Privacy (DP))**.** *A randomized algorithm $\mathcal{M} : \mathcal{X}^n \to \Theta$ is (ϵ, δ)-DP if for every pair of neighboring datasets $X, X' \in \mathcal{X}^n$, and every possible (measurable) output set $E \subseteq \Theta$ the following inequality holds:* $\Pr[\mathcal{M}(X) \in E] \leq e^{\epsilon} \Pr[\mathcal{M}(X') \in E] + \delta$.

The definition ensures that it is information-theoretically impossible for an adversary to infer whether the input dataset is X or X' beyond a certain confidence. Here, ϵ, δ are what we call privacy budget parameters and the smaller they are, the stronger the privacy guarantee is. A helpful way to work with differential privacy is in terms of tail bounds on the *privacy loss random variable*. Let $\mathcal{M}(X)$ and $\mathcal{M}(X')$ be the probability distribution induced by $\mathcal{M}$ on neighboring datasets X and X' respectively. The *privacy loss random variable* is defined as: $L_{\mathcal{M}}^{X,X'} := \log(\mathcal{M}(X)(\theta)/\mathcal{M}(X')(\theta))$ where $\theta \sim \mathcal{M}(X)$. Up to constant factors, (ϵ, δ)-DP (Definition 1) is equivalent to requiring that the probability of the privacy loss random variable being greater than ϵ is at most δ for all neighboring datasets X, X'.[3]

Classical design of differentially private mechanisms takes the privacy parameters ϵ, δ as inputs and then the algorithm carefully introduces some randomness to satisfy the privacy constraint (Definition 1), while simultaneously trying to achieve good utility. However, this paradigm has shifted a bit recently as it has become apparent that a more fine-grained analysis tailored for specific mechanisms can yield more favorable privacy-utility trade-offs and better privacy budget parameters under composition (See, e.g., Dwork and Rothblum, 2016; Abadi et al., 2016; Balle and Wang, 2018).

A common technique for achieving differential privacy while working with a real-valued function $f : \mathcal{X}^n \to \mathbb{R}$ is via addition of noise calibrated to f's sensitivity S_f, which is defined as the maximum of the absolute distance $|f(X) - f(X')|$ where X, X' are adjacent inputs.[4] In this paradigm, the Gaussian mechanism is defined as: $\mathcal{G}(X) := f(X) + \mathcal{N}(0, S_f^2 \sigma^2)$. A single application of the Gaussian mechanism to a function f with sensitivity S_f satisfies (ϵ, δ)-differential privacy if[5] $\delta \geq 0.8 \cdot \exp(-(\sigma\epsilon)^2/2)$ and $\epsilon \leq 1$ (Dwork and Roth, 2013, Theorem 3.22)

Stochastic Gradient Descent and Subsampling Lemma. A popular way of designing differentially private machine learning models is to use Stochastic Gradient Descent (SGD) with differentially private releases of (sometimes clipped) gradients evaluated on mini-batches of a dataset (Song et al., 2013; Wang et al., 2015; Bassily et al., 2014; Foulds et al., 2016; Abadi et al., 2016). Algorithmically, these methods are nearly the same and are all based on the NoisySGD idea presented in (1). They differ primarily in how they keep track of their privacy loss. Song et al. (2013) uses a sequence of disjoint mini-batches to ensure each data point is used only once in every data pass. The results in (Bassily et al., 2014; Wang et al., 2016; Foulds et al., 2016) make use of the privacy amplification lemma to take advantage of the randomness introduced by subsampling. The first privacy amplification lemma appeared in (Kasiviswanathan et al., 2011; Beimel et al., 2013), with many subsequent improvements in different settings. For the case of (ϵ, δ)-DP, (Balle et al., 2018) provide a unified account of privacy amplification techniques for different types of subsampling and dataset neighboring relations. In this paper, we work in the subsampling without replacement setup, which satisfies the following privacy amplification lemma for (ϵ, δ)-DP.

Definition 2 (Subsample)**.** *Given a dataset X of n points, the procedure* subsample *selects a random sample from the uniform distribution over all subsets of X of size m. The ratio $\gamma := m/n$ is defined as the sampling parameter of the* subsample *procedure.*

Lemma 3 (Ullman (2017)[6])**.** *If $\mathcal{M}$ is (ϵ, δ)-DP, then the subsampled mechanism $\mathcal{M} \circ \mathsf{subsample}$ obeys (ϵ', δ')-DP with $\epsilon' = \log\big(1 + \gamma(e^{\epsilon} - 1)\big)$ and $\delta' = \gamma\delta$.*

Roughly, the lemma says that subsampling with probability $\gamma < 1$ amplifies an (ϵ, δ)-DP algorithm to an $(O(\gamma\epsilon), \gamma\delta)$-DP algorithm for a sufficiently small choice of ϵ. The overall differential privacy guarantees in (Wang et al., 2015; Bassily et al., 2014; Foulds et al., 2016) were obtained by keeping track of the

[3] For meaningful guarantees, δ is typically taken to be "cryptographically" small.

[4] The restriction to a scalar-valued function is intended to simplify this presentation, but is not essential.

[5] Balle and Wang (2018) show that a more complicated relation between ϵ and δ yields an if and only if statement.

[6] This result follows from Ullman's proof, though the notes state a weaker result. See also (Balle et al., 2018).

privacy loss over each iterative update of the model parameters using the *strong composition theorem* in differential privacy (Dwork et al., 2010), which gives roughly $(\tilde{O}(\sqrt{k}\epsilon), \tilde{O}(k\delta))$-DP[7] for k iterations of an arbitrary (ϵ, δ)-DP algorithm (see Appendix B for a discussion about various composition results for DP).

The work of Abadi et al. (2016) was the first to take advantage of the fact that $\mathcal{M}$ is a subsampled Gaussian mechanism and used a mechanism-specific strong composition result. Their technique, referred to as *moments accountant*, is described below.

Cumulant Generating Functions, Moments Accountant and Rényi Differential Privacy. The moments accountant technique of Abadi et al. (2016) centers around the cumulant generating function (CGF, or the log of the moment generating function) of the privacy loss random variable: $K_{\mathcal{M}}^{X,X'}(\lambda) := \log \mathbb{E}[e^{\lambda L_{\mathcal{M}}^{X,X'}}]$. This function can also be written as[8]:

$$K_{\mathcal{M}}^{X,X'}(\lambda) = \log \mathbb{E}_{\theta\sim\mathcal{M}(X)}\left[\left(\frac{\mathcal{M}(X)(\theta)}{\mathcal{M}(X')(\theta)}\right)^{\lambda}\right]$$
$$= \log \mathbb{E}_{\theta\sim\mathcal{M}(X')}\left[\left(\frac{\mathcal{M}(X)(\theta)}{\mathcal{M}(X')(\theta)}\right)^{\lambda+1}\right].$$

Recall that two random variables with identical CGFs are identically distributed (almost everywhere). In other words, $K_{\mathcal{M}}^{X,X'}(\lambda)$ characterizes the distribution of the privacy loss random variable. We also define the maximum of this function over pairs of neighboring datasets as $K_{\mathcal{M}}(\lambda) := \sup_{d(X,X')\le 1} K_{\mathcal{M}}^{X,X'}(\lambda)$.

Before explaining the details behind the moments accountant technique, we introduce the notion of Rényi differential privacy (RDP) (Mironov, 2017) as a generalization of differential privacy that uses the α-Rényi divergences between $\mathcal{M}(X)$ and $\mathcal{M}(X')$.

Definition 4 (Rényi Differential Privacy)**.** *We say that a mechanism $\mathcal{M}$ is (α, ϵ)-RDP with order $\alpha \in (1, \infty)$ if $D_\alpha(\mathcal{M}(X)\|\mathcal{M}(X')) \le \epsilon$ for all neighboring datasets X, X', where*

$$D_\alpha(\mathcal{M}(X)\|\mathcal{M}(X')) := \frac{1}{\alpha-1} K_{\mathcal{M}}^{X,X'}(\alpha-1).$$

As $\alpha \to \infty$ RDP reduces to $(\epsilon, 0)$-DP (pure DP), i.e., a randomized mechanism $\mathcal{M}$ is $(\epsilon, 0)$-DP if and only if for any two adjacent inputs X and X' it satisfies $D_\infty(\mathcal{M}(X)\|\mathcal{M}(X')) \le \epsilon$. For $\alpha \to 1$, the RDP notion reduces to Kullback-Leibler based privacy notion, which is equivalent to a bound on the expectation of the privacy loss random variable. For a detailed exposition of the guarantee and properties of Rényi differential privacy that mirror those of differential privacy, see (Mironov, 2017, Section III). Here, we highlight two key properties that are relevant for this paper.

Lemma 5 (Adaptive Composition of RDP (Mironov, 2017, Proposition 1))**.** *If $\mathcal{M}_1$ that takes dataset as input obeys (α, ϵ_1)-RDP, and $\mathcal{M}_2$ that takes the dataset and the output of $\mathcal{M}_1$ as input obeys (α, ϵ_2)-RDP, then their composition obeys $(\alpha, \epsilon_1 + \epsilon_2)$-RDP.*

Lemma 6 (RDP to DP conversion (Mironov, 2017, Proposition 3))**.** *If $\mathcal{M}$ obeys (α, ϵ)-RDP, then $\mathcal{M}$ obeys $(\epsilon + \log(1/\delta)/(\alpha-1), \delta)$-DP for all $0 < \delta < 1$.*

RDP Functional View. While RDP for each fixed α can be used as a standalone privacy measure, we emphasize its *functional view* in which ϵ is a function of α for $1 \le \alpha \le \infty$, and this function is completely determined by $\mathcal{M}$. This is denoted by $\epsilon_{\mathcal{M}}(\alpha)$, and with this notation, mechanism $\mathcal{M}$ satisfies $(\alpha, \epsilon_{\mathcal{M}}(\alpha))$-RDP in Definition 4. In other words,

$$\sup_{X,X':d(X,X')\le 1} D_\alpha(\mathcal{M}(X)\|\mathcal{M}(X')) \le \epsilon_{\mathcal{M}}(\alpha).$$

Here $\epsilon_{\mathcal{M}}(\alpha)$ is referred to as the RDP parameter. We use $\epsilon_{\mathcal{M}}(\infty)$ to denote the case where $\alpha = \infty$, which indicates that the mechanism $\mathcal{M}$ is $(\epsilon, 0)$-DP (pure DP) with $\epsilon = \epsilon(\infty)$. We drop the subscript from $\epsilon_{\mathcal{M}}$ when $\mathcal{M}$ is clear from the context.

Our goal is, given a mechanism $\mathcal{M}$ that satisfies $(\alpha, \epsilon(\alpha))$-RDP, to investigate the RDP parameter of the subsampled mechanism $\mathcal{M} \circ \mathsf{subsample}$, i.e., to get a bound on $\epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha)$ such that the mechanism $\mathcal{M} \circ \mathsf{subsample}$ satisfies $(\alpha, \epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha))$-RDP.

Note that $\epsilon_{\mathcal{M}}(\alpha)$ is equivalent to the CGF $K_{\mathcal{M}}(\lambda)$ up to a scaling transformation (with $\alpha = \lambda + 1$) as noted by the following remark.

Remark 7 (RDP $\Leftrightarrow$ CGF)**.** *A randomized mechanism $\mathcal{M}$ obeys $(\lambda + 1, K_{\mathcal{M}}(\lambda)/\lambda)$-RDP for all λ.*

The idea of moments accountant (Abadi et al., 2016) is to essentially keep track of the evaluations of CGF at a list of fixed locations through Lemma 5 and then Lemma 6 allows one to find the smallest ϵ given a desired δ or vice versa using:

$$\delta \Rightarrow \epsilon: \qquad \epsilon(\delta) = \min_{\lambda} \frac{\log(1/\delta) + K_{\mathcal{M}}(\lambda)}{\lambda}, \tag{2}$$
$$\epsilon \Rightarrow \delta: \qquad \delta(\epsilon) = \min_{\lambda} e^{K_{\mathcal{M}}(\lambda) - \lambda\epsilon}. \tag{3}$$

Using the convexity of $K_{\mathcal{M}}(\lambda)$ and monotonicity of $K_{\mathcal{M}}(\lambda)/\lambda$ in λ (Van Erven and Harremos, 2014, Corollary 2, Theorem 3), we observe that the optimization problem in (3) is log-convex and the optimization problem (2) is unimodal/quasi-convex. Therefore, the optimization problem in (2) (similarly, in (3)) can be solved

[7] The $\tilde{O}(\cdot)$ notation hides various logarithmic factors.

[8] The second identity follows from a change of measure.

to an arbitrary accuracy τ in time $\log(\lambda^*/\tau)$ using the bisection method, where λ^* is the optimal value for λ from (2) (similarly, (3)). The same result holds even if all we have is (possibly noisy) blackbox access to $K_{\mathcal{M}}(\cdot)$ or its derivative (see more details in Appendix G).

For other useful properties of the CGF and an elementary proof of its convexity and how it implies the monotonicity of the Rényi divergence, see Appendix H.

Other Related Work. A closely related notion to RDP is that of *zero-concentrated differential privacy* (zCDP) introduced in (Bun and Steinke, 2016) (see also (Dwork and Rothblum, 2016)). zCDP is related to CGF of the privacy loss random variable as we note here.

Remark 8 (Relation between CGF and Zero-concentrated Differential Privacy)**.** *If randomized mechanism $\mathcal{M}$ obeys (ξ,ρ)-zCDP for some parameters ξ,ρ, then the CGF satisfies $K_{\mathcal{M}}(\lambda)\leq\lambda\xi+\lambda(\lambda+1)\rho$. On the other hand, if $\mathcal{M}$'s privacy loss r.v. has CGF $K_{\mathcal{M}}(\lambda)$, then $\mathcal{M}$ is also (ξ,ρ)-zCDP for all (ξ,ρ) such that the quadratic function $\lambda\xi+\lambda(\lambda+1)\rho\geq K_{\mathcal{M}}(\lambda)$.*

In general, the RDP view of privacy is broader than the CDP view as it captures finer information. For CDP, subsampling does not improve the privacy parameters (Bun et al., 2018). A truncated variant of the zCDP has been very recently proposed by Bun et al. (2018) and they studied the effect of subsampling in tCDP. While this independent work attempts to solve a problem closely related to ours, they are not directly comparable in that they deal with the amplification properties of tCDP while we deal with that of Rényi DP (and therefore CDP without truncation). A simple consequence of this difference is that the popular subsampled Gaussian mechanism explained above, that is covered by our analysis, is not directly covered by the amplification properties of tCDP.

3 Our Results

In this section, we present our main result, an amplification theorem for Rényi Differential Privacy via subsampling. We first provide the upper bound, and then discuss the optimality of this bound. Based on these bounds, in Section 3.2, we discuss an implementation of a data structure that can efficiently track RDP privacy parameters under composition.

3.1 Privacy Amplification for RDP

We start with our main theorem that bounds $\epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha)$ for the mechanism $\mathcal{M}\circ\mathsf{subsample}$ in terms of $\epsilon_{\mathcal{M}}(\alpha)$ of the mechanism $\mathcal{M}$ and sampling parameter γ used in the $\mathsf{subsample}$ procedure. Missing details from this Section are collected in Appendix C.

Theorem 9 (RDP for Subsampled Mechanisms)**.** *Given a dataset of n points drawn from a domain $\mathcal{X}$ and a mechanism $\mathcal{M}$ that takes an input from $\mathcal{X}^m$ for $m\leq n$, let the randomized algorithm $\mathcal{M}\circ\mathsf{subsample}$ be defined as: (1) subsample without replacement m datapoints of the dataset (sampling parameter $\gamma=m/n$), and (2) apply $\mathcal{M}$ to the subsampled dataset. For all integers $\alpha\geq 2$, if $\mathcal{M}$ obeys $(\alpha,\epsilon(\alpha))$-RDP, then the subsampled mechanism $\mathcal{M}\circ\mathsf{subsample}$ obeys $(\alpha,\epsilon'(\alpha))$-RDP where,*

$$\begin{aligned}\epsilon'(\alpha)\leq\frac{1}{\alpha-1}\log\Bigg(1+\gamma^2\binom{\alpha}{2}\min\Big\{4(e^{\epsilon(2)}-1),\\ e^{\epsilon(2)}\min\{2,(e^{\epsilon(\infty)}-1)^2\}\Big\}\\ +\sum_{j=3}^{\alpha}\gamma^j\binom{\alpha}{j}e^{(j-1)\epsilon(j)}\min\{2,(e^{\epsilon(\infty)}-1)^j\}\Bigg).\end{aligned}$$

The bound in the above theorem might appear complicated, and this is partly because of our efforts to get a precise non-asymptotic bound (and not just a $O(\cdot)$ bound). Some additional practical considerations related to evaluating the bound in this theorem such as computational resources needed, numerical stability issues, etc., are discussed in Appendix G. The phase transition behavior of this bound, noted in the introduction, is probably most easily observed through Figure 3 (Appendix A), where we empirically illustrates the behavior of this bound for the commonly used subsampled mechanisms. Before discussing the proof idea, we make a few remarks about this result.

Generality. Our results cover any Rényi differentially private mechanism, including those based on any exponential family distribution (see Geumlek et al., 2017, and our exposition in Appendix I). As mentioned earlier, previously such a bound (even asymptotically) was only known for the special case of the subsampled Gaussian mechanism (Abadi et al., 2016).

Pure DP. In particular, Theorem 9 also covers pure-DP mechanisms (such as Laplace and randomized response mechanisms) with a bounded $\epsilon(\infty)$. In this case, we can upper bound everything within the logarithm of Theorem 9 with a binomial expansion:

$$1+\sum_{j=1}^{\alpha}\gamma^j\binom{\alpha}{j}e^{j\epsilon(\alpha)}(e^{\epsilon(\infty)}-1)^j=\big(1+\gamma e^{\epsilon(\alpha)}(e^{\epsilon(\infty)}-1)\big)^{\alpha},$$

which results in a bound of the form

$$\epsilon'(\alpha)\leq\frac{\alpha}{\alpha-1}\log\big(1+\gamma e^{\epsilon(\alpha)}(e^{\epsilon(\infty)}-1)\big).$$

This bound converges to $\log\big(1+\gamma e^{\epsilon(\infty)}(e^{\epsilon(\infty)}-1)\big)$ as $\alpha\to\infty$, which gives quantitatively the same result as the privacy amplification result in Lemma 3 for the pure $(\epsilon,0)$-DP, modulo an extra $e^{\epsilon(\infty)}$ factor which becomes negligible as $\epsilon(\infty)$ gets smaller.

Bound under Additional Assumptions. The bound in Theorem 9 could be strengthened under additional assumptions on the RDP guarantee. We defer a detailed discussion on this topic to Appendix C.2 (see Theorem 21), but note that a consequence of this is that one can replace $e^{(j-1)\epsilon(j)}\min\{2,(e^{\epsilon(\infty)}-1)^j\}$ in the above bound with an exact evaluation given by the forward finite difference operator of some appropriately defined functional. We also note that these additional assumptions hold for the Gaussian mechanism.

In particular, with subsampled Gaussian mechanism for functions with sensitivity 1 (i.e., $\epsilon(\alpha)=\alpha/(2\sigma^2)$) the dominant part of the upper bound on $\epsilon'(\alpha)$ arises from the term $\min\{4(e^{\epsilon(2)}-1), e^{\epsilon(2)}\min\{2,(e^{\epsilon(\infty)}-1)^2\}\}$. Firstly, since the Gaussian mechanism does not have a bounded $\epsilon(\infty)$ term, this term can be simplified as $\min\{4(e^{\epsilon(2)}-1), 2e^{\epsilon(2)}\}$. Let us consider the regimes: (a) σ^2 "large", (b) σ^2 "small". When σ^2 is large, $4(e^{\epsilon(2)}-1)=4(e^{1/\sigma^2}-1)\leq 8/\sigma^2$ becomes the dominant term in $\min\{4(e^{\epsilon(2)}-1), 2e^{\epsilon(2)}\}$. In this case, for small α and γ, the overall $\epsilon'(\alpha)$ bound simplifies to $O(\gamma^2\alpha/\sigma^2)$ (matching the asymptotic bound given in Appendix C.7). When σ^2 is small, then $2e^{\epsilon(2)}=2e^{1/\sigma^2}$ becomes the dominant term in $\min\{4(e^{\epsilon(2)}-1), 2e^{\epsilon(2)}\}$. Note the small σ^2 regime is not covered by the results of Abadi et al. (2016).

Integer to Real-valued α. The above calculations rely on a binomial expansion and thus only work for integer α's. To apply it to any real-valued, we can use the relation between RDF and CGF mentioned in Remark 7, and the fact that CGF is a convex function (see Lemma 37 in Appendix H). The convexity of $K_{\mathcal{M}}(\cdot)$ implies that a piecewise linear interpolation yields a valid upper bound for all $\alpha\in(1,\infty)$.

Corollary 10. *Let $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denotes the floor and ceiling operators. Then, $K_{\mathcal{M}}(\lambda)\leq(1-\lambda+\lfloor\lambda\rfloor)K_{\mathcal{M}}(\lfloor\lambda\rfloor)+(\lambda-\lfloor\lambda\rfloor)K_{\mathcal{M}}(\lceil\lambda\rceil)$.*

The bound on $K_{\mathcal{M}}(\lambda)$ can be translated into a RDP parameter bound as noted in Remark 7.

Proof Idea. The proof of Theorem 9 is given in Appendix C.1; here we give an overview of the proof strategy. Let $\mathcal{M}'=\mathcal{M}\circ\textsf{subsample}$, $p(\theta)=\mathcal{M}'(X)(\theta)$ and $q(\theta)=\mathcal{M}'(X')(\theta)$. The starting observation is that, by the binomial expansion, for integer $\alpha\geq 2$ the CGF $K^{X,X'}_{\mathcal{M}'}(\alpha-1)$ can be written as

$$\mathbb{E}_q\left[\left(\frac{p}{q}\right)^\alpha\right]=\sum_{j=0}^{\alpha}\binom{\alpha}{j}\mathbb{E}_q\left[\left(\frac{p}{q}-1\right)^j\right]$$
$$\leq 1+\sum_{j=2}^{\alpha}\binom{\alpha}{j}\mathbb{E}_q\left[\left|\frac{p}{q}-1\right|^j\right],$$

where we used $\mathbb{E}_q[\frac{p}{q}-1]=0$. Given this expansion, we note that the expectation terms correspond to the classical Pearson-Vajda divergence of order j between p and q; the rest of the analysis then reduces to bounding these divergences in terms of the Rényi DP parameter of $\mathcal{M}$ for any neighboring X and X'. This is achieved by introducing a *ternary* divergence between triples of distributions, and showing that a new privacy definition (*ternary-$|\chi|^\alpha$-DP*) in terms of our ternary divergence exhibits a simple amplification by subsampling property (Proposition 16, Appendix C.1). To leverage this powerful observation we propose a number of ways of converting a ternary-$|\chi|^\alpha$-differential privacy guarantee back to RDP (Lemmas 17, 18, 19, Appendix C.1). Each of these conversion strategies yield different coefficients in the sum inside the logarithm defining $\alpha'(\epsilon)$; our bound accounts for all these strategies at once by taking the minimum of these coefficients.

A Lower Bound. We now discuss whether the bound in Theorem 9 can be improved. First, we provide a short answer: it cannot be improved in general.

Proposition 11. *Let $\mathcal{M}$ be a randomized algorithm that takes a dataset in $\mathcal{X}^{\gamma n}$ as an input. If $\mathcal{M}$ obeys $(\alpha,\epsilon(\alpha))$-RDP for a function $\epsilon:\mathbb{R}_+\to\mathbb{R}_+$ and that there exists $x,x'\in\mathcal{X}$ such that $\epsilon(\alpha)=D_\alpha(\mathcal{M}([x,\ldots,x,x'])\|\mathcal{M}([x,\ldots,x,x]))$ for all integer $\alpha\geq 1$ (e.g., this condition is true for all output perturbation mechanisms for counting queries), then the RDP function ϵ' for $\mathcal{M}\circ\textsf{subsample}$ obeys the following lower bound for all integers $\alpha\geq 1$:*

$$\epsilon'(\alpha)\geq\frac{\alpha}{\alpha-1}\log(1-\gamma)+\frac{1}{\alpha-1}\log\left(1+\alpha\frac{\gamma}{1-\gamma}\right.$$
$$\left.+\sum_{j=2}^{\alpha}\binom{\alpha}{j}\left(\frac{\gamma}{1-\gamma}\right)^j e^{(j-1)\epsilon(j)}\right).$$

The idea behind the proof of this result is simple: for datasets of size n of the form $X=[x,\ldots,x,x']$ and $X'=[x,\ldots,x,x]$ it is possible to apply a variation of the binomial expansion in the proof of Theorem 9 to obtain a closed-form expression for $D_\alpha(\mathcal{M}'(X)\|\mathcal{M}'(X'))$ in terms of $D_j(\mathcal{M}(\tilde{X})\|\mathcal{M}(\tilde{X}'))$ for $j=2,\ldots,\alpha$. Here $\tilde{X}$ and $\tilde{X}'$ are datasets obtained by removing the first $(1-\gamma)n$ elements from X and X' respectively. See Appendix C.6 for the detailed calculation.

Let us compare the above lower bound to our upper bound in Theorem 9 in two regimes. When $\alpha\gamma e^{\epsilon(\alpha)}\ll 1$, such that $\alpha^2\gamma^2 e^{\epsilon(2)}<1$ is the dominating factor in the summation, we can use the bounds $x/(1+x)\leq\log(1+x)\leq x$ to get that both the upper and lower bound are $\Theta(\alpha\gamma^2 e^{\epsilon(2)})$. In other words, they match up to a constant multiplicative factor. For other parameter configurations, note that $\gamma/(1-\gamma)>\gamma$, our bound in Theorem 9 (with the $2e^{(j-1)\epsilon(j)}$) is tight up to an additive factor $\frac{\alpha}{\alpha-1}\log((1-\gamma)^{-1})+\frac{\log(2)}{\alpha-1}$ which goes to 0 as $\gamma \to 0$ and $\alpha \to \infty$. We provide explicit comparisons of the upper and lower bounds in the numerical experiments presented in Section 4.

The longer answer to this question of optimality is more intricate. The RDP bound can be substantially improved when we consider more fine-grained per-instance RDP in the same flavor as the per-instance (ϵ, δ)-DP (Wang, 2018). The only difference from the standard RDP is that now ϵ is parameterized by a pair of fixed adjacent datasets. This point is illustrated in Appendix C.7, where we discuss an asymptotic approximation of the Rényi divergence for the subsampled Gaussian mechanism.

3.2 Analytical Moments Accountant

The results above allow us to build an *analytical moments accountant* supporting composition and subsampling of differentially private mechanisms. This is a data structure that tracks the CGF function of a sequence of mechanisms in symbolic form, some of which can be subsampled mechanisms. The data structure allows data analysts to query the smallest ϵ from a given δ (or vice versa) for (ϵ, δ)-DP using (2) (or (3)).

More formally, the data structure maintains the CGF $K = K_{\mathcal{M}_1} + \ldots + K_{\mathcal{M}_k}$ corresponding to the composition of the (potentially adaptive) sequence of mechanisms $\mathcal{M}_1, \mathcal{M}_2, .., \mathcal{M}_k$ applied to the same dataset. The function K is represented in symbolic form, and the data structure offers three operations: (i) composition of an additional mechanism to update K; (ii) evaluation of the RDP parameter $\epsilon(\alpha)$ of the composed mechanism; and (iii) conversion to an (ϵ, δ)-DP guarantee. The conversion to RDP is straightforward using the one-to-one relationship between CGF and RDP (see Remark 7) with the exception of RDP at $\alpha = 1$ (Kullback-Leibler privacy) and $\alpha = +\infty$ (pure DP), which we keep track of separately. The conversion to (ϵ, δ)-DP is obtained by solving the univariate optimization problems described in (2) and (3).

We design the data structure to be numerically stable, and efficient in both space and time. In particular, it tracks CGFs with $O(1)$ time to compose a new mechanism and uses space only linear in the number of *unique* mechanisms applied (rather than the number of total mechanisms applied). Using the convexity of CGFs and the monotonicity of RDP, we are able to provide $\delta \Rightarrow \epsilon$ conversion to (ϵ, δ)-DP to within accuracy τ in oracle complexity $O(\log(\lambda^*/\tau))$, where λ^* is the optimal value for λ. Similarly, for $\epsilon \Rightarrow \delta$ queries.

Note that for subsampled mechanisms the direct evaluation $\epsilon_{\mathcal{M}\circ\mathsf{subsample}}(\alpha)$ of the upper bounds in Theorem 9 is already polynomial in α. To make the data structure truly scalable, we devise a number of ways to approximate the bounds that takes only $O(\log(\alpha))$ evaluations of $\epsilon_{\mathcal{M}}(\cdot)$. More details about our analytical moments accountant and substantiations to the above claims are provided in Appendix G.

Practically, our analytical moments accountant is better than the moment accountants proposed by Abadi et al. (2016) in several noteworthy ways: (1) it allows one to keep track the CGF's of all $\lambda \geq 1$ in symbolic form without paying infinite memory, whereas moments account (Abadi et al., 2016) requires a predefined list of λ's and pays a memory proportional to the size of the list; (2) it completely avoids numerical integration used by moments account; and finally (3) it supports subsampling for generic RDP mechanisms while the original moments accountant was built for supporting only Gaussian mechanisms. All of this translates into an efficient and accurate way for tracking ϵ's and δ's when composing differentially private mechanisms.

4 Experimental Evaluation

In this section, we present numerical experiments to demonstrate our upper and lower bounds of RDP for subsampled mechanisms and the usage of analytical moments accountant. Due to space constraints, we only consider the Gaussian mechanism here; further experiments with Laplace and randomized response mechanisms are presented in Appendix A.

The RDP guarantees of the Gaussian mechanism are given by $\epsilon_{\text{Gaussian},\sigma}(\alpha) = \alpha/2\sigma^2$, where σ^2 represents the variance of the Gaussian perturbation[9]. The experiments in this section are performed under the "low privacy regime" $\sigma = 5$, which corresponds to $(0.2\sqrt{2\log(1.25/\delta)}, \delta)$-DP; the "high privacy regime" is considered in Appendix A. The subsampling ratio γ is taken to be 0.001.

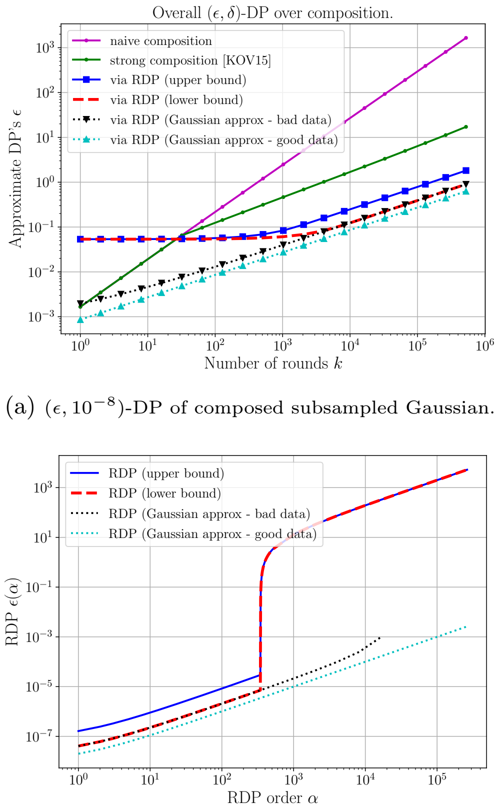
In Figure 1a, we plot the overall (ϵ, δ)-DP for $\delta = 10^{-8}$ as we compose the subsampled Gaussian mechanism 600,000 times. The ϵ is obtained as a function of δ for each k separately by calling the $\delta \Rightarrow \epsilon$ query in our analytical moments accountant. Our results are compared to the algorithm-independent techniques for differential privacy including naïve composition and strong composition. The strong composition baseline is carefully calibrated for each k by choosing an appropriate pair of $(\tilde{\epsilon}, \tilde{\delta})$ for $\mathcal{M}$ such that the overall (ϵ, δ)-DP guarantee that comes from composing k rounds of $\mathcal{M} \circ \mathsf{subsample}$ using Kairouz et al. (2015) obeys that $\delta < 10^{-8}$ and ϵ is minimized. Each round is described by the $(\log(1 + \gamma(e^{\tilde{\epsilon}} - 1)), \gamma\tilde{\delta})$-DP guarantee using the standard subsampling lemma (Lemma 3) and $\tilde{\epsilon}$ is obtained as a function of $\tilde{\delta}$ via (2).

Not surprisingly, both our approach and strong compo-

[9] We assume the underlying function has sensitivity 1.

sition give a $O(\sqrt{k})$ scaling while the naïve composition exhibits $O(k)$ scaling. An interesting observation for the subsampled Gaussian mechanism is that the RDP approach initially performs worse than the naïve composition and strong composition with the standard subsampling lemma. Our RDP lower bound certifies that this is not due to an artifact of our analysis but rather a fundamental limitation of the approach that uses RDP to obtain (ϵ, δ)-DP guarantees. We believe this is a manifestation of the same phenomenon that leads to the sub-optimality of the classical analysis of the Gaussian mechanism (Balle and Wang, 2018), which also relies on the conversion of a bound on the CGF of the privacy loss into an (ϵ, δ)-DP guarantee, and might be addressed using the necessary and sufficient condition for (ϵ, δ)-DP in terms of tail probabilities of the privacy loss random variable given in (Balle and Wang, 2018, Theorem 5). Luckily, such an artifact does not affect the typical usage of RDP: as the number of rounds of composition continues to grow, we end up having about five orders of magnitude smaller ϵ.

We also conducted numerical comparisons between the upper and lower bounds for $\epsilon'(\alpha)$ given in Theorem 9 and Proposition 11 respectively. Our evaluation shows these bounds are tight up to constant factors. As we can see from the Figure 1b, the upper and lower bounds match up to a multiplicative constant. There is a phase transition in both the upper and lower bound, which occurs at about $\gamma \alpha e^{\epsilon(\alpha)} \approx 1$. We also plot an asymptotic approximation obtained under the assumption that the size of the input dataset grows $n \to \infty$ while the subsampling ratio $\gamma = m/n$ is kept constant. In fact, we derive two asymptotic approximations: one in the case of "good" data and one for "bad" data. The approximations and the definitions of "good" and "bad" data can be found in Appendix C.7. The "bad" data approximation matches almost exactly with lower bound up to the phase transition point. The Gaussian approximation for the "good" data is smaller than the lower bound, especially in the low-privacy regime, highlighting that we could potentially obtain additional gains by performing a dataset dependent analysis.



(a) $(\epsilon, 10^{-8})$-DP of composed subsampled Gaussian.

(b) $\epsilon(\alpha)$-RDP of subsampled Gaussian.

Figure 1: Experimental results for subsampled Gaussian mechanism with $\sigma = 5$ and $\gamma = 0.001$. (a) DP guarantees after composition of k mechanisms obtained with RDP composition, strong composition and naïve composition. (b) RDP guarantees as a function of α obtained with our upper and lower bounds, as well as asymptotic approximations for two data regimes.

5 Conclusion

In this paper, we have studied the effect of subsampling (without replacement) in amplifying Rényi differential privacy (RDP). Specifically, we established a tight upper and lower bound for the RDP parameter for the randomized algorithm $\mathcal{M} \circ \mathsf{subsample}$ that first subsamples the data set then applies $\mathcal{M}$ to the subsample, in terms of the RDP parameter of $\mathcal{M}$. In addition, we designed a data structure called *analytical moments accountant* which composes RDP for randomized algorithm (including subsampled ones) in symbolic forms and allows efficiently conversion of RDP to (ϵ, δ)-DP for any δ (or ϵ) of choice. These results substantially expands the scope of the mechanisms with RDP guarantees to cover subsampled versions of Gaussian mechanism, Laplace mechanism, Randomized Responses, posterior sampling and so on, which facilitates flexible differentially private algorithm design. We compared our approach to the standard approaches that use subsampling lemma on (ϵ, δ)-DP directly and then applies strong composition, and in our experiments we notice an order of magnitude improvement in the privacy parameters with our bounds when we compose the subsampled Gaussian mechanism over multiple rounds.

Bibliography

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep learning with differential privacy. In *ACM SIGSAC Conference on Computer and Communications Secu-*

rity (CCS-16), pages 308–318. ACM.

Apple, D. (2017). Learning with privacy at scale. *Apple Machine Learning Journal*.

Balle, B., Barthe, G., and Gaboardi, M. (2018). Privacy amplification by subsampling: Tight analyses via couplings and divergences. In *NeurIPS*.

Balle, B. and Wang, Y.-X. (2018). Improving gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. *International Conference in Machine Learning (ICML)*.

Bassily, R., Smith, A., and Thakurta, A. (2014). Private empirical risk minimization: Efficient algorithms and tight error bounds. In *Foundations of Computer Science (FOCS-14)*, pages 464–473. IEEE.

Beimel, A., Nissim, K., and Stemmer, U. (2013). Characterizing the sample complexity of private learners. In *Conference on Innovations in Theoretical Computer Science (ITCS-13)*, pages 97–110. ACM.

Bobkov, S., Chistyakov, G., and Götze, F. (2016). R\'enyi divergence and the central limit theorem. *arXiv preprint arXiv:1608.01805*.

Bun, M., Dwork, C., Rothblum, G. N., and Steinke, T. (2018). Composable and versatile privacy via truncated cdp. In *to appear in STOC-18*.

Bun, M., Nissim, K., Stemmer, U., and Vadhan, S. (2015). Differentially private release and learning of threshold functions. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 634–649. IEEE.

Bun, M. and Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer.

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. (2006a). Our data, ourselves: Privacy via distributed noise generation. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer.

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

Dwork, C. and Roth, A. (2013). The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 9(3-4):211–407.

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

Dwork, C., Rothblum, G. N., and Vadhan, S. (2010). Boosting and differential privacy. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 51–60. IEEE.

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

Foulds, J., Geumlek, J., Welling, M., and Chaudhuri, K. (2016). On the theory and practice of privacy-preserving bayesian data analysis. In *Conference on Uncertainty in Artificial Intelligence (UAI-16)*, pages 192–201. AUAI Press.

Geumlek, J., Song, S., and Chaudhuri, K. (2017). Renyi differential privacy mechanisms for posterior sampling. In *Advances in Neural Information Processing Systems*, pages 5295–5304.

Gil, M., Alajaji, F., and Linder, T. (2013). Rényi divergence measures for commonly used univariate continuous distributions. *Information Sciences*, 249:124–131.

Kairouz, P., Oh, S., and Viswanath, P. (2015). The composition theorem for differential privacy. In *International Conference on Machine Learning (ICML-15)*.

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

Lukacs, E. (1970). *Characteristic functions*. Griffin.

Mironov, I. (2017). Rényi differential privacy. In *Computer Security Foundations Symposium (CSF), 2017 IEEE 30th*, pages 263–275. IEEE.

Murtagh, J. and Vadhan, S. (2016). The complexity of computing the optimal composition of differential privacy. In *Theory of Cryptography Conference*, pages 157–175. Springer.

Nielsen, F. and Nock, R. (2014). On the chi square and higher-order chi distances for approximating f-divergences. *IEEE Signal Processing Letters*, 21(1):10–13.

Song, S., Chaudhuri, K., and Sarwate, A. D. (2013). Stochastic gradient descent with differentially private updates. In *Conference on Signal and Information Processing*.

Uber Security (2017). Uber releases open source project for differential privacy. https://medium.com/uber-security-privacy/differential-privacy-open-source-7892c82c42b6.

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

Vajda, I. (1973). χ^{α}-divergence and generalized fisher information. In *Prague Conference on Information*

Theory, Statistical Decision Functions and Random Processes, page 223. Academia.

Van Erven, T. and Harremos, P. (2014). Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820.

Wang, Y.-X. (2018). Per-instance differential privacy. *Journal of Confidentiality and Privacy, to appear.*

Wang, Y.-X., Fienberg, S., and Smola, A. (2015). Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *International Conference on Machine Learning (ICML-15)*, pages 2493–2502.

Wang, Y.-X., Lei, J., and Fienberg, S. E. (2016). Learning with differential privacy: Stability, learnability and the sufficiency and necessity of erm principle. *Journal of Machine Learning Research*, 17(183):1–40.