Optimal Differential Privacy Composition for Exponential Mechanisms

Jinshuo Dong [*1 2] **David Durfee** [*2] **Ryan Rogers** [*2]

Abstract

Composition is one of the most important properties of differential privacy (DP), as it allows algorithm designers to build complex private algorithms from DP primitives. We consider precise composition bounds of the overall privacy loss for exponential mechanisms, one of the fundamental classes of mechanisms in DP. Exponential mechanism has also become a fundamental building block in private machine learning, e.g. private PCA and hyperparameter selection. We give explicit formulations of the optimal privacy loss for both the adaptive and non-adaptive composition of exponential mechanism. For the non-adaptive setting in which each mechanism has the same privacy parameter, we give an efficiently computable formulation of the optimal privacy loss. In the adaptive case, we derive a recursive formula and an efficiently computable upper bound. These precise understandings about the problem lead to a 40% saving of the privacy budget in a practical application. Furthermore, the algorithm-specific analysis shows a difference in privacy parameters of adaptive and non-adaptive composition, which was widely believed to not exist based on the evidence from general analysis.

1. Introduction

Differential privacy (DP) has emerged as the leading privacy benchmark in machine learning as well as data analytics on sensitive data sets. The basic idea is to inject noise into training algorithms so that it masks individual level of information while still preserves statistical efficiency. One of the fundamental DP primitives is exponential mechanism, which has various application in machine learning, such as hyper parameter selection (Liu and Talwar, 2019), private PCA (Chaudhuri et al., 2013), synthetic data generation (Hardt and Rothblum, 2010), and so on. As a building block, it is often used repeatedly/recursively. The overall privacy guarantee was often handled by a general composition theorem in DP, which is one of the most important features of DP that leads to its success. However, the general theorem (Dwork et al., 2010) is suboptimal and not making use of any specific structure of the algorithm, hence often underestimates the true privacy guarantee of the algorithm. A possible negative consequence is algorithm designers may decide to inject more noise until general composition theorem says its private, hence incurring unnecessary drop in model accuracy. If the precise privacy guarantee was known, we could have added the exact amount of noise for desired level of privacy, and not waste any model accuracy.

That being said, exact and optimal privacy characterization is important for private machine learning. Recently, there have been extensive works in this direction. Kairouz et al. (2017); Murtagh and Vadhan (2016) provides optimal composition theorem when all that is know is each component being (ε, δ)-DP.

Although these *black box* composition theorems give optimal privacy parameters over multiple rounds of general DP algorithms, one should be able to improve by making use of the specific structure of the components. That is, *white box* composition can further improve privacy analysis. An example of this type of analysis is the *moments accountant* considered in Abadi et al. (2016). They analyzed the privacy of noisy stochastic gradient descent as a composition of subsampled Gaussian mechanism and achieved the first reasonably private MNIST classifier without significant drop in accuracy compared to non-private baselines. However, no optimality is known for their privacy analysis.

Dong et al. (2019) invented techniques that manage to exhibit exact and optimal analysis of white box composition for a wide range of algorithms, including most building block algorithms of DP such as Laplace mechanism, Gaussian mechanism and their subsampled versions (Bu et al., 2019). However, their technique does not directly apply to exponential mechanism. Durfee and Rogers (2019) introduced *bounded range* (BR) as a property for DP algorithms

[*]Equal contribution [1]Applied Mathematics and Computational Sciences, University of Pennsylvania [2]Data Science Applied Research, LinkedIn. Correspondence to: Jinshuo Dong <djs.pku@gmail.com>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

and performed improved analysis for composition of exponential mechanisms.

The primary focus of this paper is to complement the story of white-box composition by answering the following question: **what is the optimal DP composition bound over the class of exponential mechanisms?**

Surprisingly, the answer to this question depends on whether the exponential mechanisms is adaptively chosen at each round or not. This is the first of its kind in the context of DP composition. Because of this reason we provide two main results: a formula of the optimal parameters in the non-adaptive case, and an iterative algorithm to compute optimal parameters in the adaptive case.

Both the non-adaptive and adaptive setting will have practical importance. The most straightforward one is the setting considered in (Durfee and Rogers, 2019), where their private top-k algorithms (main ingredient of which is exponential mechanisms) is repeatedly used to answer queries like "What are the most popular articles in the last 30 days among data scientists working in Bay area?" In particular, the non-adaptive formula can be applied in a *dashboard setting*, where the set of queries is predetermined. The iterative algorithm for the adaptive setting can be applied in an *API setting*, where the analyst adaptively interacts with the DP system. It is important to be aware of the distinction and use optimal parameters respectively.

While the improvement we give here in the overall privacy parameters are not asymptotically significant, it increases the number of allowable queries by a constant factor when the privacy parameters for each of the queries are considered fix. This can have a substantial impact on practical deployments. From our results in Figure 1, our non-adaptive composition formula allows for about four times more queries than what black-box composition. Furthermore, this optimal composition allows for about two times more queries than the improved bounds given in (Durfee and Rogers, 2019). Additionally, in some settings our improvement for the adaptive composition bound of exponential mechanisms allows for about three times more queries than both the optimal composition for DP mechanisms and the improved bounds in (Durfee and Rogers, 2019).

Remark Although we have presented the exponential mechanism as a specific DP mechanism, it is also important to discuss its generality. In particular, there is the folklore knowledge that any (pure) DP mechanism can be written in terms of an exponential mechanism with a particular quality score, i.e. the log-density of the mechanism (McSherry and Talwar, 2007). Hence, it might seem that exponential mechanisms and pure DP mechanisms are the same thing. As a consequence, optimal composition over the two classes might also seem to be identical and hence well-understood because of the works of Kairouz et al. (2017) and Murtagh and Vadhan (2016). Realizing the distinction of the two classes is an important first step of our improvement. We will clarify this in Section 2.2.

1.1. Our Contributions

We summarize our main contributions here. They will be formally stated in later sections once we have set up the requisite notation.

1. As remarked above, we properly identify and parametrize the class of exponential mechanisms. This is explained in Section 2.2.

2. In the non-adaptive (NA) model, for any $\varepsilon \geqslant 0$, there is a smallest δ such that all k-fold non-adaptive composition of properly parametrized exponential mechanisms satisfy (ε, δ)-DP. Denote this smallest δ by $\delta_k^{\mathrm{NA}}(\varepsilon)$. We give an explicit formula of $\delta_k^{\mathrm{NA}}(\varepsilon)$.

3. Similarly in the adaptive (A) model, let $\delta_k^{\mathrm{A}}(\varepsilon)$ be the smallest δ such that all k-fold adaptive composition of properly parametrized exponential mechanisms satisfy (ε, δ)-DP. We give an recursive formula of $\delta_k^{\mathrm{A}}(\varepsilon)$.

4. Beyond these optimality results, we provide an efficient approach to bound the adaptive parameter $\delta_k^{\mathrm{A}}(\varepsilon)$, and compare all these findings numerically with previous results.

2. Preliminaries

The two central concepts of this paper are composition and exponential mechanism. We settle the definition and notation in this section.

We first cover the standard differential privacy definition from (Dwork et al., 2006b;a), where we will say that two datasets $x, x' \in \mathcal{X}$ are neighbors if they differ in the addition or deletion of one individual's data, sometimes denoted as $x \sim x'$.

Definition 2.1. *A mechanism $M : \mathcal{X} \to \mathcal{Y}$ is (ε, δ)-differentially-private (DP) if the following holds for any neighboring dataset x, x' and $S \subseteq \mathcal{Y}$:*

$$\Pr[M(x) \in S] \leqslant \mathrm{e}^{\varepsilon} \Pr[M(x') \in S] + \delta.$$

Also if $\delta = 0$, we simply write ε-DP.

We remark that all results remain valid for other neighboring relations such as replacing individuals, except the claim that involves counting queries (see the discussion following Proposition 2).

2.1. Composition

We will consider two kinds of composition: adaptive and non-adaptive. The following presentation follows the recent line of work (Mironov, 2017; Bun and Steinke, 2016; Dong et al., 2019).

In the non-adaptive case, component mechanisms are $M_i : X \to Y_i, i = 1, 2, \ldots, k$. Since M_i are randomized, $M_i(x)$ is a distribution over Y_i. The composition $M : X \to Y_1 \times \cdots \times Y_k$ simply outputs an outcome sampled from the product distribution $M_1(x) \times \cdots \times M_k(x)$.

In the adaptive case, component mechanisms take in previous outputs and look like $M_i : X \times Y_1 \times \cdots \times Y_{i-1} \to Y_i, i = 1, 2, \ldots, k$. The output $(y_1, y_2, \ldots, y_k)$ of the composition $M : X \to Y_1 \times \cdots \times Y_k$ satisfies

$$\begin{aligned} y_1 &= M_1(x), \\ y_2 &= M_2(x, y_1), \\ &\cdots \\ y_k &= M_k(x, y_1, y_2, \ldots, y_{k-1}). \end{aligned}$$

2.2. Exponential Mechanism and Bounded Range Property

We are interested in composition where each component comes from the class of exponential mechanisms, but as noted in Section 1, we have to parameterize properly. It turns out that this is a tricky question. We will in the end identify the class of interest as the class of range bounded mechanisms, first introduced in (Durfee and Rogers, 2019).

Recall that exponential mechanism is defined in terms of a quality score $u : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Traditionally, the sensitivity of the quality score $\Delta u := \max_{y \in \mathcal{Y}} \max_{x \sim x'} |u(x, y) - u(x', y)|$ plays a crucial role in the theory. Here we take a slightly less traditional presentation by moving the sensitivity from the mechanism to the privacy guarantee.

Definition 2.2 (Exponential Mechanism (McSherry and Talwar, 2007))**.** *A randomized algorithm $M_{u,\varepsilon} : X \to Y$ is called the exponential mechanism with quality score u and parameter ε, if the outcome y is sampled with probability proportional to $\mathrm{e}^{\varepsilon u(x,y)}$, i.e.*

$$P[M_{u,\varepsilon}(x) = y] = \frac{\mathrm{e}^{\varepsilon u(x,y)}}{\sum_{y \in Y} \mathrm{e}^{\varepsilon u(x,y)}}.$$

We know from (McSherry and Talwar, 2007) that

Theorem 1. *$M_{u,\varepsilon}$ is $2\Delta u \cdot \varepsilon$-DP .*

Remember our goal is to study optimal composition of exponential mechanisms. For that purpose, we would like to focus on the case when every component is in the following class:

$$\begin{aligned} \{&\text{Exponential mechanisms} \mid \text{parameter is } \varepsilon \\ &\text{quality score has sensitivity} \leqslant L\} \end{aligned}$$

However, the class is ambiguous since the quality score is not uniquely determined by the mechanism. The following two clarification both make sense but often cause confusion. For simplicity let us assume $L = 1$.

$$\begin{aligned} \mathbb{M}_1^\varepsilon = \big\{&M : X \to Y \mid \Delta v \leqslant 1 \text{ where} \\ &v(x, y) = \tfrac{1}{\varepsilon} \ln P[M(x) = y]\big\} \\ \mathbb{M}_2^\varepsilon = \big\{&M : X \to Y \mid \exists v(x, y) \text{ s.t.} \\ &\Delta v \leqslant 1, P[M(x) = y] \sim \mathrm{e}^{\varepsilon v(x,y)}\big\} \end{aligned}$$

Observe the following facts:

1. $\mathbb{M}_1^\varepsilon = \{M : X \to Y \mid M \text{ is } \varepsilon\text{-DP}\}$;
2. $\mathbb{M}_1^\varepsilon \subsetneqq \mathbb{M}_2^\varepsilon \subsetneqq \mathbb{M}_1^{2\varepsilon}$.

Common folklore that exponential mechanisms are "universal" in ε-DP refers to the first fact. However, our paper focuses on the class $\mathbb{M}_2^\varepsilon$, not only because the composition in $\mathbb{M}_1^\varepsilon$ is well-understood (Kairouz et al., 2017; Murtagh and Vadhan, 2016; Dong et al., 2019), but also because quality scores exist *a priori* in practice. On the other hand, it is not obvious that the function $\frac{1}{\varepsilon} \ln P[M(x) = y]$ is meaningful in any sense.

If we were to use $\mathbb{M}_1^\varepsilon$ as a proxy, then the second fact suggest that we can get a 2-approximation of the truely optimal result. However, a factor of 2 can be of vital importance in practice as mentioned in Section 1.

Having justified the importance of the class $\mathbb{M}_2^\varepsilon$, we note that this class is quite challenging to study. The reason is best illustrated by the following example.

Example 1. Consider two quality scores $u(x, y)$ and $u'(x, y) = u(x, y) + f(x)$ where $f : X \to \mathbb{R}$ is an arbitrary function. The two quality scores lead to the same mechanism since

$$\frac{e^{\varepsilon u(x,y)}}{\sum_y e^{\varepsilon u(x,y)}} = \frac{e^{\varepsilon u'(x,y)}}{\sum_y e^{\varepsilon u'(x,y)}}.$$

However, it is very common that $\Delta u \neq \Delta u'$. For example let $X = Y = \{0, 1\}$ and $u(x, y) = x + y, f(x) = 10x$ and hence $u'(x, y) = 11x + y$. Clearly $\Delta u = 1$ and $\Delta u' = 11$.

This example shows that even if a mechanism has a very sensitive quality score u', it may still belong to $\mathbb{M}_2^\varepsilon$, because a less sensitive quality score u may lead to the same mechanism.

In this respect, we propose to use the following more refined quantity to replace sensitivity.

Definition 2.3. *Given a quality score $u : X \times Y \to \mathbb{R}$, its range $\widetilde{\Delta}u$ is defined as*

$$\widetilde{\Delta}u = \sup_{x \sim x'} \big\{ \max_y \{u(x', y) - u(x, y)\} - \min_y \{u(x', y) - u(x, y)\} \big\}.$$

Let's first examine the above example. Since the only neighbors are $x = 0$ and $x' = 1$, we have $\widetilde{\Delta}u = \max_y\{1 + y\} - \min_y\{1 + y\} = 1$ and $\widetilde{\Delta}u' = \max_y\{11 + y\} - \min_y\{11 + y\} = 1$. This is true in general. We state this fact and some more in the following proposition.

Proposition 1. *The range $\widetilde{\Delta}$ has the following properties*

(a) $\widetilde{\Delta}u = \widetilde{\Delta}u'$ *when* $u'(x, y) = u(x, y) + f(x)$;

(b) $\widetilde{\Delta}u \leqslant 2 \cdot \Delta u$;

The exponential mechanism with quality score u and parameter ε, as in Definition 2.2, has the following privacy property:

Proposition 2. *$M_{u,\varepsilon}$ is $\widetilde{\Delta}u \cdot \varepsilon$-DP.*

Theorem 1 is a corollary of this proposition and property (b). This shows that the range $\widetilde{\Delta}$ can lead to more refined privacy analysis than the previously accepted notion of sensitivity Δ. Indeed, for counting queries[1] we have $\widetilde{\Delta}u = \Delta u$. For this important class, Proposition 2 manages to improve by an (important!) factor of 2 on Theorem 1 with the help of the newly introduced notion.

Back to the class of exponential mechanisms that we are interested in, the new definition using range is

$$\widetilde{\mathbb{M}}_1^\varepsilon = \big\{M : X \to Y \mid \widetilde{\Delta}u \leqslant 1 \text{ where } u(x, y) = \tfrac{1}{\varepsilon} \ln P[M(x) = y]\big\}$$

$$\widetilde{\mathbb{M}}_2^\varepsilon = \big\{M : X \to Y \mid \exists u(x, y) \text{ s.t. } \widetilde{\Delta}u \leqslant 1, P[M(x) = y] \sim \mathrm{e}^{\varepsilon u(x,y)}\big\}$$

Now we can claim another property, continuing Proposition 2.

(c) $\widetilde{\mathbb{M}}_1^\varepsilon = \widetilde{\mathbb{M}}_2^\varepsilon$.

Because of the focal position of this class in the rest of the paper, it is worth of a definition.

Definition 2.4. *A mechanism $M : X \to Y$ is called ε-bounded-range (BR) if the log likelihood function $u(x, y) = \ln P[M(x) = y]$ has range at most ε.*

This is exactly the notion introduced in (Durfee and Rogers, 2019). It's not hard to see the following formulations are equivalent:

[1]The function u is a counting query if $u(x, y)$ is the number of people in dataset x that has property y.

Proposition 3. *The followings are equivalent:*

(1) *M is ε-BR;*

(2) *$M \in \widetilde{\mathbb{M}}_1^\varepsilon = \widetilde{\mathbb{M}}_2^\varepsilon$;*

(3) *For each pair of neighboring datasets x, x' there exists some $t \in [0, \varepsilon]$ such that for any outcome $y \in Y$ we have*

$$t - \varepsilon \leqslant \ln\left(\frac{\Pr[M(x) = y]}{\Pr[M(x') = y]}\right) \leqslant t.$$

Moreover, the equivalence between (1) and (2) identifies the class of ε-BR mechanisms and $\widetilde{\mathbb{M}}_2^\varepsilon$. As a consequence, the following terms will be used interchangably in the rest of the paper:

> "ε-BR mechanisms"
> =
> "exponential mechanisms whose quality score has range at most 1 and has parameter ε"

2.3. Composition of BR Mechanisms

Now that we have identified the class of exponenetial mechanisms we are interested in as the class of ε-BR mechanisms, the primary focus of the paper becomes adaptive and non-adaptive compositions of BR mechanisms. More precisely, a mechanism $M : X \to Y_1 \times \cdots \times Y_k$ is a k-fold non-adaptive composition of ε-BR mechanisms if there are ε-BR mechanisms $M_i : X \to Y_i, i = 1, 2 \ldots, k$ such that $M(x) = \big(M_1(x), \ldots, M_k(x)\big)$. Similarly, M is a k-fold adaptive composition of ε-BR mechanisms if there are mechanisms $M_i : X \times Y_1 \times \cdots \times Y_{i-1} \to Y_i$ such that M is the composition of $M_1, \ldots, M_k$ as defined in Section 2.1, and $M_i(\cdot, y_1, \ldots, y_{i-1}) : X \to Y_i$ is ε-BR for each $y_1 \in Y_1, \ldots, y_{i-1} \in Y_{i-1}$. Again, the only difference is whether latter mechanisms are allowed to see former outputs.

On the practice side, as briefly discussed in the introduction, we are mainly interested in answering queries like "What are the most popular articles in the last 30 days among data scientists working in Bay area?" privately. For the next question, she may consider changing the time window. This is a typical case where the composition is adaptive.

All proofs of the statements in this section are quite straightforward, so they are relegated to the appendix.

3. Main Results

We set up a few notations before we state the results.

The following two formula are useful through out the rest of the paper. For $\varepsilon \geqslant 0$ and $t \in [0, \varepsilon]$, let

$$p_{t,\varepsilon} = \frac{\mathrm{e}^{-t} - \mathrm{e}^{-\varepsilon}}{1 - \mathrm{e}^{-\varepsilon}}, \quad q_{t,\varepsilon} = \mathrm{e}^{t} p_{t,\varepsilon} = \frac{1 - \mathrm{e}^{t-\varepsilon}}{1 - \mathrm{e}^{-\varepsilon}} \tag{1}$$

The ε subscript is usually dropped when the meaning is clear from the context. It might be helpful to notice that when t increases from 0 to ε, both p_t and q_t decrease from 1 to 0.

We follow the convention of (Murtagh and Vadhan, 2016) and use ε to denote the privacy parameters of a single mechanism and $(\varepsilon_g, \delta_g)$ for the global privacy parameters incurred by the composition. Although typical use case fixes $\delta_g \in [0, 1]$ and computes the dependence $\varepsilon_g(\delta_g)$, it is often easier ((Kairouz et al., 2017; Murtagh and Vadhan, 2016; Balle and Wang, 2018), see also Lemma 4.2) to write δ_g as an explicit formula of ε_g. Because of the obvious monotone dependence of δ_g on ε_g, a binary search can invert the function conveniently.

Now we are ready for the formal statements.

For non-adaptive composition, we have

Theorem 2. *If M is a k-fold non-adaptive composition of ε-BR mechanisms, then it is $\big(\varepsilon_g, \delta_k^{\mathrm{NA}}(\varepsilon_g)\big)$-DP where*

$$\delta_k^{\mathrm{NA}}(\varepsilon_g) = \max_{0 \leqslant \ell \leqslant k} \sum_{i=0}^{k} \binom{k}{i} p_{t_\ell^*}^{k-i} (1 - p_{t_\ell^*})^i (\mathrm{e}^{kt_\ell^* - i\varepsilon} - \mathrm{e}^{\varepsilon_g})_+,$$

where $(a)_+$ is defined as $\max\{a, 0\}$ and $t_\ell^* = \frac{\varepsilon_g + (\ell+1)\varepsilon}{k+1}$ where if $t_\ell^* \notin [0, \varepsilon]$, then we round it to the closest point in $[0, \varepsilon]$.

Privacy parameter $\delta_k^{\mathrm{NA}}(\varepsilon_g)$ cannot be improved. Furthermore, it can be evaluated in $O(k^2)$ time.

We remark that the seemingly complicated expression comes from a significant simplification to an even more colossal optimization problem, which enjoys all possible challenging features: high-dimensional, non-convex/concave, non-smooth. We manage to show the symmetry of its maximizer and simplify it to the shape above.

In the adaptive case, we have a recursive formula.

Theorem 3. *If M is a k-fold adaptive composition of ε-BR mechanisms, then it is $\big(\varepsilon_g, \delta_k^{\mathrm{A}}(\varepsilon_g)\big)$-DP where $\{\delta_j^{\mathrm{A}}\}_{0 \leqslant j \leqslant k}$ are recursively defined as*

$$\delta_0^{\mathrm{A}}(\varepsilon_g) = \max\{1 - \mathrm{e}^{\varepsilon_g}, 0\}$$
$$\delta_{j+1}^{\mathrm{A}}(\varepsilon_g) = \max_{t \in [0,\varepsilon]} q_t \delta_j^{\mathrm{A}}(\varepsilon_g - t) + (1 - q_t)\delta_j^{\mathrm{A}}(\varepsilon_g - t + \varepsilon).$$

Privacy parameter $\delta_k^{\mathrm{A}}(\varepsilon_g)$ is not improvable.

This recursive formula is not provably efficient, because the value of the $j+1$-th function at ε_g not only depends on values of previous functions at ε_g, but also at other locations. However, the following is a heuristic approach: we first discretize the domain of t and ε_g, and restrict δ_j^{A} to the grid of ε_g to get an array of function values. Now we use the recursion to compute the next array using the previous array. Running time of this heuristic algorithm is at most kN^2 where N is the grid size.

In the appendix we show that $\delta_k^{\mathrm{NA}}(\varepsilon_g)$ and $\delta_k^{\mathrm{A}}(\varepsilon_g)$ are not equal for a wide range of ε_g. Together with the optimality of δ_k^{NA} and δ_k^{A}, it implies the following explicit distinction between adaptive composition and non-adaptive composition of BR mechanisms.

Theorem 4. *For each $k \geqslant 4$ and $\varepsilon_g \in [0, (k-3)\varepsilon]$, let $\delta = \delta_k^{\mathrm{NA}}(\varepsilon_g)$. Then*

1. *Every k-fold non-adaptive composition of ε-BR mechanisms is (ε_g, δ)-DP;*
2. *There exists a k-fold adaptive composition of ε-BR mechanisms that is not (ε_g, δ)-DP.*

For the heterogenous case where BR parameters $\varepsilon_1, \ldots, \varepsilon_k$ are allowed to be different, see the appendix for results. Unfortunately in this case there are reasons to believe that optimal formula are #P-hard to evaluate, similar to the hardness result in (Murtagh and Vadhan, 2016). Instead, we provide an efficiently computable bound on the optimal privacy parameter by exploiting moment generating function of the privacy loss.

3.1. Efficiently Computable Bounds for Adaptive Jeterogenous Composition

Both results presented in the following can be considered as improvements of the previous result in (Durfee and Rogers, 2019). To understand the source of these improvements, it is helpful to recall their technique: they followed a similar approach to (Dwork et al., 2010), applying both an Azuma-Hoeffding bound (on the variance) and a KL divergence bound (on the bias) to achieve a reasonably simple upper bound on the optimal composition. However, this work only considered using the BR property to improve the bound from Azuma-Hoeffding and did not consider improving the KL divergence bound. Using the reduction to be introduced in the next section, the supremum of the KL divergence can be computed exactly. This yields the following result:

Proposition 4. *If M_i is ε_i-BR for $i = 1, 2, \ldots, k$, then their adaptive composition is $\big(\varepsilon_g(\delta_g), \delta_g\big)$-DP with*

$$\varepsilon_g(\delta) = \min\left\{ \sum_{i=1}^{k} \varepsilon_i, \sum_{i=1}^{k} \Big(\frac{\varepsilon_i}{1 - \mathrm{e}^{-\varepsilon_i}} - 1 - \ln\Big(\frac{\varepsilon_i}{1 - \mathrm{e}^{-\varepsilon_i}}\Big)\Big) + \sqrt{\frac{1}{2} \sum_{i=1}^{k} \varepsilon_i^2 \ln(1/\delta)} \right\}.$$

This improvement can be substantial in some settings (See Figure 1), but we will further improve this bound. For that purpose, we backtrack one more step and use the same techniques from the proof of Azuma-Hoeffding but apply our more exact characterization.

Theorem 5. *If M_i is ε_i-BR for $i = 1, 2, \ldots, k$, then their adaptive composition is $(\varepsilon_g, \delta^{\text{MGF}}(\varepsilon_g))$-DP for any $\varepsilon_g \geq 0$ with*

$$\delta^{\text{MGF}}(\varepsilon_g) = \inf_{\lambda > 0} \ \mathrm{e}^{-\lambda\varepsilon_g + \sum_i h(\lambda;\varepsilon_i)}.$$

where $h(\lambda;\varepsilon) := \sup_{t\in[0,\varepsilon]} \lambda(\varepsilon - t) + \ln\left(1 + p_{\varepsilon,t}(\mathrm{e}^{-\lambda\varepsilon} - 1)\right)$ with $p_{\varepsilon,t} = \frac{\mathrm{e}^{-t} - \mathrm{e}^{-\varepsilon}}{1 - \mathrm{e}^{-\varepsilon}}$ as in (1).

We present plots of our results in Figure 1 for the homogeneous case, plotting ε_g as a function of k. As stated earlier, we label "ε-DP OptComp" as the optimal composition bound for DP mechanisms from (Murtagh and Vadhan, 2016), "DR19" as the composition bound for ε-BR mechanisms from (Durfee and Rogers, 2019), and "BR OptComp" as the composition bound in Theorem 2, which only applies in the non-adaptive setting. Furthermore, we label "OptKL" as the bound from Proposition 4 and "MGF" as the bound in Theorem 5. To compare our bounds with simply using the optimal DP composition bound with a half the actual privacy parameter, we also plot the DP optimal composition bound with $\varepsilon/2$ with label "$\varepsilon/2$-DP OptComp". This last curve highlights the fact that ε-BR is almost the same as $\varepsilon/2$-DP when applying composition.

4. Overview of Techniques

4.1. Reduction to Generalized Randomized Response

Essentially all of the results are obtained by first identifying the "worst-case" mechanism for the class of BR mechanisms and then reason about them. Similar to (Kairouz et al., 2017; Murtagh and Vadhan, 2016), "worst-case" means that any BR mechanism can be simulated through post-processing of this worst-case mechanism. For the class of ε-DP mechanisms, the worst-case mechanism was shown to be randomized response (Kairouz et al., 2017; Murtagh and Vadhan, 2016). Fixing the neighboring datasets x and x', randomized response returns two Bernoulli distributions with parameters $\frac{1}{1+\mathrm{e}^{\varepsilon}}$ and $\frac{\mathrm{e}^{\varepsilon}}{1+\mathrm{e}^{\varepsilon}}$ respectively. The "generalized randomized response", when applied to neighboring datasets, returns two Bernoulli distributions with parameters p_t and q_t as defined in (1). In fact, the two parameters are specifically defined for this purpose. More precisely,

Lemma 4.1. *If a mechanism $M : \mathcal{X} \to \mathcal{Y}$ is ε-BR and $x, x' \in \mathcal{X}$ are neighboring datasets, there exists some $t = t(M, x^0, x^1) \in [0, \varepsilon]$ and a randomized function $\mathrm{Proc} : \{0,1\} \to \mathcal{Y}$, such that the following distributional equalities hold:*

$$\mathrm{Proc}\big(\mathrm{Bern}(p_t)\big) = M(x), \qquad \mathrm{Proc}\big(\mathrm{Bern}(q_t)\big) = M(x').$$

This closely mirrors the reduction of ε-DP mechanisms reduce to randomized response result. To take a closer look and understand the expressions of p_t and q_t, recall that a mechanism M is ε-DP if and only if for any y, the log likelihood ratio (or equivalently, privacy loss) lies in a closed interval

$$\ln\left(\frac{\Pr[M(x) = y]}{\Pr[M(x') = y]}\right) \in [-\varepsilon, \varepsilon].$$

Randomized response, namely $M(x) = \mathrm{Bern}(\frac{1}{1+\mathrm{e}^{\varepsilon}})$ and $M(x) = \mathrm{Bern}(\frac{\mathrm{e}^{\varepsilon}}{1+\mathrm{e}^{\varepsilon}})$, satisfies a more restrained version

$$\ln\left(\frac{\Pr[M(x) = y]}{\Pr[M(x') = y]}\right) \in \{-\varepsilon, \varepsilon\}.$$

In parallel, we know from Corollary 3 that a mechanism $M : \mathcal{X} \to \mathcal{Y}$ is ε-BR, if and only if there is $t = t(x, x') \in [0, \varepsilon]$ such that for any y,

$$\ln\left(\frac{\Pr[M(x) = y]}{\Pr[M(x') = y]}\right) \in [t - \varepsilon, t].$$

It is straightforward to verify that the log likelihood ratio of $\mathrm{Bern}(p_t)$ and $\mathrm{Bern}(q_t)$ only takes values $t - \varepsilon$ and t. In particular, this allows a hypothesis testing interpretation of ε-BR and the powerful Blackwell's theorem[2] (Blackwell, 1950) kicks in. A similar argument as in (Kairouz et al., 2017; Murtagh and Vadhan, 2016) yields Lemma 4.1.

With this reduction, it then follows that composition of BR mechanisms (adaptive or non-adaptive) can be reduced to simply considering composition of this worst-case mechanism, allowing for explicit description of the optimal composition.

While the result explained so far is largely unsurprising, we point out that the additional parameter t yields significant difficulty. Roughly speaking, since t is only an intermediate object that cannot appear in the theorems, we need to find the worst collection of $\{t_i : i = 1, 2, \ldots, k\}$ where k is the number of components in the composition. It is this optimization step that makes the problem challenging. In addition, interactivity allows the adversary to "choose" the next t based on the previous results he has seen. This additional power is absent in all previous works, which explains why this setting admits the first separation result on adaptive and non-adaptive composition.

[2] It is possible to avoid this heavy machinery and only utilize elementary results in hypothesis testing. See (Dong et al., 2019).

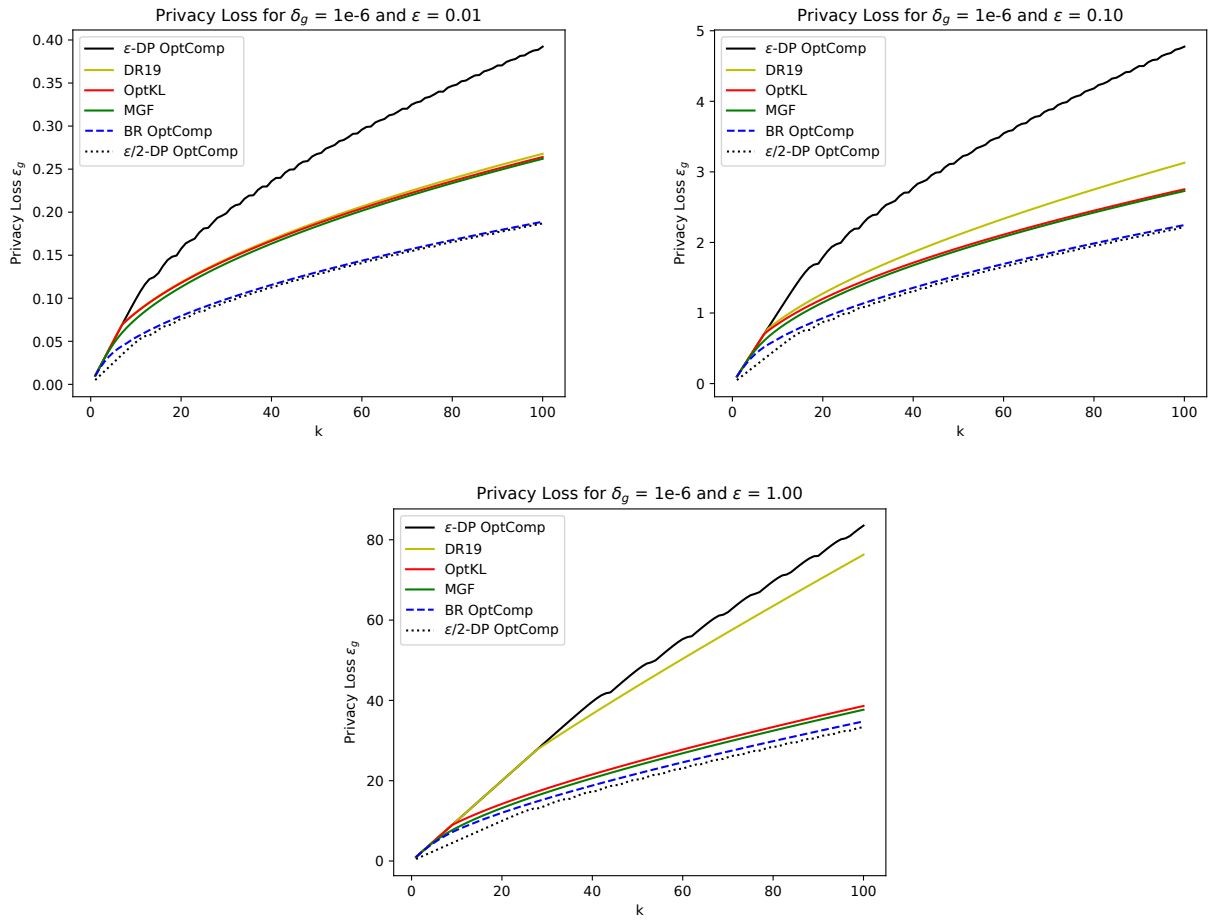

Figure 1. Comparison of optimal DP composition with the BR composition bounds in this work and in Durfee and Rogers (2019). BR OptComp (dashed blue) corresponds to Theorem 2, one of our main theorem. It is almost as private as optimal DP composition (dotted black) with parameter cut in half. We present results for $\delta_g = 10^{-6}$ and $\varepsilon \in \{0.01, 0.1, 1\}$.

To finish on reduction, we note that the BR property is not preserved under randomization. On the other hand, randomization is common in private machine learning. Consequently, the right object to study is composition in the class of randomized BR mechanisms, instead of composition of pure BR mechanisms. However, it turns out there's no difference in terms of their optimal (ε, δ) parameters. Interested readers can read about the simple argument in the appendix.

4.2. Non-adaptive Composition

From Section 2.1 we see that non-adaptive composition simply yields product distributions. Using the reduction introduced above, the "worst-case" of each ε-BR component is some generalized randomized response. To obtain the optimal composition guarantee, it suffices to assume the composition M operate on neighboring datasets x, x' as follows:

$$M(x) = \mathrm{Bern}(p_{t_1}) \times \cdots \times \mathrm{Bern}(p_{t_k})$$
$$M(x') = \mathrm{Bern}(q_{t_1}) \times \cdots \times \mathrm{Bern}(q_{t_k})$$

We first recall a useful tool that computes for a given mechanism M, the optimal δ such that M is (ε, δ)-DP. Formally, given a mechanism $M : \mathcal{X} \to \mathcal{Y}$ and any $\varepsilon \in \mathbb{R}$, let

$$\delta_{\text{opt}}(M, \varepsilon) := \inf \left\{ \delta : M \text{ is } (\varepsilon, \delta)\text{-DP} \right\}$$

Lemma 4.2.

$$\delta_{\text{opt}}(M, \varepsilon) = \sup_{x \sim x'} \mathbb{E}_{y \sim M(x')} \left[(1 - \mathrm{e}^{\varepsilon - L(y; x, x')})_+ \right]$$

where $L(y; x, x') = \ln \frac{P[M(x')=y]}{P[M(x)=y]}$ *is the log likelihood ratio function and* a_+ *denotes* $\max\{a, 0\}$.

Plugging in Bernoulli products into Lemma 4.2 and optimizing over $\mathbf{t} = (t_1, \ldots, t_k)$ yields the following preliminary version of Theorem 2.

Lemma 4.3. *If M is a k-fold non-adaptive composition of ε-BR mechanisms, then it is $\left(\varepsilon_g, \delta_k^{\mathrm{NA}}(\varepsilon_g)\right)$-DP where*

$$\delta_k^{\mathrm{NA}}(\varepsilon_g) = \sup_{\mathbf{t}\in[0,\varepsilon]^k} \sum_{S\subseteq\{1,\dots,k\}} \max\Big\{0, \prod_{i\notin S} q_{t_i} \prod_{i\in S}(1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i\notin S} p_{t_i} \prod_{i\in S}(1-p_{t_i})\Big\}.$$

Note this formulation closely mirrors the following result from Murtagh and Vadhan (2016).

Theorem 6 (Theorem 1.5 from Murtagh and Vadhan (2016))**.** *$M_1, \dots, M_k$ be mechanisms such that M_i is ε_i-DP, $i = 1, 2, \dots, k$. Their adaptive composition is $\left(\varepsilon_g, \delta^{\mathrm{DP}}(\varepsilon_g)\right)$-DP with*

$$\delta^{\mathrm{DP}}(\varepsilon_g) = \sum_{S\subseteq\{1,\dots,k\}} \max\Big\{0, \prod_{i\notin S} \tilde{q}_i \prod_{i\in S}(1-\tilde{q}_i) - \mathrm{e}^{\varepsilon_g} \prod_{i\notin S} \tilde{p}_i \prod_{i\in S}(1-\tilde{p}_i)\Big\}$$

where $\tilde{p}_i = \frac{1}{1+\mathrm{e}^{\varepsilon_i}}, \tilde{q}_i = \frac{\mathrm{e}^{\varepsilon_i}}{1+\mathrm{e}^{\varepsilon_i}}$. In addition, evaluating $\delta^{\mathrm{DP}}(\varepsilon_g)$ is #P-complete.

The major difference between these two results is that Lemma 4.3 involves an additional high dimensional optimization, with highly non-convex and non-smooth objective. This seems to suggest that δ_k^{NA} is even harder to evaluate than δ^{DP}. With the strong hardness conclusion of Theorem 6, it seems quite challenging, if not hopeless, to tackle δ_k^{NA}.

It turns out that we can overcome the difficulty, as stated in Theorem 2. Our simplification will require significant technical work that will ultimately be done in two key steps: 1) we show that the supremum is achieved when all $t_i = t_j$ for $i \neq j$, and 2) we show that the supremum is achieved by a certain value $t_i = t^* \in [0, \varepsilon]$ contained in a set of at most k possible values. This will then yield an explicit and efficiently computable formulation of the optimal non-adaptive composition of range-bounded mechanisms.

We remind the readers that this result focuses on the homogeneous setting where $\varepsilon_1, \cdots, \varepsilon_k = \varepsilon$, while Theorem 6 is about the general heterogenous case. Unfortunately, our technique does not extend to either the heterogenous setting, for which we conjecture the same #P-hardness, or the adaptive setting, pointing to the natural question of whether there is in fact further privacy loss when the adversary is given power to choose the mechanism based upon previous responses.

4.3. Additional Power of Interactivity

Rigorous justification of the gap relies heavily upon having obtained optimality results for adaptive and non-adaptive compositions, plus intensive calculation. Even an intuitive explanation is heavy in terms of notation, so we relegate everything to the appendix. Here we make some comments on why this can be particularly interesting to the DP community.

For the existing DP composition theorems, adaptivity in the choice of DP algorithm did not affect the overall privacy parameters. Rogers et al. (2016) show that there is an asymptotic gap in the privacy loss bound when the privacy parameters $\{\varepsilon_i\}_{i=1}^k$ are fixed in advance versus when an analyst can adaptively select the privacy parameters ε_i at each round i based on previous outcomes before i. However, we focus on the traditional view of DP that fixes all the privacy parameters up front. Role of interactivity and adaptivity in learning algorithms and estimation tasks have been studied in (Kasiviswanathan et al., 2011; Smith et al., 2017; Joseph et al., 2019; Duchi and Rogers, 2019) in the model of local DP.

5. Conclusion and Future Directions

In this work, we studied the privacy parameters when composing multiple exponential mechanisms, which is fundamental in private machine learning and private data analysis. We considered both cases when the exponential mechanisms can be adaptively selected at each round and when they are all selected in advance, and provide optimal results in both cases. Based on these results, we showed a separation of privacy parameters between adaptive composition and non-adaptive composition, which to our knowledge is a first of its kind result.

We then provided improved and computationally efficient composition bounds for the adaptive and inhomogeneous case by tailoring concentration bounds for our particular setting. In order to better understand the adaptive composition bound, one potential direction for future work is to understand the asymptotics of the privacy loss bound, as $k \to \infty$. We conjecture that the asymptotic gap collapses between the optimal composition bound for the adaptive and nonadaptive cases, and leave that as future work to study. Furthermore, in the non-asymptotic setting we believe that the gap between adaptive and non-adaptive is quite small, and also leave proving a strong upper bound on this gap to future work.

We demonstrate by Figure 1 that our improved analysis let the top-k algorithm in (Durfee and Rogers, 2019) run twice as many rounds as in the original analysis. An exciting direction is to apply the analysis to other machine learning algorithms such as private PCA (Chaudhuri et al., 2012) to see how much privacy budget we can save, or equivalently, how much accuracy we can buy while retaining the same privacy level.

References

M. Abadi, A. Chu, I. Goodfellow, B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *23rd ACM Conference on Computer and Communications Security (ACM CCS)*, pages 308–318, 2016. URL https://arxiv.org/abs/1607.00133.

B. Balle and Y.-X. Wang. Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 394–403, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018. PMLR. URL http://proceedings.mlr.press/v80/balle18a.html.

D. Blackwell. Comparison of experiments. Technical report, HOWARD UNIVERSITY Washington United States, 1950.

Z. Bu, J. Dong, Q. Long, and W. J. Su. Deep learning with gaussian differential privacy. *arXiv preprint arXiv:1911.11607*, 2019.

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

K. Chaudhuri, A. Sarwate, and K. Sinha. Near-optimal algorithms for differentially-private principal components. In *Advances in Neural Information Processing Systems 25*, 2012. URL http://arxiv.org/abs/1207.2812.

K. Chaudhuri, A. D. Sarwate, and K. Sinha. A near-optimal algorithm for differentially-private principal components. *The Journal of Machine Learning Research*, 14 (1):2905–2943, 2013.

J. Dong, A. Roth, and W. J. Su. Gaussian differential privacy. *CoRR*, abs/1905.02383, 2019. URL http://arxiv.org/abs/1905.02383.

J. Duchi and R. Rogers. Lower bounds for locally private estimation via communication complexity. In A. Beygelzimer and D. Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1161–1191, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/duchi19a.html.

D. Durfee and R. Rogers. Practical differentially private top-k selection with pay-what-you-get composition. *CoRR*, abs/1905.04273, 2019. URL http://arxiv.org/abs/1905.04273.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology (EUROCRYPT 2006)*, 2006a.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Theory of Cryptography Conference*, pages 265–284, 2006b.

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

M. Hardt and G. N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *51st Annual Symposium on Foundations of Computer Science*, pages 61–70, 2010.

M. Joseph, J. Mao, S. Neel, and A. Roth. The role of interactivity in local differential privacy. *CoRR*, abs/1904.03564, 2019. URL http://arxiv.org/abs/1904.03564.

P. Kairouz, S. Oh, and P. Viswanath. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6):4037–4049, June 2017. ISSN 0018-9448. doi: 10.1109/TIT.2017.2685505.

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

J. Liu and K. Talwar. Private selection from private candidates. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 298–309, 2019.

F. McSherry and K. Talwar. Mechanism design via differential privacy. In *48th Annual Symposium on Foundations of Computer Science*, 2007.

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

J. Murtagh and S. Vadhan. The complexity of computing the optimal composition of differential privacy. In *Proceedings, Part I, of the 13th International Conference on Theory of Cryptography - Volume 9562*, TCC 2016-A, pages 157–175, Berlin, Heidelberg, 2016. Springer-Verlag. ISBN 978-3-662-49095-2. doi: 10.1007/978-3-662-49096-9_7. URL https://doi.org/10.1007/978-3-662-49096-9_7.

R. M. Rogers, A. Roth, J. Ullman, and S. P. Vadhan. Privacy odometers and filters: Pay-as-you-go composition. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 1921–1929, 2016.

A. Smith, A. Thakurta, and J. Upadhyay. Is interaction necessary for distributed private learning? In *IEEE Symposium on Security and Privacy*, 2017.