Bounding, Concentrating, and Truncating: Unifying Privacy Loss Composition for Data Analytics

Mark Cesar MCESAR@LINKEDIN.COM
LinkedIn Corporation

Ryan Rogers RROGERS386@GMAIL.COM
LinkedIn Corporation

Editors: Vitaly Feldman, Katrina Ligett and Sivan Sabato

Abstract

We unify existing privacy loss composition bounds for special classes of differentially private (DP) algorithms along with general DP composition bounds. In particular, we provide strong privacy loss bounds when an analyst may select pure DP, bounded range (e.g. exponential mechanisms), or concentrated DP mechanisms in any adaptively selected order. We also provide optimal privacy loss bounds that apply when an analyst can select pure DP and bounded range mechanisms in a batch, i.e. non-adaptively. Further, when an analyst selects mechanisms within each class adaptively, we show a difference in privacy loss between different, predetermined orderings of pure DP and bounded range mechanisms. Lastly, we compare the composition bounds of Laplace and Gaussian mechanisms and provide new private mechanisms for top-k using truncated Gaussian noise.

Keywords: Differential privacy, privacy preserving data analysis, adaptive composition

1. Introduction

Differential privacy (DP) provides a mathematical formalism to an intuitive notion of what it means for a computation to be private — the computation should produce similar results with or without any one individual's data. With this formalism, we can quantify the privacy loss of a computation that injects noise when evaluated on a sensitive dataset. This allows us to determine which DP algorithms are more private than others, i.e. which has smaller privacy loss. Furthermore, if multiple computations are done on the same dataset we can still quantify the privacy loss over the entire interaction with the dataset, i.e. DP composes. As opposed to measuring utility empirically, e.g. prediction accuracy of a classification task, privacy loss in DP requires analytical bounds over worst case datasets and outcomes. Improvements to the privacy loss bounds show that a given algorithm might actually be more private than originally proposed, with no changes to the algorithm itself.

Hence, there have been many works in precisely bounding the overall privacy loss. There are multiple composition bounds to use, including bounds that hold for any DP algorithms (Dwork et al., 2006b, 2010; Kairouz et al., 2017; Murtagh and Vadhan, 2016), as well as improved composition bounds that only apply to specific types of DP algorithms (Abadi et al., 2016; Mironov, 2017; Bun and Steinke, 2016; Dong et al., 2019; Durfee and Rogers, 2019; Dong et al., 2020). In the design of a DP system, we would like to provide the best possible bounds on the privacy loss that apply for combinations of general DP algorithms as well as specific types of DP algorithms that enjoy much better composition bounds. One can simply use the most general formulations to provide a loose

bound on the overall privacy loss, but this neglects the improvements that can be made, which allow for more queries or more accurate results.

Consider a privacy system for data analytics, supporting tasks such as counting queries and exploratory analysis. These general tasks typically use the Laplace mechanism (Dwork et al., 2006b) or Gaussian mechanism (Dwork et al., 2006a) to provide noisy counts, as well as exponential mechanisms (McSherry and Talwar, 2007), a general class of DP algorithms that have been shown to achieve much improved composition bounds. In particular, Dong et al. (2020) showed that one can query nearly four times more exponential mechanisms for the same overall privacy bound as what can be achieved with using the general, optimal DP composition bounds (Kairouz et al., 2017). This improvement is because Durfee and Rogers (2019) defined a stronger condition that exponential mechanisms satisfy, called *bounded range* (BR). In particular ε-BR implies ε-DP, whereas ε-DP implies 2ε-BR. Further, the Gaussian mechanism does not satisfy (pure) ε-DP, but rather a slight relaxation called *concentrated* DP (CDP) (Dwork and Rothblum, 2016) or zero-mean concentrated DP (zCDP) (Bun and Steinke, 2016).

To see that combining DP and BR mechanisms can arise naturally, consider a privacy system that allows for general top-k queries. One would typically use a two phase approach to ensure DP. The first phase would use a series of exponential mechanisms to *discover* the domain of elements in the top-k. Given the discovered set, the second phase goes back to the dataset to add noise to the true counts of the discovered elements via the Laplace mechanism and then release the noisy counts. In fact, most DP top-k algorithms use this two phase approach, see for example Bhaskar et al. (2010) and Durfee and Rogers (2019). One approach to bounding the privacy loss of such an interaction would be to simply use the general DP composition bounds, but this ignores the improved composition bounds that are possible via the BR analysis. Another approach would be to analyze the composition bounds via BR, but this results in doubling the privacy parameter for each Laplace mechanism, as was done in LinkedIn's privacy system that handles top-k queries (Rogers et al., 2020). We then follow a line of research proposed in Dong et al. (2020), studying the privacy loss bounds that combine both general DP bounds and improved BR bounds.

We make several contributions in this work, starting with providing a unified framework for analyzing the accumulated privacy loss of multiple variants of DP under the least stringent privacy definition we consider, zero-mean concentrated DP (zCDP) (Bun and Steinke, 2016). Although BR and (pure) DP can be parameterized as zCDP, it is important to point out that existing strong privacy loss bounds were not shown to apply if the class of mechanism (BR, DP, zCDP) were adaptively selected at each round, despite the parameters being fixed in advance. This is because the choice of zCDP parameters at each round is chosen adaptively from a predetermined set. Nevertheless, we show existing bounds still do apply even in this more general setting thus allowing for arbitrary orderings of mechanisms.

There are several approaches to reducing the overall privacy loss of a system: improving general bounds for existing private algorithms, designing private algorithms with improved utility, or a combination of both. For more specific scenarios, we seek to improve the privacy loss bounds from the fully generic bounds. Our next contribution is that we provide the optimal privacy loss bound when $k-m$ of the k ε-DP mechanisms are ε-BR and are non-adaptively selected in the homogenous privacy parameter setting, i.e. all privacy parameters are ε. With these bounds, we can interpolate between the two extremes of only composing ε-DP mechanisms (Kairouz et al., 2017; Murtagh and Vadhan, 2016) and composing only ε-BR mechanisms (Dong et al., 2020). In the adaptive setting, we show that the ordering between ε-BR and ε-DP mechanisms matters in certain settings.

Hence the privacy loss can differ between an analyst adaptively selecting exponential mechanisms after using Laplace mechanisms and an analyst that alternates between exponential mechanisms and Laplace mechanisms, as one would do with private top-k.

We then seek to improve existing private algorithms to reduce the amount of added noise for the same level of privacy. Rather than using Laplace noise mechanisms for releasing histograms privately, we present Gaussian based variants of mechanisms presented in LinkedIn's recently deployed privacy system (Rogers et al., 2020). When using Gaussian noise, we can solve a constrained least squares problem to provide privatized counts in a top-k query subject to the ranked ordering provided by a series of exponential mechanisms. We then present a zCDP mechanism using truncated Gaussian noise for releasing top-k results that enjoy smaller thresholds when discovering elements and releasing their counts from an unknown domain than Laplace noise mechanisms (Durfee and Rogers, 2019).

2. Preliminaries

We begin by defining differential privacy, which considers two *neighboring* datasets x, x' from some data universe $\mathcal{X}$, i.e. x is the same as x' except one user's data has been removed or added, sometimes denoted as $x \sim x'$. Note that DP is parameterized by the *privacy loss* parameter $\varepsilon > 0$ and a small probability of privacy failure $\delta \in [0, 1)$.

Definition 1 (Dwork et al. (2006b,a)) *A randomized algorithm $M : \mathcal{X} \to \mathcal{Y}$ that maps input set $\mathcal{X}$ to some arbitrary outcome set $\mathcal{Y}$ is (ε, δ)-differentially private (DP) if for any neighboring datasets x, x' and outcome set $S \subseteq \mathcal{Y}$, $\Pr[M(x) \in S] \leq e^{\varepsilon} \Pr[M(x') \in S] + \delta$. When $\delta = 0$, we typically say that M is ε-DP or* pure *DP.*

One of the most useful properties of DP is that it composes, meaning that if one were to repeatedly use different DP algorithms, which can be adaptively selected at each round, then the result will remain DP, although with a slightly worse privacy loss parameter. A class of mechanisms that enjoys improved composition bounds are bounded range (BR) mechanisms (Durfee and Rogers, 2019; Dong et al., 2020). Roughly, composing ε-BR mechanisms in a batch (i.e. non-adaptively) is nearly the same, in terms of the accumulated privacy loss, as composing $\varepsilon/2$-DP mechanisms.

Definition 2 (Durfee and Rogers (2019)) *A randomized algorithm $M : \mathcal{X} \to \mathcal{Y}$ that maps input set $\mathcal{X}$ to some arbitrary outcome set $\mathcal{Y}$ is ε-bounded range (BR) if for any neighbors x, x', there exists a $t := t(x, x', M) \in [0, \varepsilon]$ such that $t - \varepsilon \leq \ln\left(\frac{\Pr[M(x)=y]}{\Pr[M(x')=y]}\right) \leq t$.*

In order to prove the optimal composition bounds for DP or BR mechanisms, it is sufficient to consider a simpler mechanism, based on randomized response (Warner, 1965). This mechanism takes a single bit and returns a bit, subject to DP or BR. We then define the generalized version of randomized response from Dong et al. (2020).

Definition 3 (Generalized Random Response) *For any $\varepsilon \geq 0$ and $t \in [0, \varepsilon]$, let $RR_{\varepsilon,t} : \{0, 1\} \to \{0, 1\}$ be a randomized mechanism in terms of probability $q_{\varepsilon,t}$ such that $RR_{\varepsilon,t}(0) = 0$ w.p. $\frac{1-e^{t-\varepsilon}}{1-e^{-\varepsilon}} =: q_{\varepsilon,t}$ and $RR_{\varepsilon,t}(1) = 0$ w.p. $\frac{e^{-t}-e^{-\varepsilon}}{1-e^{-\varepsilon}} = e^{-t} q_{\varepsilon,t} =: p_{\varepsilon,t}$.*

Note that $\texttt{RR}_{2\varepsilon,\varepsilon}(\cdot)$ is the standard randomized response mechanism from DP and is ε-DP. When considering the worst case optimal composition bounds of k adaptively or non-adaptively chosen ε-DP mechanisms, we need only consider the overall privacy loss of k repeated instances of $\texttt{RR}_{2\varepsilon,\varepsilon}(\cdot)$ (Kairouz et al., 2017). For obtaining the worst case optimal composition bound of ε-BR mechanisms, it gets more complicated. The main difficulty is in how each $t_i \in [0, \varepsilon]$ in $\texttt{RR}_{\varepsilon,t_i}(\cdot)$ is selected at each round $i \in [k]$. If all the t_i's are non-adaptively selected then Dong et al. (2020) provide an efficiently computable formula for the optimal composition bound. However, if each t_i can be selected as a function of previous outcomes, then they provide a recursive formula for computing the optimal bound on the privacy loss, which is conjectured to be computationally hard to evaluate, even in the homogenous parameter setting, i.e. $\varepsilon_i = \varepsilon$ for all i.

We will write $\mathcal{M}_{\texttt{BR}}(\varepsilon)$ to denote the class of all ε-BR mechanisms, and similarly $\mathcal{M}_{\texttt{DP}}(\varepsilon)$ will denote the class of all ε-DP mechanisms. As discussed earlier, we will need to differentiate our bounds when the mechanisms can be adaptively selected at each round or not. We then write $\mathcal{M}_1 \times \mathcal{M}_2 = \{(M_1(\cdot), M_2(\cdot)) : M_1 \in \mathcal{M}_1, M_2 \in \mathcal{M}_2\}$ as the non-adaptively selected class of mechanisms from $\mathcal{M}_1$ and $\mathcal{M}_2$. Alternatively, we will write $(\mathcal{M}_1, \mathcal{M}_2) = \{(M_1(\cdot), M_2(\cdot, M_1(\cdot))) : M_1 \in \mathcal{M}_1, M_2 \in \mathcal{M}_2\}$ to denote the class of mechanisms that can be adaptively selected at each round, based on previous outcomes.

Since DP has two privacy parameters, we will typically fix a global privacy loss parameter ε_g and write the best possible δ as a function of ε_g. Hence, we will use the following definition for the optimal privacy parameter as a function of ε_g.

Definition 4 (Optimal Privacy Parameters) *Given a mechanism $M : \mathcal{X} \to \mathcal{Y}$ and any $\varepsilon_g \in \mathbb{R}$, we define the optimal δ to be $\delta_{OPT}(M; \varepsilon_g) := \inf\{\delta \geq 0 : M \text{ is } (\varepsilon_g, \delta)\text{-DP}\}$. Further, if $\mathcal{M}$ is a class of mechanisms $M : \mathcal{X} \to \mathcal{Y}$, then for any $\varepsilon_g \in \mathbb{R}$, we define $\delta_{OPT}(\mathcal{M}; \varepsilon_g) := \sup_{M \in \mathcal{M}} \delta_{OPT}(M; \varepsilon_g)$.*

Kairouz et al. (2017) provides a formula for $\delta_{\texttt{OPT}}((\mathcal{M}_{\texttt{DP}}(\varepsilon_1), \cdots, \mathcal{M}_{\texttt{DP}}(\varepsilon_k)); \varepsilon_g)$, the optimal privacy parameter for a given $\varepsilon_g \geq 0$ for the homogeneous privacy parameter case. Further, Dong et al. (2020) presents a closed form formula for the optimal privacy parameters $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon_1) \times \cdots \times \mathcal{M}_{\texttt{BR}}(\varepsilon_k); \varepsilon_g)$, again in the homogeneous case but the choice of BR mechanisms at each round are non-adaptively selected.

We also provide the definition of zero-mean CDP (zCDP) from Bun and Steinke (2016). It is based on the Rényi divergence of order $\alpha > 1$ between two distributions P and Q over the same domain, denoted as $D_\alpha(P||Q)$ where $(\alpha - 1)D_\alpha(P||Q) := \ln \mathbb{E}_{z \sim P}\left[(P(z)/Q(z))^{\alpha-1}\right]$.

Definition 5 (Zero-mean Concentrated Differential Privacy) *A randomized algorithm $M : \mathcal{X} \to \mathcal{Y}$ is δ-approximately (ξ, ρ)-zCDP if for any neighbors $x, x' \in \mathcal{X}$, there exists events E and E', such that $\Pr[E], \Pr[E'] \geq 1 - \delta$ and for every $\alpha > 1$ we have the following bound in terms of the Rényi divergence $D_\alpha(\cdot||\cdot)$ of order α*

$$D_\alpha(M(x)|_E||M(x')|_{E'}) \leq \alpha\rho + \xi \qquad \text{and} \qquad D_\alpha(M(x')|_{E'}||M(x)|_E) \leq \alpha\rho + \xi.$$

where $M(x)|_E$ is the distribution of $M(x)$ conditioned on event E and similarly for $M(x')|_{E'}$. If $\delta = 0$, then we say M is (ξ, ρ)-zCDP and if $\xi = 0$ we write ρ-zCDP.

Note that ϵ-DP implies 2ϵ-BR and ϵ-BR implies ϵ-DP and we have the following connection between zCDP and DP from Bun and Steinke (2016).

Lemma 6 *If M is ρ-zCDP then M is also $(\rho + 2\sqrt{\rho \ln(1/\delta)}, \delta)$-DP for any $\delta > 0$.*

3. Set-wise Adaptive Composition

We now present a general way to interact with a privacy system that allows the analyst lots of freedom in selecting different private mechanisms. We allow the analyst to select zCDP parameters from a set $\mathcal{E} := \{(\delta_i, \xi_i, \rho_i) : i \in [k]\}$ at each round, but the order need not be predetermined. Allowing an analyst to adaptively select the order of parameters arises naturally in practice. Consider a setting where an analyst does heavy hitter selection (with BR mechanisms) and then decides whether she wants counts returned with either Laplace noise (DP) or with Gaussian noise (zCDP) to do some post processing (see later section that describes such a post processing mechanism), or she may decide to run a new heavy hitter query. This extra freedom to the analyst was not considered in previous composition bounds and stops short of the complete freedom of adaptive parameter selection in *privacy odometers* (Rogers et al., 2016), which does not require a preregistered set of privacy parameters but defines a different privacy guarantee than traditional DP.

We now detail the experiment between the analyst and the private mechanisms. The analyst selects zCDP parameters $(\delta_1, \xi_1, \rho_1)$ and then a δ_1-approximately (ξ_1, ρ_1)-zCDP mechanism with those parameters, while also updating $\mathcal{E} \leftarrow \mathcal{E} \setminus \{(\delta_1, \xi_1, \rho_1)\}$. Similar to other works on privacy loss composition, see Dwork et al. (2010), the analyst may then select neighboring datasets $x_1^{(0)}, x_1^{(1)}$. Once the analyst sees the outcome from the selected mechanism evaluated on the (unknown) data set $x_1^{(b)}$ where $b \in \{0, 1\}$, the analyst gets to adaptively select zCDP parameters from the remaining set $\mathcal{E}$, a corresponding mechanism, and neighboring datasets $x_2^{(0)}, x_2^{(1)}$. We then delete the selected privacy parameters from $\mathcal{E}$ and continue. The interaction proceeds until $\mathcal{E} = \emptyset$. The information that is hidden from the analyst is the bit $b \in \{0, 1\}$ that determines which of the neighboring datasets is used at each round.

We refer to this protocol between the analyst and the privacy system as the *$\mathcal{E}$-set-wise adaptive composition experiment b*. As in Dwork et al. (2010), we refer to $V^{(b)}$ as the *view* of the analyst in the $\mathcal{E}$-set-wise adaptive composition experiment b. Hence, we want to be able to show that the two experiments with b and $1 - b$ are similar.[1]

Definition 7 *We say that $\mathcal{E}$ is* (ε_g, δ)-differentially private under set-wise adaptive composition *if for any outcome set S for the $\mathcal{E}$-set-wise adaptive composition experiment b, we have* $\Pr[V^{(b)} \in S] \leq e^{\varepsilon_g} \Pr[V^{(1-b)} \in S] + \delta$.

We start with a general concentration bound result with full proof in Appendix A that will be used to bound the overall privacy loss in this experiment.

Lemma 8 *Let $(\Omega, \mathcal{F}, \Pr)$ be a probability triple where $\emptyset := \mathcal{F}_0 \subseteq \mathcal{F}_1, \subseteq \cdots \subseteq \mathcal{F}$ is an increasing sequence of σ-algebras. Let X_i be a real-valued $\mathcal{F}_i$-measurable random variabse and $X_0 = 0$. Assume that there exists B_i that are $\mathcal{F}_{i-1}$ measurable for $i \geq 1$ and $B_0 = 0$ such that $\sum_{i=0}^{k} B_i^2 \leq b^2$ a.s. for some constant b. If for all $\lambda > 0$, we have $\mathbb{E}[e^{\lambda X_i} \mid \mathcal{F}_{i-1}] \leq e^{\lambda^2 B_i^2/2}$ a.s. $\forall i$, then we have for all $\beta > 0$, $\Pr\left[\sum_{i=1}^{k} X_i \geq \beta\right] \leq \exp\left(\frac{-\beta^2}{2b^2}\right)$.*

1. As presented, we do not allow for the analyst to randomize over different choices of mechanisms at each round in the experiment. Hence, the analyst deterministically selects the next mechanism based on the previous outcomes. However, the differential privacy guarantees for the experiment would not change if we allow the analyst to randomize over choices of mechanisms at each round, as was shown in Lemma 4.3 in (Dong et al., 2020)

Many concentration bounds rely on a similar subgaussian bound. This particular result allows the analyst to adaptively select different classes of mechanisms at each round, where the subgaussian parameter changes based on the class that is selected, and the order of parameters need not be fixed in advance.

Note that the *advanced composition* bound from Dwork et al. (2010) uses Azuma's inequality for a concentration bound on the privacy loss and Durfee and Rogers (2019) uses the more general Azuma-Hoeffding bound. Here, we use Lemma 8 to bound the privacy loss, which shows that existing bounds work in a more general setting (full proof in Appendix A).

Lemma 9 *Let $\mathcal{E} = \{(\delta_i, \xi_i, \rho_i) : i \in [k]\}$ and $\delta > 0$. The set $\mathcal{E}$ is $(\varepsilon_g, \delta + \sum_{i=1}^{k} \delta_i)$-differentially private under set-wise adaptive composition, where zCDP mechanisms are selected at each round and we have*

$$\varepsilon_g = \sum_{i\in[k]} (\xi_i + \rho_i) + 2\sqrt{\sum_{i\in[k]} \rho_i \ln(1/\delta)}. \tag{1}$$

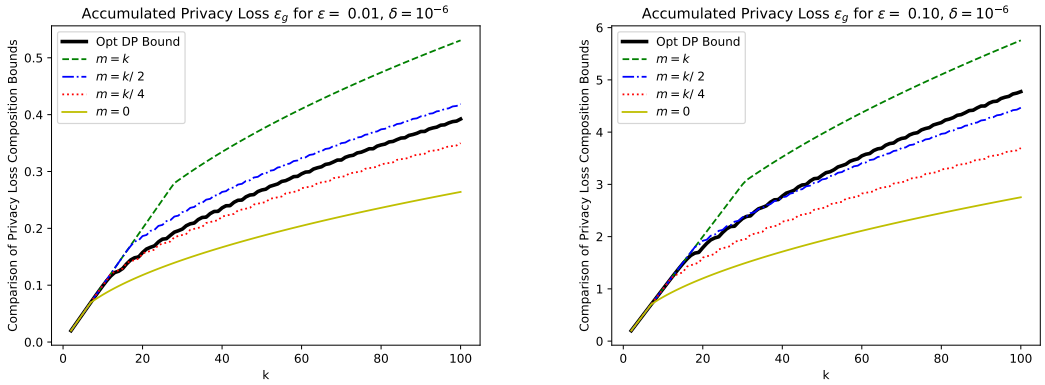

Figure 1: We plot the bounds on ε_g from (1) for k many ε-DP mechanisms, of which $k - m$ are ε-BR. We also compare with the optimal DP bound, "Opt DP Bound", from Kairouz et al. (2017).

We know from Bun and Steinke (2016) that ϵ-DP implies $\varepsilon^2/2$-zCDP and from Dong et al. (2020), we know that ε-BR is $(\varepsilon^2/8 + O(\varepsilon^4), \varepsilon^2/8)$-zCDP (see Appendix A for more detail). In Figure 1, we present the family of curves using the formula in (1) for different numbers of DP and BR mechnaisms, and compare it with the optimal DP composition bound from (Kairouz et al., 2017). Note that applying the optimal DP bound is almost the same as using the given formula in (1) when half are ε-BR mechanisms and the remaining are ε-DP.

4. Optimal Non-adaptive Composition Bounds

In this section we will present optimal bounds on the privacy loss when we combine both ε-BR ($\mathcal{M}_{\text{BR}}$) and ε-DP ($\mathcal{M}_{\text{DP}}$) mechanisms where the BR mechanisms are all preselected, prior to any interaction. Note that our analysis does allow for an analyst to select any order of ε-BR and ε-DP mechanisms in advance, and the ε-DP mechanisms can be adaptively selected, but the ε-BR mechanisms cannot.

Theorem 11 below generalizes the result in Dong et al. (2020) when m of the k mechanisms can be ε-DP, rather than all being ε-BR. We now provide a sketch of this analysis (details are in Appendix B). The following result that allows for heterogeneous ε_i at each $i \in [k]$ gives the starting point for the analysis.

Lemma 10 (Lemma 3.1 in Dong et al. (2020)) *Recall from Definition 3 we have $p_{\varepsilon_i,t_i}, q_{\varepsilon_i,t_i}$. Then we have $\delta_{OPT}(\mathcal{M}_{BR}(\varepsilon_1) \times \cdots \times \mathcal{M}_{BR}(\varepsilon_k); \varepsilon_g)$ is equal to:*

$$\sup_{\mathbf{t} \in \prod_{i\in[k]}[0,\varepsilon_i]} \sum_{S \subseteq \{1,\ldots,k\}} \left[\prod_{i \notin S} q_{\varepsilon_i,t_i} \prod_{i \in S} (1 - q_{\varepsilon_i,t_i}) - e^{\varepsilon_g} \prod_{i \notin S} p_{\varepsilon_i,t_i} \prod_{i \in S} (1 - p_{\varepsilon_i,t_i}) \right]_+ . \tag{2}$$

From here the analysis proceeds with two results from Dong et al. (2020): 1) $\delta_{\mathtt{OPT}}$ is invariant under the permutation of its inputs, 2) the expression in $\delta_{\mathtt{OPT}}$ can be maximized by a single value of $t > 0$. We can then reduce our expression for $\delta_{\mathtt{OPT}}(\mathcal{M}_{\mathtt{BR}} \times \cdots \times \mathcal{M}_{\mathtt{BR}} \times \mathcal{M}_{\mathtt{DP}} \times \cdots \times \mathcal{M}_{\mathtt{DP}}; \varepsilon_g)$ with $k - m$ many $\mathcal{M}_{\mathtt{BR}}(\varepsilon)$ and m many $\mathcal{M}_{\mathtt{DP}}(\varepsilon)$ mechanisms to the sup of $t \in [0, \varepsilon]$ of the following:

$$\sum_{i=0}^{k-m} \sum_{j=0}^{m} \binom{k-m}{i} \binom{m}{j} q_{2\varepsilon,\varepsilon}^{m-j} (1 - q_{2\varepsilon,\varepsilon})^j \cdot q_{\varepsilon,t}^{k-m-i} (1 - q_{\varepsilon,t})^i \left[1 - e^{\varepsilon_g - \varepsilon(m-2j-i) - t(k-m)} \right]_+ .$$

We then factor in terms $q_{\varepsilon,t}$ for rounds with ε-BR and $q_{2\varepsilon,\varepsilon}$ for rounds with ε-DP. We then maximize the expression by maximizing a differentiable function that matches with the optimal privacy parameter function. Due to some convenient cancellations, we obtain the following simplified formula. Full details of the proof can be found in Appendix B.

Theorem 11 *Consider the non-adaptive sequence of m many $\mathcal{M}_{DP}(\varepsilon)$ mechanisms and $k - m$ many $\mathcal{M}_{BR}(\varepsilon)$ mechanisms. We define $t_\ell^* = \frac{\varepsilon_g + \varepsilon(\ell+1-m)}{k-m+1}$ if $t_\ell^* \in [0, \varepsilon]$, otherwise we round it to the closest point in $\{0, \varepsilon\}$, for $\ell \in \{0, \cdots, k+m\}$. Then for $\varepsilon, \varepsilon_g > 0$ we have the formula for $\delta_{OPT}(\mathcal{M}_{BR}(\varepsilon_1) \times \cdots \times \mathcal{M}_{BR}(\varepsilon_{k-m}) \times \mathcal{M}_{DP}(\varepsilon_{k-m+1}) \times \cdots \times \mathcal{M}_{DP}(\varepsilon_k); \varepsilon_g)$ with $\varepsilon_i = \varepsilon$.*

$$\max_{\ell \in \{0,\cdots,k+m\}} \sum_{i=0}^{k-m} \sum_{j=0}^{m} \binom{k-m}{i} \binom{m}{j} q_{2\varepsilon,\varepsilon}^{m-j} (1 - q_{2\varepsilon,\varepsilon})^j \cdot q_{\varepsilon,t_\ell^*}^{k-m-i} (1 - q_{\varepsilon,t_\ell^*})^i \left[1 - e^{\varepsilon_g - \varepsilon(m-2j-i) - t_\ell^*(k-m)} \right]_+ .$$

Note that when $m = k$ we recover the optimal privacy loss bounds for DP mechanisms from Kairouz et al. (2017) and Murtagh and Vadhan (2016), and the expression in Theorem 11 becomes independent of t_ℓ^*. Hence, we do not need to do a max over $2k + 1$ terms, and it can be computed in $O(k)$ time. When $m = 0$, we recover the expression from Dong et al. (2020) and it can be computed in $O(k^2)$ time. In the case when $m = \Theta(k)$ our formula can take $O(k^3)$ time to calculate. We present the family of curves for various values of m in Figure 2 (left plot). These bounds allow us to interpolate between the two previous optimal composition bounds for ε-DP mechanisms from Kairouz et al. (2017) (setting $m = k$) and ε-BR mechanisms from Dong et al. (2020) (setting $m = 0$).

5. Optimal Adaptive Composition Bounds

We next turn to the adaptive setting, where each ε-BR and ε-DP mechanism can be selected as a function of previous outcomes. From Dong et al. (2020), we know that the privacy loss can strictly increase when compared to the non-adaptive setting. Note that we can use the privacy loss bounds from Lemma 9 here, even in the heterogenous privacy parameter case. However, we seek to better understand the adaptive setting by studying whether the ordering of DP and BR alters the total privacy loss. As in the previous section, we will only consider a homogenous parameter, $\varepsilon > 0$, and leave the heterogenous case to future work.

We start by showing the worst case ordering is to select the BR mechanisms all at the end of the entire interaction. Hence, we might be able to decrease the overall privacy loss if it is known that the BR mechanisms will not all be selected at the end. We defer all proofs in this section to Appendix C.

Proposition 1 *Let $\boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{M}}'$ be adaptively selected sequences of $\mathcal{M}_{BR}$ and $\mathcal{M}_{DP}$, where either may be empty, then $\delta_{OPT}(\boldsymbol{\mathcal{M}}, \mathcal{M}_{BR}, \mathcal{M}_{DP}, \boldsymbol{\mathcal{M}}'; \varepsilon_g) \leq \delta_{OPT}(\boldsymbol{\mathcal{M}}, \mathcal{M}_{DP}, \mathcal{M}_{BR}, \boldsymbol{\mathcal{M}}'; \varepsilon_g), \forall \varepsilon_g \geq 0$.*

Although we have a worst case ordering of $\mathcal{M}_{\text{BR}}$ and $\mathcal{M}_{\text{DP}}$ mechanisms, we still want to know if the ordering of these mechanisms leads to *strictly* larger privacy losses. The following result shows that the placement of a single ε-BR mechanism in a sequence of ε-DP mechanisms does not effect δ_{OPT}. So, to detect ordering dependence we will need to examine a sequence containing at least two ε-BR mechanisms.

Proposition 2 *Let $\varepsilon > 0, N \in \mathbb{N}$, and let $\boldsymbol{\mathcal{A}}_N, \boldsymbol{\mathcal{B}}_N$ be sequences of mechanisms $(\mathcal{A}_1 \dots \mathcal{A}_N)$, $(\mathcal{B}_1 \dots \mathcal{B}_N)$ for which $\exists k_{\mathcal{A}}, k_{\mathcal{B}} \in [N]$ s.t. $\mathcal{A}_{k_{\mathcal{A}}}, \mathcal{B}_{k_{\mathcal{B}}} = \mathcal{M}_{BR}$ and for $i \neq k_{\mathcal{A}}, j \neq k_{\mathcal{B}}$ $\mathcal{A}_i, \mathcal{B}_j = \mathcal{M}_{DP}$. Then for all $\varepsilon_g \in \mathbb{R}$ and all $\boldsymbol{\mathcal{A}}_N, \boldsymbol{\mathcal{B}}_N$: $\delta_{OPT}(\boldsymbol{\mathcal{A}}_N; \varepsilon_g) = \delta_{OPT}(\boldsymbol{\mathcal{B}}_N; \varepsilon_g)$.*

The following corollary states that we can use the non-adaptive, optimal DP composition formula from Theorem 11 with $m = 1$, even when the single BR mechanism can be adaptively selected at any round.

Corollary 12 *Let $\boldsymbol{\mathcal{A}}$ be any sequence of $k-1$ adaptively selected $\mathcal{M}_{DP}$ mechanisms and a single $\mathcal{M}_{BR}$ that can be adaptively selected, e.g. $\boldsymbol{\mathcal{A}} = (\mathcal{M}_{DP}, \cdots, \mathcal{M}_{DP}, \mathcal{M}_{BR}, \mathcal{M}_{DP}, \cdots, \mathcal{M}_{DP})$. Then $\delta_{OPT}(\boldsymbol{\mathcal{A}}; \varepsilon_g)$ can be computed with Theorem 11.*

It is sufficient, however, to examine the orderings of one DP and two BR mechanisms to see that the privacy loss can strictly increase for different pre-determined orderings of BR and DP mechanisms. We have the following result.

Lemma 13 *If $0 \leq \varepsilon_g < \varepsilon$ then $\delta_{OPT}(\mathcal{M}_{DP}, \mathcal{M}_{BR}, \mathcal{M}_{BR}; \varepsilon_g) > \delta_{OPT}(\mathcal{M}_{BR}, \mathcal{M}_{DP}, \mathcal{M}_{BR}; \varepsilon_g)$. If $\varepsilon_g > \varepsilon$ then the two terms are equal.*

Knowing that there is a difference in the overall privacy loss when an analyst changes the order of BR and DP mechanisms, we then plot the ratio between $\delta_{\text{OPT}}(\mathcal{M}_{\text{BR}}, \mathcal{M}_{\text{DP}}, \mathcal{M}_{\text{BR}}; \varepsilon_g)$ and $\delta_{\text{OPT}}(\mathcal{M}_{\text{DP}}, \mathcal{M}_{\text{BR}}, \mathcal{M}_{\text{BR}}; \varepsilon_g)$ in Figure 2 (right plot). See Appendix C for more details.

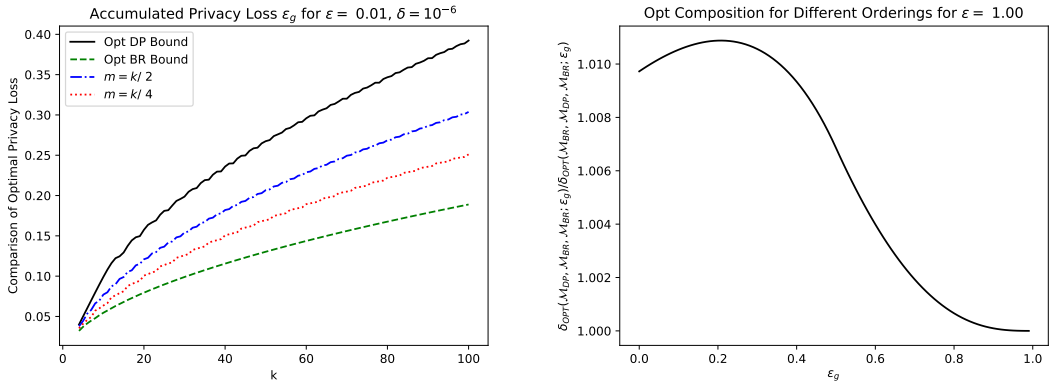

Figure 2: (Left) Plotting $\delta_{\texttt{OPT}}$ from Theorem 11 with $k-m$ many being ε-BR. (Right) Ratio between $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}; \varepsilon_g)$ and $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{BR}}; \varepsilon_g)$ with $\varepsilon = 1.0$ and $\varepsilon_g < \varepsilon$.

6. Comparing Gaussian and Laplace Mechanisms for Private Histograms

We have shown how we can provide overall privacy loss bounds for an analyst that adaptively selects different classes of private mechanisms (BR, pure DP, zCDP). However, in designing a privacy system it might be better to focus on one class of private mechanism over another if it can lead to a smaller privacy loss. We then compare the Laplace mechanism (a pure DP mechanism) and the Gaussian mechanism (a zCDP mechanism) over multiple rounds of composition. The main distinguishing feature of which noise to use, Laplace or Gaussian, for histograms mainly relies on which sensitivity bound we have on the quantity we want to add noise to. We define ℓ_p-sensitivity of a given function $f : \mathcal{X} \to \mathbb{R}^d$ as $\Delta_p(f) := \max_{x \sim x'} ||f(x) - f(x')||_p$.

Typically, if we have a bound on ℓ_1-sensitivity we would use the Laplace mechanism, whereas if we have a bound on ℓ_2-sensitivity we would use the Gaussian mechanism, since each mechanism adds noise with standard deviation proportional to the ℓ_1 or ℓ_2 sensitivity, respectively, to ensure DP. In machine learning applications, one typically normalizes high dimensional vectors by some ℓ_2 bound, so it is then natural to use Gaussian noise, not to mention the *moment accounting* composition that also takes advantage of subsampling at each round of stochastic gradient descent (Abadi et al., 2016).

However, when one wants to privatize counts from a histogram, one typically does not have an ℓ_1 or ℓ_2 sensitivity bound. Rather, one has a bound on the number of distinct counts a user can contribute to in the histogram, i.e. an ℓ_0-sensitivity bound, and a bound on the amount a user can impact a single element's count, i.e, an ℓ_∞-sensitivity bound. Obviously, we can convert these bounds to ℓ_1 or ℓ_2 sensitivity bounds. Let τ be the ℓ_∞-sensitivity and Δ be the ℓ_0-sensitivity of the histogram $h \in \mathbb{N}^d$, computed on the input data. We then have $\Delta_1(h) = \tau\Delta$ and $\Delta_2(h) = \tau\sqrt{\Delta}$.

Hence, it would seem better to use the Gaussian mechanism to release counts, since the ℓ_2-sensitivity can be much lower than the ℓ_1-sensitivity, thus leading to more accurate counts (see Appendix D for a more detailed comparison). We then propose variants of existing private top-k mechanisms that use Laplace noise (see Rogers et al. (2020) for some examples) but now use Gaussian noise instead.

For histogram inputs where one user can impact an arbitrary number of counts (also known as the unrestricted sensitivity setting), the exponential mechanism (using Gumbel noise) can first

discover the elements in the top-k and then use the Laplace mechanism with noise $\texttt{Lap}(\tau/\varepsilon)$ to release the counts on the discovered elements. The privacy loss then only scales with $\varepsilon\sqrt{k\ln(1/\delta)}$. A simple modification of this algorithm is to still use exponential mechanisms to discover the elements but then use Gaussian noise on the resulting counts. However, simply adding noise ignores a useful detail that the first phase of exponential mechanisms is giving a ranked list of elements. Hence, we propose a simple post-processing function of the Gaussian mechanism that will respect the order given by the domain discovery. We solve a constrained least squares problem to return the maximum likelihood estimator for the true counts given an ordering, with results in Appendix D.

This MLE is a post-processing function of the Gaussian mechanism, hence the privacy loss remains the same. Note that we could have similarly used ℓ_1 loss to find the MLE when using Laplace noise, however this would not guarantee a unique solution. An additional advantage of using the ℓ_2 loss is that we can use standard constrained least squares numerical methods, e.g. 'nnls' in scipy.

Next, we consider the unknown domain setting and histogram inputs having restricted ℓ_0-sensitivity of $\Delta \ll d$. The procedure $\texttt{UnkLap}^{\Delta,\bar{d},\tau}$ from Durfee and Rogers (2019); Rogers et al. (2020) adds Laplace noise to the $\bar{d} < d$ counts that it has access to from the true histogram and adds $\texttt{Lap}(\Delta\tau/\epsilon)$ to each count. Then it includes a data dependent threshold, which also has Laplace noise added to it and only returns the elements above the noisy threshold. The resulting algorithm is (ϵ, δ)-DP. We propose a variant of this algorithm in Algorithm 1, using a symmetric truncated Gaussian, which we write as $\texttt{TruncGauss}^{\Delta,\bar{d},\tau}$ with truncation value $T > 0$ that has the following density: f_T for $\mu - T \leq z \leq \mu + T$,

$$f_T(z;\mu,\sigma) = \frac{1}{\sigma}\frac{\phi\left(\frac{z-\mu}{\sigma}\right)}{\Phi\left(\frac{T}{\sigma}\right)-\Phi\left(\frac{-T}{\sigma}\right)},$$

and otherwise $f_T(z;\mu,\sigma) = 0$. Further the parameter $\bar{d}$ needs to be larger than the domain of the histogram to ensure the full histogram is available to add counts to. In related work, Geng et al. (2018) studied the truncated Laplace mechanism and showed the benefits of it over Gaussian noise, however they did not study composition nor top-k.

Algorithm 1 $\texttt{TruncGauss}^{\Delta,\bar{d},\tau}$; Truncated Gaussian mechanism over unknown domain with access to largest $\bar{d}$ elements, ℓ_∞-sensitivity τ, and Δ-restricted sensitivity.

Input: Sorted histogram $h_{(1)} \geq h_{(2)} \geq \cdots \geq h_{(\bar{d}+1)} = 0$, Δ sensitivity, and τ, σ.
Output: Ordered set of indices and counts.
Initialize $S = \emptyset$ and set T to solve the following expression: $\delta = \Delta\left(1 - \frac{\Phi\left(\frac{T}{\tau\sigma}\right)-\Phi\left(\frac{\tau-T}{\tau\sigma}\right)}{\Phi\left(\frac{T}{\tau\sigma}\right)-\Phi\left(\frac{-T}{\tau\sigma}\right)}\right)$.
for $i \leq \bar{d}$ **do**
  Set $v_i = \texttt{N}_T\left(h_{(i)}, \tau^2\sigma^2\right)$
  if $v_i > \tau + T$ **then**
    $S \leftarrow S \cup \{(i, v_i)\}$
  end
end
Return S

In comparison to $\texttt{UnkLap}^{\Delta,\bar{d},\tau}$, we need only add noise proportional to roughly $\tau\sigma \approx \sqrt{\Delta}\tau/\varepsilon$ to ensure (ε, δ)-DP. Further, the threshold T grows with $\approx \sqrt{\Delta}\ln(1/\delta)/\varepsilon$ rather than $\Delta\ln(1/\delta)/\varepsilon$ as in $\texttt{UnkLap}^{\Delta,\bar{d},\tau}$. We also have the following privacy result.

Lemma 14 *The procedure* $\texttt{TruncGauss}^{\Delta,\bar{d},\tau}$ *is* δ*-approximately* $\left(0, \frac{\Delta}{2\sigma^2}\right)$*-zCDP.*

Proof We begin by proving that for a single count $\mathtt{N}_T\left(h_i, \tau^2\sigma^2\right)$ ensures δ/Δ-approximate ρ-zCDP. This follows a similar analysis as in Geng et al. (2018) to prove truncated Laplace noise ensures approximate DP. WLOG, let $h_i' = h_i + \tau$. We consider the event $E_i = \{h_i' - T \leq \mathtt{N}_T\left(h_i, \tau^2\sigma^2\right) \leq h_i + T\}$ and compute its probability,

$$\Pr[E_i] = \int_{h_i'-T}^{h_i+T} f_T(z; h_i, \tau\sigma)dz = \frac{\Phi\left(\frac{T}{\tau\sigma}\right) - \Phi\left(\frac{\tau-T}{\tau\sigma}\right)}{\Phi\left(\frac{T}{\tau\sigma}\right) - \Phi\left(\frac{-T}{\tau\sigma}\right)}.$$

Further, we consider the event $E_i' = \{h_i' - T \leq \mathtt{N}_T\left(h_i', \tau^2\sigma^2\right) \leq h_i + T\}$. Due to symmetry, we have $\Pr[E_i] = \Pr[E_i']$. Note that we set T so that for a given $\delta > 0$ we solve (12), hence $\Pr[E_i] = \Pr[E_i'] \geq 1 - \delta/\Delta$. We then consider the Rényi divergence between two truncated Gaussians conditioned on the events that their supports overlap. Consider the scale $\alpha > 1$ and set constant $c = \frac{1}{\sqrt{2\pi\tau^2\sigma^2}} \frac{1}{\Phi\left(\frac{T}{\tau\sigma}\right) - \Phi\left(\frac{\tau-T}{\tau\sigma}\right)}$.

$$\begin{aligned}
&\exp\left((\alpha-1)D_\alpha(\mathtt{N}_T\left(h_i, \tau^2\sigma^2\right)|_{E_i} || \mathtt{N}_T\left(h_i', \tau^2\sigma^2\right)|_{E_i'})\right) \\
&= c \cdot \int_{h_i'-T}^{h_i+T} \exp\left(-\alpha\left(\frac{(z-h_i)^2}{2\tau^2\sigma^2}\right) - (1-\alpha)\left(\frac{(z-h_i')^2}{2\tau^2\sigma^2}\right)\right) dz \\
&= c \cdot \exp\left(\frac{\alpha(\alpha-1)}{2\sigma^2}\right) \cdot \int_{h_i'-T}^{h_i+T} \exp\left(\frac{-1}{2\tau^2\sigma^2}\left(z - (\alpha h_i + (1-\alpha)h_i')\right)^2\right) dz \\
&= \exp\left(\frac{\alpha(\alpha-1)}{2\sigma^2}\right) \frac{\Phi\left(\frac{T-(1-\alpha)\tau}{\tau\sigma}\right) - \Phi\left(\frac{\alpha\tau-T}{\tau\sigma}\right)}{\Phi\left(\frac{T}{\tau\sigma}\right) - \Phi\left(\frac{\tau-T}{\tau\sigma}\right)} \leq \exp\left(\frac{\alpha(\alpha-1)}{2\sigma^2}\right).
\end{aligned}$$

The last inequality holds due to the fact that the numerator is smaller than the denominator for all $\alpha > 1$. We next consider switching the order in the divergence. Let $c' = \frac{1}{\sqrt{2\pi\tau^2\sigma^2}} \frac{1}{\Phi\left(\frac{T-\tau}{\tau\sigma}\right) - \Phi\left(\frac{-T}{\tau\sigma}\right)}$

$$\begin{aligned}
&\exp\left((\alpha-1)D_\alpha(\mathtt{N}_T\left(h_i', \tau^2\sigma^2\right)|_{E_i'} || \mathtt{N}_T\left(h_i, \tau^2\sigma^2\right)|_{E_i})\right) \\
&= c' \cdot \int_{h_i'-T}^{h_i+T} \exp\left(-\alpha\left(\frac{(z-h_i')^2}{2\tau^2\sigma^2}\right) - (1-\alpha)\left(\frac{(z-h_i)^2}{2\tau^2\sigma^2}\right)\right) dz \\
&= c' \cdot \exp\left(\frac{\alpha(\alpha-1)}{2\sigma^2}\right) \cdot \int_{h_i'-T}^{h_i+T} \exp\left(\frac{-1}{2\tau^2\sigma^2}\left(z - (\alpha h_i' + (1-\alpha)h_i)\right)^2\right) dz \\
&= \exp\left(\frac{\alpha(\alpha-1)}{2\sigma^2}\right) \frac{\Phi\left(\frac{T-\alpha\tau}{\tau\sigma}\right) - \Phi\left(\frac{(1-\alpha)\tau-T}{\tau\sigma}\right)}{\Phi\left(\frac{T-\tau}{\tau\sigma}\right) - \Phi\left(\frac{-T}{\tau\sigma}\right)} \leq \exp\left(\frac{\alpha(\alpha-1)}{2\sigma^2}\right).
\end{aligned}$$

Hence, truncation only reduces the Rényi divergence between two shifted Gaussians, i.e.

$$D_\alpha(\mathtt{N}_T\left(h_i, \tau^2\sigma^2\right)|_{E_i} || \mathtt{N}_T\left(h_i', \tau^2\sigma^2\right)|_{E_i'}) \leq \frac{\alpha}{2\sigma^2}.$$

Now consider the setting where we have access to the full histogram and use the truncated Gaussian noise for each count. In this case, we can apply zCDP composition on the Δ counts that changed in neighboring histograms, hence by considering good events E_i, E_i' that are independent across all counts, the result is a δ-approximate $\frac{\Delta}{2\sigma^2}$-zCDP mechanism. We then apply post-processing to only return elements that are above the threshold $\tau + T$. Recall that post-processing does not increase the privacy parameters. Note that any element whose true count is less than τ can never appear in the result, due to the truncated noise. In a neighboring dataset a user can cause an element whose true count is 0 to have new count of at most τ, and thus cannot appear in the result from the neighboring dataset. ■

7. Conclusion

In this work, we have considered the impact to the overall privacy loss when an analyst can select different classes of private mechanisms, including pure DP, BR, and zCDP. We developed more general bounds on the privacy loss when the analyst can select the type of mechanism and the privacy parameters adaptively at each round, as long as the privacy parameters are predetermined and selected without replacement, thus generalizing previous works on bounding the privacy loss under composition. We then considered the optimal DP composition bounds when an analyst may select k pure DP mechanisms of which $k - m$ mechanisms are BR, both with the same privacy parameter, i.e. the homogeneous case. In the non-adaptive setting, we computed the optimal DP composition bound, which allows us to smoothly interpolate between the bounds from Kairouz et al. (2017), i.e. $m = k$ and the bounds from Dong et al. (2020), i.e. $m = 0$. In the adaptive homogeneous setting, we showed that the placement of a single BR mechanism does not change the composition bound, but provided an example with two BR mechanisms where ordering does impact privacy loss. We then studied variants of existing private top-k algorithms that are based on zCDP (Gaussian) mechanisms to obtain improved accuracy for the same privacy loss as using existing DP (Laplace) mechanisms.

We see lots of interesting directions for future work. As discussed in Dong et al. (2020), determining the computational complexity of the optimal DP composition for adaptively selected BR mechanisms, even in the homogenous case, is incredibly interesting. We found that with 3 mechanisms, 2 of which are BR, the $\arg\sup$ values were weighted sums of ε and ε_g. Does this generalize beyond $k = 3$? Further, for composition of heterogenous DP mechanisms, does the ordering of the privacy parameters matter if they can be adaptively selected?

Acknowledgments

We would like to thank Adrian Cardoso, Koray Mancuhan, Guillaume Saint Jacques, Reza Hosseini, and Seunghyun Lee for their helpful feedback on this work. Also, special thanks to David Durfee and Jinshuo Dong for early conversations about this work.

References

Martin Abadi, Andy Chu, Ian Goodfellow, Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li 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.

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

Raghav Bhaskar, Srivatsan Laxman, Adam Smith, and Abhradeep Thakurta. Discovering frequent patterns in sensitive data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 503–512, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0055-1. doi: 10.1145/1835804.1835869. URL http://doi.acm.org/10.1145/1835804.1835869.

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

Jinshuo Dong, Aaron Roth, and Weijie J. Su. Gaussian differential privacy, 2019.

Jinshuo Dong, David Durfee, and Ryan Rogers. Optimal differential privacy composition for exponential mechanisms. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 2597–2606. PMLR, 13–18 Jul 2020. URL http://proceedings.mlr.press/v119/dong20a.html.

David Durfee and Ryan M. Rogers. Practical differentially private top-k selection with pay-what-you-get composition. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 3527–3537, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/b139e104214a08ae3f2ebcce149cdf6e-Abstract.html.

Cynthia Dwork and Guy Rothblum. Concentrated differential privacy. *arXiv:1603.01887 [cs.DS]*, 2016.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology (EUROCRYPT 2006)*, 2006a.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Theory of Cryptography Conference*, pages 265–284, 2006b.

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

Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Truncated laplacian mechanism for approximate differential privacy. *CoRR*, abs/1810.00877, 2018. URL http://arxiv.org/abs/1810.00877.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Exponential line-crossing inequalities, 2018.

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.

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

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

Jack Murtagh and Salil 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.

Ryan Rogers, Aaron Roth, Jonathan Ullman, and Salil Vadhan. Privacy odometers and filters: Pay-as-you-go composition. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 1929–1937, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

Ryan Rogers, Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kumar Kancha, Shraddha Sahay, and Parvez Ahammad. Linkedin's audience engagements API: A privacy preserving data analytics system at scale, 2020.

Stanley Warner. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

Appendix A. Omitted Proofs of Section 3

The general result in Lemma 8 follows from Howard et al. (2018), but we include a self contained proof here for completeness.

Proof [Proof of Lemma 8] We first show the following for $\lambda > 0$:

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=0}^{k} X_i\right)\right] \leq \exp(\lambda^2 b^2/2) \tag{3}$$

By our hypothesis, we have that

$$\mathbb{E}[e^{\lambda X_i - \lambda^2 B_i^2/2} \mid \mathcal{F}_{i-1}] \leq 1 \quad \forall i$$

Furthermore, we have that $M_k = \exp\left(\lambda \sum_{i=0}^{k} X_i - \lambda^2/2 \sum_{i=0}^{k} B_i^2\right)$ is a supermartingale, due to the following

$$\begin{aligned}
\mathbb{E}[M_k|\mathcal{F}_{k-1}] &= \mathbb{E}\left[\exp\left(\lambda \sum_{i=0}^{k} X_i - \lambda^2/2 \sum_{i=0}^{k} B_i^2\right) \mid \mathcal{F}_{k-1}\right] \\
&= \exp\left(\lambda \sum_{i=0}^{k-1} X_i - \lambda^2/2 \sum_{i=0}^{k-1} B_i^2\right) \mathbb{E}\left[e^{\lambda X_k - \lambda^2 B_k^2/2} \mid \mathcal{F}_{k-1}\right] \\
&\leq \exp\left(\lambda \sum_{i=0}^{k-1} X_i - \lambda^2/2 \sum_{i=0}^{k-1} B_i^2\right) = M_{k-1}.
\end{aligned}$$

Hence, we have $\mathbb{E}[M_k] \leq 1$. Using our assumption that $\sum_{i=0}^{k} B_i^2 \leq b$, we have the following

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=0}^{k} X_i - \lambda^2 b^2/2\right)\right] \leq \mathbb{E}\left[\exp\left(\lambda \sum_{i=0}^{k} X_i - \lambda^2 \sum_{i=0}^{k} B_i^2/2\right)\right] \leq 1$$

and hence (3) holds. We can now prove the result using a standard argument involving Markov's inequality.

$$\Pr[\sum_{i=0}^{k} X_i > \beta] = \Pr[e^{\lambda \sum_{i=0}^{k} X_i - \lambda^2 b^2/2} \geq e^{\lambda\beta - \lambda^2 b^2/2}] \leq \frac{1}{e^{\lambda\beta - \lambda^2 b^2/2}}$$

We then set $\lambda = \frac{\beta}{b^2}$ to get the result. ■

We then define the privacy loss random variable, which we will bound with high probability to get a bound on the overall privacy loss. At each round i, the analyst has all the necessary information from the previous outcomes to decide on which neighboring datasets to use, which class of mechanism to select, which of the remaining privacy parameters to choose, and which specific mechanism M_i to pick. The previous outcomes and choices of the analyst up to round i are random variables from the sigma algebra $\mathcal{F}_{i-1}$. The selected randomized mechanism M_i takes the input

to some outcome set $\mathcal{Y}_i$ and for each $y_i \in \mathcal{Y}_i$ we define the following for neighboring datasets $x^{(0)}, x^{(1)}$

$$L_i^{(b)}(y_i) := \ln\left(\frac{\Pr[M_i(x^{(b)}) = y_i \mid \mathcal{F}_{i-1}]}{\Pr[M_i(x^{(1-b)}) = y_i \mid \mathcal{F}_{i-1}]}\right) \qquad L_i := L_i^{(b)}(Y_i) \text{ where } Y_i \sim M_i(x^{(b)}) \mid \mathcal{F}_{i-1}.$$

We then consider the full privacy loss over the entire $\mathcal{E}$-set-wise adaptive composition experiment, $\sum_{i=1}^k L_i$. Similar to Dwork et al. (2010), we aim to bound the accumulated privacy loss with high probability, i.e. $\Pr[\sum_{i=1}^k L_i \geq \varepsilon_g] \leq \delta$, so that the $\mathcal{E}$-set-wise adaptive composition experiment is (ε_g, δ)-DP.

Proof [Proof of Lemma 9] Consider the privacy loss $L_i^{(b)}$ at each round i conditioned on the events $E_i^{(0)}, E_i^{(1)}$ where $\Pr[E_i^{(b)}] \geq 1 - \delta_i$ for $b \in \{0, 1\}$ and each round $i \in [k]$.

$$L_i^{(b)}(y_i) := \ln\left(\frac{\Pr[M_i(x_i^{(b)}) = y_i \mid E_i^{(b)}, \mathcal{F}_{i-1}]}{\Pr[M_i(x_i^{(1-b)} = y_i \mid E_i^{(1-b)}, \mathcal{F}_{i-1}]}\right)$$
$$L_i^{(b)} := L_i^{(b)}(Y_i) \text{ where } Y_i \sim M_i(x_i^{(b)}) \mid E_i^{(b)}, \mathcal{F}_{i-1}.$$

We use Lemma 8 with $X_i = L_i^{(b)} - \xi_i - \rho_i$ and $B_i^2 = 2\rho_i$. From the definition of zCDP we have the following for any $\lambda = \alpha - 1 > 0$

$$\mathbb{E}[\exp\left(\lambda(L_i^{(b)} - \xi_i - \rho_i)\right) \mid \mathcal{F}_{i-1}] \leq \exp\left(\lambda^2 \rho_i\right)$$
$$\implies \mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^k (L_i^{(b)} - \xi_i - \rho_i) - \lambda^2 \sum_{i=1}^k \rho_i\right)\right] \leq 1.$$

Recall that at each round $i \in [k]$ the index of the parameters is a random variable, based on the outcomes of previous results, but the resulting sum of parameters is predetermined. Hence, we have

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^k (L_i^{(b)} - \xi_i - \rho_i)\right)\right] \leq \exp\left(\lambda^2 \sum_{i=1}^k \rho_i\right).$$

The rest of the analysis is identical to the proof of Lemma 3.5 from Bun and Steinke (2016), which gives the conversion of zCDP to approximate DP. Hence, we have

$$\Pr[V^{(b)} \in S \mid \cup_{i=1}^k E_i^{(b)}] \leq e^{\varepsilon_g} \Pr[V^{(1-b)} \in S \mid \cup_{i=1}^k E_i^{(1-b)}] + \delta$$

We then follow Lemma 8.8 from Bun and Steinke (2016). Without loss of generality we set $\Pr[\cup_i E_i^{(b)}] = 1 - \delta'$ and $\Pr[\cup_i E_i^{(1-b)}] \geq 1 - \delta'$ where $\delta' \leq \sum_i \delta_i$. We then have

$$\Pr[V^{(b)} \in S] \leq \Pr[V^{(b)} \in S \mid \cup_{i=1}^k E_i^{(b)}](1 - \delta') + \delta'$$

and

$$\Pr[V^{(1-b)} \in S] \geq \Pr[V^{(1-b)} \in S \mid \cup_{i=1}^k E_i^{(1-b)}](1 - \delta')$$

Putting everything together, we have

$$\begin{aligned}\Pr[V^{(b)} \in S] &\leq \Pr[V^{(b)} \in S \mid \cup_{i=1}^{k} E_i^{(b)}](1-\delta') + \delta' \\ &\leq \left(e^{\varepsilon_g} \Pr[V^{(1-b)} \in S \mid \cup_{i=1}^{k} E_i^{(1-b)}] + \delta\right)(1-\delta') + \delta' \\ &\leq e^{\varepsilon_g} \Pr[V^{(1-b)} \in S] + \delta(1-\delta') + \delta' \\ &\leq e^{\varepsilon_g} \Pr[V^{(1-b)} \in S] + \delta + \sum_{i=1}^{k} \delta_i.\end{aligned}$$

■

We can also use concentrated differential privacy (CDP) Dwork and Rothblum (2016) as a privacy definition. Note that there are other variants related to CDP, including Rényi DP (RDP) Mironov (2017). We first define the privacy loss random variable in terms of two random variables Y and Z over the same support

$$L_{Y||Z} := \ln\left(\frac{\Pr[Y=y]}{\Pr[Z=y]}\right) \qquad \text{where } y \sim Y.$$

Definition 15 (Concentrated Differential Privacy) *A randomized algorithm M is (μ,τ)-CDP if for all neighboring inputs x, x', we have $\mathbb{E}[L_{M(x)||M(x')}] \leq \mu$ and for any $\lambda \in \mathbb{R}$ we have $\mathbb{E}\left[\exp\left(\lambda\left(L_{M(x)||M(x')} - \mathbb{E}[L_{M(x)||M(x')}]\right)\right)\right] \leq e^{\lambda^2\cdot\tau^2/2}$.*

If one wanted to use *concentrated differential privacy* from Dwork and Rothblum (2016) as the class of private mechanisms, we then focus on the following mean zero random variable,

$$X_i = L_i - \mathbb{E}[L_i \mid \mathcal{F}_{i-1}]. \tag{4}$$

We then have the following connection between CDP and zCDP.

Lemma 16 (Bun and Steinke (2016)) *If M is (μ,τ)-CDP, then M is also $(\mu-\tau^2/2, \tau^2/2)$-zCDP.*

We then need bounds on $\mathbb{E}[L_i \mid \mathcal{F}_{i-1}]$ and on the subgaussian parameter for each X_i and this can then be converted into zCDP parameters.

Lemma 17 *For X_i given in (4), if the analyst selects an ε-DP mechanism M_i given $\mathcal{F}_{i-1}$, then for all $\lambda \in \mathbb{R}$ we have,*

$$\mathbb{E}[\exp(\lambda X_i) \mid \mathcal{F}_{i-1}] \leq e^{\lambda^2\varepsilon^2/2} \qquad \& \qquad \mathbb{E}[L_i \mid \mathcal{F}_{i-1}] \leq \varepsilon\left(\frac{e^\varepsilon - 1}{e^\varepsilon + 1}\right).$$

If the analyst selects an α-BR mechanism $\mathcal{M}_i$ given $\mathcal{F}_{i-1}$ then for all $\lambda \in \mathbb{R}$ we have

$$\mathbb{E}[\exp(\lambda X_i) \mid \mathcal{F}_{i-1}] \leq e^{\lambda^2\alpha^2/8} \qquad \& \qquad \mathbb{E}[L_i \mid \mathcal{F}_{i-1}] \leq \frac{\alpha}{e^\alpha - 1} - 1 - \ln\left(\frac{\alpha}{e^\alpha - 1}\right).$$

If the analyst selects a (μ,τ)-CDP mechanism M_i given $\mathcal{F}_{i-1}$, then for all $\lambda \in \mathbb{R}$, we have,

$$\mathbb{E}[\exp(\lambda X_i) \mid \mathcal{F}_{i-1}] \leq e^{\lambda^2\tau^2/2} \qquad \& \qquad \mathbb{E}[L_i \mid \mathcal{F}_{i-1}] \leq \mu.$$

Proof The statement about the CDP mechanism is simply due to the definition of (μ, τ)-CDP Dwork and Rothblum (2016). The expectations of the privacy losses for ε-DP and α-BR are from Kairouz et al. (2017) and Dong et al. (2020), respectively. From Hoeffding (1963), we know that for bounded random variables $X_i \in [a, b]$ that X is subgaussian with parameter $(b-a)^2/4$. Hence, for ε-DP, we have $b - a \leq 2\varepsilon$ and for α-BR, we have $b - a \leq \alpha$. ■

Consider the setting where the analyst is allowed to select m_{DP} pure-DP mechanisms, m_{BR} BR mechanisms, and m_{CDP} concentrated DP mechanisms Dwork and Rothblum (2016). Furthermore, there are preregistered privacy parameters $\mathcal{E}^{\text{DP}} = \{\varepsilon_i = \varepsilon > 0 : i \in [m_{\text{DP}}]\}$ for pure DP mechanisms, $\mathcal{E}^{\text{BR}} = \{\alpha_i = \alpha > 0 : i \in [m_{\text{BR}}]\}$ for BR mechanisms, and $\mathcal{E}^{\text{CDP}} = \{(\mu_i, \tau_i) = (\mu, \tau) > (0, 0) : i \in [m_{\text{CDP}}]\}$ for CDP mechanisms. We allow the analyst to adaptively select the class of mechanism at each round i adaptively and to also select the privacy parameter from the corresponding class of mechanisms, without replacement. We then apply Lemma 9 and Lemma 8 to replace the formula in (1) to get

$$\varepsilon_g = m_{\text{DP}}\varepsilon\left(\frac{e^\varepsilon - 1}{e^\varepsilon + 1}\right) + m_{\text{BR}}\left(\frac{\alpha}{e^\alpha - 1} - 1 - \ln\left(\frac{\alpha}{e^\alpha - 1}\right)\right) + m_{\text{CDP}}\mu + \sqrt{2\left(m_{\text{DP}}\varepsilon^2 + \frac{m_{\text{BR}}}{4}\alpha^2 + m_{\text{CDP}}\tau^2\right)\ln(1/\delta)}. \tag{5}$$

Note that when $m_{\text{DP}} = k$, we get the traditional *advanced* composition bound from Dwork et al. (2010) with refinement from Kairouz et al. (2017) and when $m_{\text{BR}} = k$, we get the bound for composing ε-BR mechanisms from Corollary 3.1 in Dong et al. (2020).

We also have the following formula for the optimal composition bound for homogenous, adaptively selected pure DP mechanisms.

Theorem 18 (Optimal Homogeneous DP Composition Kairouz et al. (2017)) *For every $\varepsilon > 0$ and $\varepsilon_g \geq 0$, we have the following where $\varepsilon_i = \varepsilon$ for all $i \in [k]$*

$$\delta_{OPT}((\mathcal{M}_{DP}(\varepsilon_1), \cdots, \mathcal{M}_{DP}(\varepsilon_k)); \varepsilon_g) = \frac{1}{(1+e^\varepsilon)^k} \sum_{\ell=\left\lceil \frac{\varepsilon_g + k\varepsilon}{2\varepsilon} \right\rceil}^{k} \binom{k}{\ell} \left(e^{\ell\varepsilon} - e^{\varepsilon_g + (k-\ell)\varepsilon}\right).$$

In Figure 1, we present the family of curves from (5) for different values of $m = m_{\text{DP}}$, i.e. the number of ε-DP mechanisms, $k - m = m_{\text{BR}}$ while fixing $m_{\text{CDP}} = 0$, and compare it with the optimal DP composition bound from Theorem 18. Note that applying the optimal DP bound is almost the same as using the given formula when half are BR mechanisms and the remaining half are DP.

Appendix B. Omitted Proofs of Section 4

We start by defining the elements that we take the sup over in (2) for general $(t_1, \cdots, t_k)$ and corresponding privacy parameters $(\varepsilon_1, \cdots, \varepsilon_k)$

$$\delta((t_1, \varepsilon_1) \times \cdots \times (t_k, \varepsilon_k); \varepsilon_g) := \sum_{S \subseteq \{1,\ldots,k\}} \left[\prod_{i \notin S} q_{\varepsilon_i, t_i} \prod_{i \in S}(1 - q_{\varepsilon_i, t_i}) - e^{\varepsilon_g} \prod_{i \notin S} p_{\varepsilon_i, t_i} \prod_{i \in S}(1 - p_{\varepsilon_i, t_i})\right]_+ . \tag{6}$$

We then have the following result that shows that the ordering of mechanisms does not modify the value of $\delta_{\texttt{OPT}}$.

Lemma 19 *Let $\pi : [k] \to [k]$ be a permutation on the indices $[k]$, we then have*

$$\delta((t_1, \varepsilon_1) \times \cdots \times (t_k, \varepsilon_k); \varepsilon_g) = \delta((t_{\pi(1)}, \varepsilon_{\pi(1)}), \cdots, (t_{\pi(k)}, \varepsilon_{\pi(k)}); \varepsilon_g)$$

Proof Note that the expression in (6) takes a summation over all possible subsets of indices. Hence, if we permute the indices, the full summation has the same terms. ■

We will focus only on the homogeneous case, where all the privacy parameters are the same and leave the heterogeneous case to future work. We have the immediate result from Lemma 19.

Lemma 20 *Let $\boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{M}}'$ be two sequences of non-adaptively selected mechanisms where m are $\mathcal{M}_{DP}(\varepsilon)$ and $k - m$ are from $\mathcal{M}_{BR}(\varepsilon)$. We then have for any $\varepsilon_g > 0$ and $\varepsilon > 0$, $\delta_{OPT}(\boldsymbol{\mathcal{M}}; \varepsilon_g) = \delta_{OPT}(\boldsymbol{\mathcal{M}}'; \varepsilon_g)$.*

Recall that the big difference between ε-BR and ε-DP mechanisms, is that ε-BR mechanisms use the generalized randomized response $\texttt{RR}_{\varepsilon,t}$ for a worst case $t \in [0, \varepsilon]$ as opposed to ε-DP mechanisms which use $\texttt{RR}_{2\varepsilon,\varepsilon}$, so that $t = \varepsilon$. We then define the following function that fixes m of the t_i values to be ε which have corresponding $\varepsilon_i = 2\varepsilon$ and corresponding sets $\mathcal{S}_1 := [k - m]$ and $\mathcal{S}_2 := \{k - m + 1, \cdots, k\}$.

$$\begin{aligned}
&\delta((t_1, \varepsilon), \cdots, (t_{k-m}, \varepsilon), (\varepsilon, 2\varepsilon), \cdots, (\varepsilon, 2\varepsilon); \varepsilon_g) \\
&\quad = \sum_{S_1 \subseteq \mathcal{S}_1} \sum_{S_2 \subseteq \mathcal{S}_2} \Bigg[\prod_{i_1 \in \mathcal{S}_1 \setminus S_1} q_{\varepsilon,t_{i_1}} \prod_{i_2 \in \mathcal{S}_2 \setminus S_2} q_{2\varepsilon,\varepsilon} \prod_{j_1 \in S_1} (1 - q_{\varepsilon,t_{j_1}}) \prod_{j_2 \in S_2} (1 - q_{2\varepsilon,\varepsilon}) \\
&\qquad\qquad - e^{\varepsilon_g} \prod_{i_1 \in \mathcal{S}_1 \setminus S_1} p_{\varepsilon,t_{i_1}} \prod_{i_2 \in \mathcal{S}_2 \setminus S_2} p_{2\varepsilon,\varepsilon} \prod_{j_1 \in S_1} (1 - p_{\varepsilon,t_{j_1}}) \prod_{j_2 \in S_2} (1 - p_{2\varepsilon,\varepsilon}) \Bigg]_+ \\
&\quad = \sum_{S_1 \subseteq \mathcal{S}_1} \sum_{S_2 \subseteq \mathcal{S}_2} \Bigg[q_{2\varepsilon,\varepsilon}^{m-|S_2|} (1 - q_{2\varepsilon,\varepsilon})^{|S_2|} \prod_{i_1 \in \mathcal{S}_1 \setminus S_1} q_{\varepsilon,t_{i_1}} \prod_{j_1 \in S_1} (1 - q_{\varepsilon,t_{j_1}}) \\
&\qquad\qquad - e^{\varepsilon_g} p_{2\varepsilon,\varepsilon}^{m-|S_2|} (1 - p_{2\varepsilon,\varepsilon})^{|S_2|} \prod_{i_1 \in \mathcal{S}_1 \setminus S_1} p_{\varepsilon,t_{i_1}} \prod_{j_1 \in S_1} (1 - p_{\varepsilon,t_{j_1}}) \Bigg]_+
\end{aligned}$$

Note that we have the simple connection with this function and the optimal δ,

$$\begin{aligned}
&\delta_{\texttt{OPT}} \left(\underbrace{\mathcal{M}_{\texttt{BR}}(\varepsilon) \times \cdots \times \mathcal{M}_{\texttt{BR}}(\varepsilon)}_{k-m} \times \underbrace{\mathcal{M}_{\texttt{DP}}(\varepsilon) \times \cdots \times \mathcal{M}_{\texttt{DP}}(\varepsilon)}_{m}; \varepsilon_g \right) \\
&\qquad = \sup_{\mathbf{t} \in \prod_{i \in \{1, \cdots, k-m\}} [0, \varepsilon]} \delta((t_1, \varepsilon), \cdots, (t_{k-m}, \varepsilon), (\varepsilon, 2\varepsilon), \cdots, (\varepsilon, 2\varepsilon))
\end{aligned}$$

We then use a key result from Dong et al. (2020) that shows that despite BR mechanisms having separate $t_i \in [0, \varepsilon]$ for each BR mechanism, the worst way to set the t_i's is to set them all equal. The following is a slight generalization of Lemma 5.4 from Dong et al. (2020).

Lemma 21 *For any* $\varepsilon, \varepsilon' > 0$, $\varepsilon_g \in \mathbb{R}$, *and* $\mathbf{t} \in [0,\varepsilon]^{k-m} \times [0,\varepsilon']^m$,

$$\begin{aligned}&\delta((t_1,\varepsilon),(t_2,\varepsilon),\cdots(t_{k-m},\varepsilon),(t_{k-m+1},\varepsilon'),\cdots,(t_k,\varepsilon');\varepsilon_g)\\&\quad\le \delta\left(\left(\frac{t_1+t_2}{2},\varepsilon\right),\left(\frac{t_1+t_2}{2},\varepsilon\right),(t_3,\varepsilon),\cdots,(t_{k-m},\varepsilon),(t_{k-m+1},\varepsilon'),\cdots,(t_k,\varepsilon');\varepsilon_g\right)\end{aligned}$$

Hence, we can simplify our expression for $\delta_{\texttt{OPT}}$ substantially to the following where we use the fact that $p_{\varepsilon,t} = e^{-t}q_{\varepsilon,t}$ and $1-p_{\varepsilon,t} = e^{\varepsilon-t}(1-q_{\varepsilon,t})$

$$\delta_{\texttt{OPT}}((\mathcal{M}_{\texttt{BR}}(\varepsilon),\cdots,\mathcal{M}_{\texttt{BR}}(\varepsilon),\mathcal{M}_{\texttt{DP}}(\varepsilon),\cdots,\mathcal{M}_{\texttt{DP}}(\varepsilon));\varepsilon_g) \tag{7}$$

$$\begin{aligned}&= \sup_{t\in[0,\varepsilon]}\sum_{S_1\subseteq\mathcal{S}_1}\sum_{S_2\subseteq\mathcal{S}_2}\Big[q_{2\varepsilon,\varepsilon}^{m-|S_2|}(1-q_{2\varepsilon,\varepsilon})^{|S_2|}q_{\varepsilon,t}^{k-m-|S_1|}(1-q_{\varepsilon,t})^{|S_1|}\\&\qquad\qquad -e^{\varepsilon_g}p_{2\varepsilon,\varepsilon}^{m-|S_2|}(1-p_{2\varepsilon,\varepsilon})^{|S_2|}p_{\varepsilon,t}^{k-m-|S_1|}(1-p_{\varepsilon,t})^{|S_1|}\Big]_+\\&= \sup_{t\in[0,\varepsilon]}\sum_{i=0}^{k-m}\sum_{j=0}^{m}\binom{k-m}{i}\binom{m}{j}q_{2\varepsilon,\varepsilon}^{m-j}(1-q_{2\varepsilon,\varepsilon})^j q_{\varepsilon,t}^{k-m-i}(1-q_{\varepsilon,t})^i\left[1-e^{\varepsilon_g-\varepsilon(m-2j-i)-t(k-m)}\right]_+.\end{aligned} \tag{8}$$

We now want to show that the $\sup_{t\in[0,\varepsilon]}$ can be decomposed into a max over a finite number of values for $t\in[0,\varepsilon]$. We define the following function for $t\in[0,\varepsilon]$:

$$\delta^{m,k}(t;\varepsilon_g) := \delta(\underbrace{(t,\varepsilon),\cdots,(t,\varepsilon)}_{k-m},\underbrace{(\varepsilon,2\varepsilon),\cdots,(\varepsilon,2\varepsilon)}_{m};\varepsilon_g).$$

We define the following function in terms of $\alpha_{i,j} = \binom{k-m}{i}\binom{m}{j}q_{2\varepsilon,\varepsilon}^{m-j}(1-q_{2\varepsilon,\varepsilon})^j$ when $i\in\{0,\cdots k-m\}$ and $j\in\{0,\cdots,m\}$ with $\alpha_{i,j}=0$ otherwise.

$$F_\ell(t) := \sum_{n=0}^{\ell}\left(1-e^{\varepsilon_g-\varepsilon(m-n)-t(k-m)}\right)\sum_{\substack{i+2j=n\\ i\in\{0,1\cdots,k-m\}\\ j\in\{0,1,\cdots,m\}}}\alpha_{i,j}q_{\varepsilon,t}^{k-m-i}(1-q_{\varepsilon,t})^j$$

The following result provides a more general version of Lemma 5.7 in Dong et al. (2020), which is central to showing that we need to only consider a few values of $t\in[0,\varepsilon]$.

Lemma 22

$$F'_\ell(t) = \frac{1}{1-e^{-\varepsilon}}\left(e^{\varepsilon_g-\varepsilon(m-\ell)-t(k-m)}-e^{t-\varepsilon}\right)\sum_{\substack{i+2j=\ell\\ i\in\{0,1\cdots,k-m\}\\ j\in\{0,1,\cdots,m\}}}\alpha_{i,j}(k-m-i)q_{\varepsilon,t}^{k-m-i-1}(1-q_{\varepsilon,t})^i$$

Proof To simplify notation, we use $k' = k-m$ and $q_{\varepsilon,t} = q_t$. We have the following recurrence relation for $F_\ell(t) = F_{\ell-1}(t) + f_\ell(t)$ where

$$f_\ell(t) = \left(1-e^{\varepsilon_g-\varepsilon(m-\ell)-tk'}\right)\sum_{\substack{i+2j=\ell\\ i\in\{0,1\cdots,k'\}\\ j\in\{0,1,\cdots,m\}}}\alpha_{i,j}q_t^{k'-i}(1-q_t)^j$$

We then prove the statement by using induction $F'_\ell(t) = F'_{\ell-1}(t) + f'_\ell(t)$. We start with the base case,

$$\begin{aligned}
F'_0(t) &= \left(k' \cdot e^{\varepsilon_g - \varepsilon m - tk'}\right) \alpha_{0,0} q_t^{k'} + \left(1 - e^{\varepsilon_g - \varepsilon m - tk'}\right) \alpha_{0,0} k' q_t^{k'-1} q'_t \\
&= \frac{k' \alpha_{0,0}}{1 - e^{-\varepsilon}} q_t^{k'-1} \left(e^{\varepsilon_g - \varepsilon m - tk'}(1 - e^{t-\varepsilon}) - e^{t-\varepsilon}\left(1 - e^{\varepsilon_g - \varepsilon m - tk'}\right)\right) \\
&= \frac{k' \alpha_{0,0}}{1 - e^{-\varepsilon}} q_t^{k'-1} \left(e^{\varepsilon - \varepsilon m - tk'} - e^{t-\varepsilon}\right)
\end{aligned}$$

We now present the derivative of $f_\ell(t)$. Note that we drop the condition in the summation where $i \in \{0, \cdots, k'\}$ and $j \in \{0, \cdots, m\}$ to ease the notation.

$$\begin{aligned}
f'_\ell(t) &= k' e^{\varepsilon_g - \varepsilon(m-\ell) - tk'} \sum_{i+2j=\ell} \alpha_{i,j} q_t^{k'-i} (1 - q_t)^i \\
&\quad + \left(1 - e^{\varepsilon_g - \varepsilon(m-\ell) - tk'}\right) \sum_{i+2j=\ell} \alpha_{i,j} \left((k' - i) q_t^{k'-i-1} (1 - q_t)^i q'_t - i q_t^{k-i} (1 - q_t)^{i-1} q'_t\right) \\
&= \sum_{i+2j=\ell} \alpha_{i,j} q_t^{k'-i-1} (1 - q_t)^{i-1} \\
&\quad \left(k' e^{\varepsilon_g - \varepsilon(m-\ell) - tk'} q_t (1 - q_t) + q'_t \left(1 - e^{\varepsilon_g - \varepsilon(m-\ell) - tk'}\right) \left((k' - i)(1 - q_t) - i q_t\right)\right)
\end{aligned}$$

We now factor out a $1/(1 - e^{-\varepsilon})^2$ and the inner term becomes.

$$\begin{aligned}
(*) :=& k' e^{\varepsilon_g - \varepsilon(m-\ell) - tk'} \left(1 - e^{t-\varepsilon}\right) \left(e^{t-\varepsilon} - e^{-\varepsilon}\right) \\
&\quad - e^{t-\varepsilon} \left(1 - e^{\varepsilon_g - \varepsilon(m-\ell) - tk'}\right) \left((k' - i)(e^{t-\varepsilon} - e^{-\varepsilon}) - i(1 - e^{t-\varepsilon})\right) \\
&= k' e^{\varepsilon_g - \varepsilon(m-\ell) - tk'} \left(e^{t-\varepsilon} - e^{-\varepsilon} - e^{2(t-\varepsilon)} + e^{t-2\varepsilon}\right) \\
&\quad - \left(e^{t-\varepsilon} - e^{\varepsilon_g - \varepsilon(m-\ell+1) - t(k'-1)}\right) \left((k' - i)(e^{t-\varepsilon} - e^{-\varepsilon}) - i(1 - e^{t-\varepsilon})\right) \\
&= k' \left(e^{\varepsilon_g - \varepsilon(m-\ell+1) - t(k'-1)} - e^{\varepsilon_g - \varepsilon(m-\ell+1) - tk'} - e^{\varepsilon_g - \varepsilon(m-\ell+2) - t(k'-2)} + e^{\varepsilon_g - \varepsilon(m-\ell+2) - t(k'-1)}\right) \\
&\quad - (k' - i) \left(e^{2(t-\varepsilon)} - e^{\varepsilon_g - \varepsilon(m-\ell+2) - t(k'-2)} - e^{t-2\varepsilon} + e^{\varepsilon_g - \varepsilon(m-\ell+2) - t(k'-1)}\right) \\
&\quad + i \left(e^{t-\varepsilon} - e^{\varepsilon_g - \varepsilon(m-\ell+1) - t(k'-1)} - e^{2(t-\varepsilon)} + e^{\varepsilon_g - \varepsilon(m-\ell+2) - t(k'-2)}\right)
\end{aligned}$$

We now use the inductive claim to prove the statement. We use the fact that $\alpha_{i,j}(k'-i) = \alpha_{i+1,j}(i+1)$

$$\begin{aligned}
F'_\ell(t) &= \frac{1}{1-e^{-\varepsilon}}\left(e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{t-\varepsilon}\right)\sum_{i+2j=\ell-1}\alpha_{i,j}(k'-i)q_{\varepsilon,t}^{k'-i-1}(1-q_{\varepsilon,t})^i \\
&\quad + \sum_{i+2j=\ell}\alpha_{i,j}\frac{q_t^{k'-i-1}(1-q_t)^{i-1}}{(1-e^{-\varepsilon})^2}\cdot(*) \\
&= \frac{1}{(1-e^{-\varepsilon})^2}\sum_{i+2j=\ell-1}(i+1)\alpha_{i+1,j}\left(e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{t-\varepsilon}\right)q_t^{k'-i-1}(1-q_t)^i(1-e^{-\varepsilon}) \\
&\quad + \sum_{i+2j=\ell}\alpha_{i,j}\frac{q_t^{k'-i-1}(1-q_t)^{i-1}}{(1-e^{-\varepsilon})^2}\cdot(*) \\
&= \frac{1}{(1-e^{-\varepsilon})^2}\sum_{i+2j=\ell}\alpha_{i,j}\left(e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{t-\varepsilon}\right)q_t^{k'-i}(1-q_t)^{i-1}(1-e^{-\varepsilon})i \\
&\quad + \frac{1}{(1-e^{-\varepsilon})^2}\sum_{i+2j=\ell}\alpha_{i,j}q_t^{k'-i-1}(1-q_t)^{i-1}\cdot(*) \\
&= \frac{1}{(1-e^{-\varepsilon})^2}\sum_{i+2j=\ell}\alpha_{i,j}q_t^{k'-i-1}(1-q_t)^{i-1}\left(i\left(e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{t-\varepsilon}\right)(1-e^{t-\varepsilon}) + (*)\right)
\end{aligned}$$

We now expand the inner term by combining like terms with the i coefficient.

$$\begin{aligned}
&i\left(\left(e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{t-\varepsilon}\right)(1-e^{t-\varepsilon}) + \left(e^{t-\varepsilon} - e^{\varepsilon_g-\varepsilon(m-\ell+1)-t(k'-1)} - e^{2(t-\varepsilon)} + e^{\varepsilon_g-\varepsilon(m-\ell+2)-t(k'-2)}\right)\right) \\
&= i\left(e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{\varepsilon_g-\varepsilon(m-\ell+2)-t(k'-1)} - e^{\varepsilon_g-\varepsilon(m-\ell+1)-t(k'-1)} + e^{\varepsilon_g-\varepsilon(m-\ell+2)-t(k'-2)}\right)
\end{aligned}$$

Note that this is the same term as the negative of the coefficient on k'. We then combine the terms with $(k'-i)$ coefficient.

$$\begin{aligned}
&-(k'-i)\left(e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{\varepsilon_g-\varepsilon(m-\ell+2)-t(k'-1)} - e^{\varepsilon_g-\varepsilon(m-\ell+1)-t(k'-1)} + e^{\varepsilon_g-\varepsilon(m-\ell+2)-t(k'-2)}\right) \\
&\qquad -(k'-i)\left(e^{2(t-\varepsilon)} - e^{\varepsilon_g-\varepsilon(m-\ell+2)-t(k'-2)} - e^{t-2\varepsilon} + e^{\varepsilon_g-\varepsilon(m-\ell+2)-t(k'-1)}\right) \\
&\qquad\quad = -(k'-i)\left(e^{2(t-\varepsilon)} - e^{t-2\varepsilon} + e^{\varepsilon_g-\varepsilon(m-\ell+1)-tk'} - e^{\varepsilon_g-\varepsilon(m-\ell+1)-t(k'-1)}\right) \\
&\qquad\qquad = (k'-i)(e^{t-\varepsilon} - e^{-\varepsilon})\left(e^{\varepsilon_g-\varepsilon(m-\ell)-tk'} - e^{t-\varepsilon}\right)
\end{aligned}$$

Putting this altogether, we have the following.

$$\begin{aligned}
F'_\ell(t) &= \frac{1}{(1-e^{-\varepsilon})^2}\sum_{i+2j=\ell}\alpha_{i,j}q_t^{k'-i-1}(1-q_t)^{i-1}(k'-i)(e^{t-\varepsilon} - e^{-\varepsilon})\left(e^{\varepsilon_g-\varepsilon(m-\ell)-tk'} - e^{t-\varepsilon}\right) \\
&= \frac{1}{1-e^{-\varepsilon}}\left(e^{\varepsilon_g-\varepsilon(m-\ell)-tk'} - e^{t-\varepsilon}\right)\sum_{i+2j=\ell}\alpha_{i,j}(k-i)q_t^{k'-i-1}(1-q_t)^i
\end{aligned}$$

This proves the statement. ■

Given this result, we can limit our search for $\sup_{t\in[0,\varepsilon]}$ to a $\max$ over $k+m$ terms.

Appendix C. Omitted Proofs of Section 5

It will be helpful in our analysis to have the following formulation of the optimal privacy loss for ε-BR mechanisms from Dong et al. (2020).

Lemma 23 *Let $\boldsymbol{\mathcal{M}}$ be a sequence of adaptively selected $\mathcal{M}_{BR}(\varepsilon_i)$ and $\mathcal{M}_{DP}(\varepsilon_i)$ where $\varepsilon_i > 0$ are fixed in advance for $i \in [k]$. For any $\varepsilon_g > 0$ and setting $\delta_{OPT}(\emptyset; \varepsilon_g) = [1 - e^{\varepsilon_g}]_+$ we have:*

$$\delta_{OPT}(\mathcal{M}_{BR}(\varepsilon_0), \boldsymbol{\mathcal{M}}; \varepsilon_g) = \sup_{t_0 \in [0, \varepsilon_0]} \{q_{\varepsilon_0, t_0} \delta_{OPT}(\boldsymbol{\mathcal{M}}; \varepsilon_g - t_0) + (1 - q_{\varepsilon_0, t_0}) \delta_{OPT}(\boldsymbol{\mathcal{M}}; \varepsilon_g + \varepsilon_0 - t_0)\}$$

$$\delta_{OPT}(\mathcal{M}_{DP}(\varepsilon_0), \boldsymbol{\mathcal{M}}; \varepsilon_g) = q_{2\varepsilon_0, \varepsilon_0} \delta_{OPT}(\boldsymbol{\mathcal{M}}; \varepsilon_g - \varepsilon_0) + (1 - q_{2\varepsilon_0, \varepsilon_0}) \delta_{OPT}(\boldsymbol{\mathcal{M}}; \varepsilon_g + \varepsilon_0)$$

We now present the proof of Lemma 1, showing that selecting BR mechanisms at the end leads to the worst ordering of mechanisms for the overall privacy loss.

Proof [Proof of Lemma 1] Let $\boldsymbol{\mathcal{M}}$ consist of k' many ε-BR mechanisms, which are in positions $\ell_1, \cdots, \ell_{k'}$. At each level ℓ, there will be $2^{\ell-1}$ many variables to maximize an expression over, which we will write the variables as $t_{j,\ell} \in [0, \varepsilon]$ where $j \in [2^{\ell-1}]$. The full set of these variables can be written as $\mathcal{T} = \{t_{i,\ell} : i \in [2^{\ell-1}], \ell \in \{\ell_1, \cdots, \ell_{k'}\}\}$. Using the recursive formulation from Lemma 23, we know that $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}, \mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}'; \varepsilon_g)$ will consist of a summation of terms with the following form, where we take a sup over all $\mathbf{t} \in \mathcal{T}$ for some coefficient $\lambda(\mathbf{t}; \varepsilon)$ and term $\alpha(\mathbf{t}; \varepsilon, \varepsilon_g)$,

$$\lambda(\mathbf{t}; \varepsilon) \cdot \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g)).$$

We then expand this term using our recurrence formula,

$$\begin{aligned}
&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g)) \\
&= \sup_{s \in [0, \varepsilon]} \{q_{\varepsilon,s} \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s) + (1 - q_{\varepsilon,s}) \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) + \varepsilon - s)\} \\
&= \sup_{s \in [0, \varepsilon]} \{q_{\varepsilon,s} q_{2\varepsilon,\varepsilon} \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s - \varepsilon) + q_{\varepsilon,s}(1 - q_{2\varepsilon,\varepsilon}) \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s + \varepsilon) \\
&\qquad + (1 - q_{\varepsilon,s}) q_{2\varepsilon,\varepsilon} \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s) + (1 - q_{\varepsilon,s})(1 - q_{2\varepsilon,\varepsilon}) \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) + 2\varepsilon - s)\}.
\end{aligned}$$

Similarly, $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}, \mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}'; \varepsilon_g)$ will consist of a summation of terms with the following form, with the same terms $\lambda(\mathbf{t}; \varepsilon)$ and $\alpha(\mathbf{t}; \varepsilon, \varepsilon_g)$

$$\lambda(\mathbf{t}; \varepsilon) \cdot \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g)).$$

Again, we use our recurrence formula to get the following

$$\begin{aligned}
&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g)) \\
&= q_{2\varepsilon,\varepsilon} \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - \varepsilon) + (1 - q_{2\varepsilon,\varepsilon}) \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) + \varepsilon) \\
&= q_{2\varepsilon,\varepsilon} \sup_{s \in [0, \varepsilon]} \{q_{\varepsilon,s} \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s - \varepsilon) + (1 - q_{\varepsilon,s}) \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s)\} \\
&\quad + (1 - q_{2\varepsilon,\varepsilon}) \sup_{s' \in [0, \varepsilon]} \left\{q_{\varepsilon,s'} \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) + \varepsilon - s') + (1 - q_{\varepsilon,s'}) \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) + 2\varepsilon - s')\right\} \\
&= \sup_{s, s' \in [0, \varepsilon]} \left\{q_{\varepsilon,s} q_{2\varepsilon,\varepsilon} \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s - \varepsilon) + q_{\varepsilon,s'}(1 - q_{2\varepsilon,\varepsilon}) \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s' + \varepsilon)\right. \\
&\qquad \left. + (1 - q_{\varepsilon,s}) q_{2\varepsilon,\varepsilon} \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) - s) + (1 - q_{\varepsilon,s'})(1 - q_{2\varepsilon,\varepsilon}) \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g) + 2\varepsilon - s')\right\}
\end{aligned}$$

Hence, $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g))$ and $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g))$ consists of the same terms, except the former has a single sup and the latter takes a sup over two terms. Hence, we must have

$$\lambda(\mathbf{t}) \cdot \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g)) \leq \lambda(\mathbf{t}; \varepsilon) \cdot \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}; \alpha(\mathbf{t}; \varepsilon, \varepsilon_g)).$$

Because this applies for each term in both $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}, \mathcal{M}_{\texttt{BR}}, \mathcal{M}_{\texttt{DP}}, \boldsymbol{\mathcal{M}}'; \varepsilon_g)$ and when we switch the BR and DP order $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{M}}, \mathcal{M}_{\texttt{DP}}, \mathcal{M}_{\texttt{BR}}, \boldsymbol{\mathcal{M}}'; \varepsilon_g)$, we have our result. ■

Before we prove Proposition 2 we need the following two lemmas. The first provides an identity for reducing elemental terms to expressions independent of t (similar to an identity used previously in Dong et al. (2020)). The next lemma provides an expansion of $\delta_{\texttt{OPT}}$ for k pure DP mechanisms $\mathcal{M}_{\texttt{DP}}(\varepsilon)$.

Lemma 24 (Reduction Identity) *Given $\alpha \in \mathbb{R}, \varepsilon > 0$, and $t \in [0, \varepsilon]$, we have the identity:*

$$q_{t,\varepsilon}\left[1 - e^{\alpha - t}\right]_+ + (1 - q_{t,\varepsilon})\left[1 - e^{\alpha+\varepsilon-t}\right]_+ = \begin{cases} 0 & \text{if } \alpha - t \geq 0 \\ q_{t,\varepsilon}\left(1 - e^{\alpha-t}\right) & \text{if } \alpha \leq t \leq \alpha + \varepsilon \\ 1 - e^{\alpha} & \text{otherwise} \end{cases}.$$

Proof When $\alpha - t \geq 0$ we have $1 - e^{\alpha - t} < 0$ and $1 - e^{\alpha+\varepsilon-t} < 0$, hence

$$q_{t,\varepsilon}\left[1 - e^{\alpha - t}\right]_+ + (1 - q_{t,\varepsilon})\left[1 - e^{\alpha+\varepsilon-t}\right]_+ = 0.$$

When $\alpha \leq t \leq \alpha + \varepsilon$ we have $1 - e^{\alpha+\varepsilon-t} < 0$, hence

$$q_{t,\varepsilon}\left[1 - e^{\alpha - t}\right]_+ + (1 - q_{t,\varepsilon})\left[1 - e^{\alpha+\varepsilon-t}\right]_+ = q_{t,\varepsilon}\left[1 - e^{\alpha-t}\right]_+.$$

When $\alpha + \varepsilon - t \leq 0$ we have

$$\begin{aligned} q_{t,\varepsilon}\left[1 - e^{\alpha - t}\right]_+ + (1 - q_{t,\varepsilon})\left[1 - e^{\alpha+\varepsilon-t}\right]_+ &= q_{t,\varepsilon}\left(1 - e^{\alpha-t}\right) + (1 - q_{t,\varepsilon})\left(1 - e^{\alpha+\varepsilon-t}\right) \\ &= q_{t,\varepsilon} + 1 - q_{t,\varepsilon} - \left[q_{t,\varepsilon} + e^{\varepsilon}\left(1 - q_{t,\varepsilon}\right)\right] e^{\alpha-t} \\ &= 1 - \left[\frac{1 - e^{t-\varepsilon}}{1 - e^{-\varepsilon}} + e^{\varepsilon}\left(1 - \frac{1 - e^{t-\varepsilon}}{1 - e^{-\varepsilon}}\right)\right] e^{\alpha-t} \\ &= 1 - \frac{e^{t} - e^{t-\varepsilon}}{1 - e^{-\varepsilon}} e^{\alpha-t} \\ &= 1 - e^{\alpha}. \end{aligned}$$

■

Lemma 25 *For $\ell \in \mathbb{N}$, $x \in \mathbb{R}$, $\varepsilon \geq 0$, and $\varepsilon_i = \varepsilon$ for $i \in \{0, \ldots, \ell\}$ we define:*

$$\delta_\ell(x) := \delta_{OPT}\left(\mathcal{M}_{DP}(\varepsilon_1), \ldots, \mathcal{M}_{DP}(\varepsilon_\ell); x\right).$$

Then we have the following identity for some constants $\lambda_{\ell,i} \in \mathbb{R}$:

$$\delta_\ell(x) = \sum_{i \in \{0,\ldots,\ell\}} \lambda_{\ell,i}\left[1 - e^{(2i-\ell)\varepsilon + x}\right]_+.$$

Proof This is a straightforward induction on ℓ. The base case $\ell = 1$ is apparent from the recurrence in Lemma 23. If we have $\ell > 1$ and the identity holds for $\ell' < \ell$ then we have:

$$\begin{aligned}
\delta_\ell(x) = \;& q_{\varepsilon,2\varepsilon}\delta_{\ell-1}(x-\varepsilon) + (1-q_{\varepsilon,2\varepsilon})\,\delta_{\ell-1}(x+\varepsilon) \\
= \;& q_{\varepsilon,2\varepsilon} \sum_{i\in\{0,\ldots,\ell-1\}} \lambda_{\ell-1,i}\left[1-e^{(2i-\ell+1)\varepsilon+x-\varepsilon}\right]_+ \\
& + \;(1-q_{\varepsilon,2\varepsilon}) \sum_{i\in\{0,\ldots,\ell-1\}} \lambda_{\ell-1,i}\left[1-e^{(2i-\ell+1)\varepsilon+x+\varepsilon}\right]_+ \\
= \;& q_{\varepsilon,2\varepsilon}\lambda_{\ell-1,0}\left[1-e^{(-\ell\varepsilon+x)}\right]_+ \\
& + \sum_{i\in\{1,\ldots,\ell-1\}} \left(q_{\varepsilon,2\varepsilon}\lambda_{\ell-1,i} + (1-q_{\varepsilon,2\varepsilon})\,\lambda_{\ell-1,i-1}\right)\left[1-e^{(2i-\ell+1)\varepsilon+x-\varepsilon}\right]_+ \\
& + \;(1-q_{\varepsilon,2\varepsilon})\lambda_{\ell-1,\ell-1}\left[1-e^{(\ell\varepsilon+x)}\right]_+ \\
= \;& \sum_{i\in\{0,\ldots,\ell\}} \lambda_{\ell,i}\left[1-e^{(2i-\ell)\varepsilon+x}\right]_+ .
\end{aligned}$$

■

We can now complete the proof of Proposition 2.

Proof [Proof of Proposition 2] To prove this we induct on the number, N, of ε-DP mechanisms. For the base case ($N = 2$) consider the expansion of the first two terms (using Lemma 23):

$$\begin{aligned}
\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{A}}_2;\varepsilon_g) = \;& \sup_{t\in[0,\varepsilon]}\Big\{q_{t,\varepsilon}\left[q_{\varepsilon,2\varepsilon}\left[1-e^{\varepsilon_g-\varepsilon-t}\right]_+ + (1-q_{\varepsilon,2\varepsilon})\left[1-e^{\varepsilon_g+\varepsilon-t}\right]_+\right] \\
& + (1-q_{t,\varepsilon})\left[q_{\varepsilon,2\varepsilon}\left[1-e^{\varepsilon_g-t}\right]_+ + (1-q_{\varepsilon,2\varepsilon})\left[1-e^{\varepsilon_g+2\varepsilon-t}\right]_+\right]\Big\} \\
\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{B}}_2;\varepsilon_g) = \;& q_{\varepsilon,2\varepsilon}\sup_{t_1\in[0,\varepsilon]}\left\{q_{t_1,\varepsilon}\left[1-e^{\varepsilon_g-\varepsilon-t_1}\right]_+ + (1-q_{t_1,\varepsilon})\left[1-e^{\varepsilon_g-t_1}\right]_+\right\} \\
& + (1-q_{\varepsilon,2\varepsilon})\sup_{t_2\in[0,\varepsilon]}\left\{q_{t_2,\varepsilon}\left[1-e^{\varepsilon_g+\varepsilon-t_2}\right]_+ + (1-q_{t_2,\varepsilon})\left[1-e^{\varepsilon_g+2\varepsilon-t_2}\right]_+\right\}
\end{aligned}$$

Where the terms with positive coefficients that depend on ε must be 0, so the $\sup_{t_2\in[0,\varepsilon]}$ term in $\delta_{\texttt{OPT}}(\mathcal{B}_2;\varepsilon_g)$ disappears and equality is apparent.

Now, consider some $N > 2$. Suppose that for $k < N$ we have $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{A}}_k;\varepsilon_g) = \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{B}}_k;\varepsilon_g)$ for all $\boldsymbol{\mathcal{A}}_k, \boldsymbol{\mathcal{B}}_k$. Then to show $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{A}}_N;\varepsilon_g) = \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{B}}_N;\varepsilon_g)$ there are two nontrivial cases to consider:

1. $\mathcal{A}_1, \mathcal{B}_1 = \mathcal{M}_{\texttt{DP}}(\varepsilon)$. Here we simply expand the first ε-DP mechanism on each side using definitions and apply the inductive hypothesis.

2. $\mathcal{A}_1 = \mathcal{M}_{\texttt{BR}}(\varepsilon)$ and $\mathcal{B}_1 = \mathcal{M}_{\texttt{DP}}(\varepsilon)$. Here the proof is slightly more involved. First we expand the first two terms of each side:

Now consider the second case.

$$\begin{aligned}
\delta_{\texttt{OPT}}\left(\boldsymbol{\mathcal{A}}_N;\varepsilon_g\right) = \quad & \sup_{t\in[0,\varepsilon]}\Big\{q_{t,\varepsilon}\left[q_{\varepsilon,2\varepsilon}\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g-\varepsilon-t\right)\right. \\
& \qquad \left.+\left(1-q_{\varepsilon,2\varepsilon}\right)\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+\varepsilon-t\right)\right] \\
& \qquad +\left(1-q_{t,\varepsilon}\right)\left[q_{\varepsilon,2\varepsilon}\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g-t\right)\right. \\
& \qquad \left.\left.+\left(1-q_{\varepsilon,2\varepsilon}\right)\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+2\varepsilon-t\right)\right]\Big\}. \\
\delta_{\texttt{OPT}}\left(\boldsymbol{\mathcal{B}}_N;\varepsilon_g\right) = \quad & q_{\varepsilon,2\varepsilon}\sup_{t_1\in[0,\varepsilon]}\Big\{q_{t_1,\varepsilon}\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g-\varepsilon-t_1\right) \\
& \qquad \left.+\left(1-q_{t_1,\varepsilon}\right)\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g-t_1\right)\right\} \\
+ \quad & \left(1-q_{\varepsilon,2\varepsilon}\right)\sup_{t_2\in[0,\varepsilon]}\Big\{q_{t_2,\varepsilon}\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+\varepsilon-t_2\right) \\
& \qquad \left.+\left(1-q_{t_2,\varepsilon}\right)\delta_{\texttt{OPT}}\left(\mathcal{M}_{\texttt{DP}}(\varepsilon),\ldots,\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+2\varepsilon-t_2\right)\right\}.
\end{aligned}$$

Where we have applied the inductive hypothesis to $\boldsymbol{\mathcal{B}}_N$ after expanding the first term to ensure that the second mechanism is ε-BR wlog. Now, applying our formula from Lemma 25 we have:

$$\begin{aligned}
\delta_{\texttt{OPT}}\left(\boldsymbol{\mathcal{A}}_N;\varepsilon_g\right) = \quad & \sup_{t\in[0,\varepsilon]}\Bigg\{q_{t,\varepsilon}\Bigg(q_{\varepsilon,2\varepsilon}\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(-1+2i-(N-2))\varepsilon-t}\right]_+ \\
& \quad + (1-q_{\varepsilon,2\varepsilon})\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(1+2i-(N-2))\varepsilon-t}\right]_+\Bigg) \\
& \quad + (1-q_{t,\varepsilon})\Bigg(q_{\varepsilon,2\varepsilon}\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(2i-(N-2))\varepsilon-t}\right]_+ \\
& \quad + (1-q_{\varepsilon,2\varepsilon})\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(2+2i-(N-2))\varepsilon-t}\right]_+\Bigg)\Bigg\}. \\
\delta_{\texttt{OPT}}\left(\boldsymbol{\mathcal{B}}_N;\varepsilon_g\right) = \quad & q_{\varepsilon,2\varepsilon}\sup_{t_1\in[0,\varepsilon]}\Bigg\{q_{t_1,\varepsilon}\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(-1+2i-(N-2))\varepsilon-t_1}\right]_+ \\
& \quad + (1-q_{t_1,\varepsilon})\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(2i-(N-2))\varepsilon-t_1}\right]_+\Bigg\} \\
& + (1-q_{\varepsilon,2\varepsilon})\sup_{t_2\in[0,\varepsilon]}\Bigg\{q_{t_2,\varepsilon}\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(1+2i-(N-2))\varepsilon-t_2}\right]_+ \\
& \quad + (1-q_{t_2,\varepsilon})\sum_{i\in[\ell]}\lambda_i\left[1-e^{\varepsilon_g+(2+2i-(N-2))\varepsilon-t_2}\right]_+\Bigg\}.
\end{aligned}$$

Comparing terms with like coefficients we see that the reduction identity applies and so each pair of sums can be reduced. The reduction identity yields an expression that depends on t only if $t \in (\alpha, \alpha+\varepsilon)$, therefore only one reduced term from each sum can depend on t as the $\alpha = \varepsilon_g + \ell\varepsilon - t$ terms increase in increments of size 2ε. Furthermore, because the coefficients of ε in the first exponent in each pair of sums (to which we are applying the reduction) have the same parity, the same term i^* depends on t in both. Thus, collecting the terms constant in t into constants $C_1(\varepsilon), C_2(\varepsilon) \in \mathbb{R}$, we can write:

$$\begin{aligned}
\delta_{\texttt{OPT}}\left(\boldsymbol{\mathcal{A}}_N;\varepsilon_g\right) = \quad & \sup_{t\in[0,\varepsilon]}\Bigg\{q_{\varepsilon,2\varepsilon}\left(C_1(\varepsilon)+q_{t,\varepsilon}\lambda_{i^*}\left[1-e^{\varepsilon_g+(-1+2i^*-(N-2))\varepsilon-t}\right]_+\right) \\
& \quad + (1-q_{\varepsilon,2\varepsilon})\left(C_2(\varepsilon)+q_{t,\varepsilon}\lambda_{i^*}\left[1-e^{\varepsilon_g+(1+2i^*-(N-2))\varepsilon-t}\right]_+\right)\Bigg\}. \quad (9) \\
\delta_{\texttt{OPT}}\left(\boldsymbol{\mathcal{B}}_N;\varepsilon_g\right) = \quad & q_{\varepsilon,2\varepsilon}\sup_{t_1\in[0,\varepsilon]}\left\{C_1(\varepsilon)+q_{t,\varepsilon}\lambda_{i^*}\left[1-e^{\varepsilon_g+(-1+2i^*-(N-2))\varepsilon-t_1}\right]_+\right\} \\
& \quad + (1-q_{\varepsilon,2\varepsilon})\sup_{t_2\in[0,\varepsilon]}\left\{C_2(\varepsilon)+q_{t_2,\varepsilon}\lambda_{i^*}\left[1-e^{\varepsilon_g+(1+2i^*-(N-2))\varepsilon-t_2}\right]_+\right\}. \quad (10)
\end{aligned}$$

Then both remaining terms in $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{B}}_N;\varepsilon_g)$ achieve the supremum for the same argument $t^* = t_1 = t_2$. Fixing $t = t^*$ in (9) and $t_1 = t_2 = t^*$ in (10) and comparing it is clear that $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{A}}_N;\varepsilon_g) \geq \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{B}}_N;\varepsilon_g)$. By the triangle inequality, we also have $\delta_{\texttt{OPT}}(\boldsymbol{\mathcal{B}}_N;\varepsilon_g) \geq \delta_{\texttt{OPT}}(\boldsymbol{\mathcal{A}}_N;\varepsilon_g)$, which proves the result. ■

To see that the privacy loss can strictly increase if we change the ordering of BR and DP mechanisms, we consider the following simple example where we compose $k = 3$ ε-DP mechanisms adaptively and 2 of them are ε-BR. In this case, we can directly compute $\delta_{\texttt{OPT}}(\cdot;\varepsilon_g)$ for each of the three possible orderings of mechanisms.

Lemma 26 *Let $\varepsilon > 0$ and $\varepsilon_g \geq 0$. If $\varepsilon_g \geq \varepsilon$, then we have*

$$\begin{aligned}\delta_{OPT}(\mathcal{M}_{BR},\mathcal{M}_{BR},\mathcal{M}_{DP};\varepsilon_g) &= \delta_{OPT}(\mathcal{M}_{BR},\mathcal{M}_{DP},\mathcal{M}_{BR};\varepsilon_g)\\ &= \delta_{OPT}(\mathcal{M}_{DP},\mathcal{M}_{BR},\mathcal{M}_{BR};\varepsilon_g).\end{aligned}$$

For $\varepsilon_g \leq \varepsilon$ we define the following functions:

$$\begin{aligned}x(t) &:= q_{2\varepsilon,\varepsilon}\left(q_{\varepsilon,t}q^2_{\varepsilon,\frac{\varepsilon_g-t}{2}}\left(1-e^{-\varepsilon}\right) + (1-q_{\varepsilon,t})q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t}{2}}\left(1-e^{-\varepsilon}\right)\right)\\ y(t) &:= q_{2\varepsilon,\varepsilon}\left(q_{\varepsilon,t}(1-e^{\varepsilon_g-\varepsilon-t}) + (1-q_{\varepsilon,t})q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t}{2}}\left(1-e^{-\varepsilon}\right)\right)\\ z(t) &:= (1-q_{2\varepsilon,\varepsilon})q_{\varepsilon,t}q^2_{\varepsilon,\varepsilon+\frac{\varepsilon_g-t}{2}}\left(1-e^{-\varepsilon}\right).\end{aligned}$$

We then can write out the following expressions:

$$\begin{aligned}\delta_{OPT}(\mathcal{M}_{DP},\mathcal{M}_{BR},\mathcal{M}_{BR};\varepsilon_g) &= \max\left\{\sup_{t\in[0,\varepsilon_g)} x(t), \sup_{t\in[\varepsilon_g,\varepsilon]} y(t)\right\} + \sup_{t'\in[\varepsilon_g,\varepsilon]} z(t')\\ \delta_{OPT}(\mathcal{M}_{BR},\mathcal{M}_{DP},\mathcal{M}_{BR};\varepsilon_g) &= \max\left\{\sup_{t\in[0,\varepsilon_g)} x(t), \sup_{t\in[\varepsilon_g,\varepsilon]} y(t)+z(t)\right\}\\ \delta_{OPT}(\mathcal{M}_{BR},\mathcal{M}_{BR},\mathcal{M}_{DP};\varepsilon_g) &= \max\left\{\sup_{t\in[0,\varepsilon_g)} x(t), \sup_{t\in[\varepsilon_g,\varepsilon]} y(t)+z(t)\right\}.\end{aligned}$$

Proof We start with the ordering with the DP mechanism at the end.

$$\begin{aligned}&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon),\mathcal{M}_{\texttt{BR}}(\varepsilon),\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g)\\ &= \sup_{t_{1,1}\in[0,\varepsilon]}\left\{q_{\varepsilon,t_{1,1}}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon),\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g-t_{1,1}) + (1-q_{\varepsilon,t_{1,1}})\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon),\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+\varepsilon-t_{1,1})\right\}\\ &= \sup_{t_{1,1}}\left\{q_{\varepsilon,t_{1,1}}\sup_{t_{1,2}}\left\{q_{\varepsilon,t_{1,2}}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g-t_{1,1}-t_{1,2}) + (1-q_{\varepsilon,t_{1,2}})\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+\varepsilon-t_{1,1}-t_{1,2})\right\}\right.\\ &\left.+(1-q_{\varepsilon,t_{1,1}})\sup_{t_{2,2}}\left\{q_{\varepsilon,t_{2,2}}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+\varepsilon-t_{1,1}-t_{2,2}) + (1-q_{\varepsilon,t_{2,2}})\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon);\varepsilon_g+2\varepsilon-t_{1,1}-t_{2,2})\right\}\right\}\\ &= \sup_{t_{1,1},t_{1,2},t_{2,2}}\left\{q_{\varepsilon,t_{1,1}}q_{\varepsilon,t_{1,2}}q_{2\varepsilon,\varepsilon}[1-e^{\varepsilon_g-\varepsilon-t_{1,1}-t_{1,2}}]_+ + q_{\varepsilon,t_{1,1}}q_{\varepsilon,t_{1,2}}(1-q_{2\varepsilon,\varepsilon})[1-e^{\varepsilon_g+\varepsilon-t_{1,1}-t_{1,2}}]_+\right.\\ &\quad +q_{\varepsilon,t_{1,1}}(1-q_{\varepsilon,t_{1,2}})q_{2\varepsilon,\varepsilon}[1-e^{\varepsilon_g-t_{1,1}-t_{1,2}}]_+ + (1-q_{\varepsilon,t_{1,1}})q_{\varepsilon,t_{2,2}}q_{2\varepsilon,\varepsilon}[1-e^{\varepsilon_g-t_{1,1}-t_{2,2}}]_+\\ &\quad \left.+(1-q_{\varepsilon,t_{1,1}})(1-q_{\varepsilon,t_{2,2}})q_{2\varepsilon,\varepsilon}[1-e^{\varepsilon_g+\varepsilon-t_{1,1}-t_{2,2}}]_+\right\}\end{aligned}$$

$$
\begin{aligned}
&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g) \\
&= q_{2\varepsilon,\varepsilon}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - \varepsilon) + (1 - q_{2\varepsilon,\varepsilon})\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g + \varepsilon) \\
&= q_{2\varepsilon,\varepsilon} \sup_{t_{1,2}\in[0,\varepsilon]} \left\{q_{\varepsilon,t_{1,2}}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - \varepsilon - t_{1,2}) + (1 - q_{\varepsilon,t_{1,2}})\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - t_{1,1})\right\} \\
&\quad + (1 - q_{2\varepsilon,\varepsilon}) \sup_{t_{2,2}\in[0,\varepsilon]} \left\{ q_{\varepsilon,t_{2,2}}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g + \varepsilon - t_{2,2}) + (1 - q_{\varepsilon,t_{2,2}}) \underbrace{\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g + 2\varepsilon - t_{2,2})}_{=0} \right\} \\
&= q_{2\varepsilon,\varepsilon} \sup_{t_{1,2}} \left\{ q_{\varepsilon,t_{1,2}} \sup_{t_{1,3}} \left\{q_{\varepsilon,t_{1,3}}[1 - e^{\varepsilon_g - \varepsilon - t_{1,2} - t_{1,3}}]_+ + (1 - q_{\varepsilon,t_{1,3}})[1 - e^{\varepsilon_g - t_{1,2} - t_{1,3}}]_+\right\} \right. \\
&\qquad \left. + (1 - q_{\varepsilon,t_{1,2}}) \sup_{t_{2,3}} \left\{q_{\varepsilon,t_{2,3}}[1 - e^{\varepsilon_g - t_{1,2} - t_{2,3}}]_+ + (1 - q_{\varepsilon,t_{2,3}})[1 - e^{\varepsilon_g + \varepsilon - t_{1,2} - t_{2,3}}]_+\right\} \right\} \\
&\quad + (1 - q_{2\varepsilon,\varepsilon}) \sup_{t_{2,2}} \left\{ q_{\varepsilon,t_{2,2}} \sup_{t_{3,3}} \left\{ q_{\varepsilon,t_{3,3}}[1 - e^{\varepsilon_g + \varepsilon - t_{2,2} - t_{3,3}}]_+ + (1 - q_{\varepsilon,t_{3,3}}) \underbrace{[1 - e^{\varepsilon_g + 2\varepsilon - t_{2,2} - t_{3,3}}]_+}_{=0} \right\} \right\} \\
&= \sup_{t_{1,2},t_{2,2},t_{1,3},t_{2,3},t_{3,3}} \left\{q_{\varepsilon,t_{1,2}}q_{\varepsilon,t_{1,3}}q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g - \varepsilon - t_{1,2} - t_{1,3}}]_+ + q_{\varepsilon,t_{2,2}}q_{\varepsilon,t_{3,3}}(1 - q_{2\varepsilon,\varepsilon})[1 - e^{\varepsilon_g + \varepsilon - t_{2,2} - t_{3,3}}]_+ \right. \\
&\qquad + q_{\varepsilon,t_{1,2}}(1 - q_{\varepsilon,t_{1,3}})q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g - t_{1,2} - t_{1,3}}]_+ + (1 - q_{\varepsilon,t_{1,2}})q_{\varepsilon,t_{2,3}}q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g - t_{1,2} - t_{2,3}}]_+ \\
&\qquad \left. + (1 - q_{\varepsilon,t_{1,2}})(1 - q_{\varepsilon,t_{2,3}})q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g + \varepsilon - t_{1,2} - t_{2,3}}]_+\right\}
\end{aligned}
$$

We next consider the ordering that alternates between BR and DP, as one would do with using the exponential mechanism to discover the k most frequent elements and then adding Laplace noise to the counts of the elements that were discovered.

$$
\begin{aligned}
&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g) \\
&= \sup_{t_{1,1}\in[0,\varepsilon]} \left\{q_{\varepsilon,t_{1,1}}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - t_{1,1}) + (1 - q_{\varepsilon,t_{1,1}})\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g + \varepsilon - t_{1,1})\right\} \\
&= \sup_{t_{1,1}\in[0,\varepsilon]} \left\{q_{\varepsilon,t_{1,1}}(q_{2\varepsilon,\varepsilon}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - t_{1,1} - \varepsilon) + (1 - q_{2\varepsilon,\varepsilon})\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - t_{1,1} + \varepsilon)) \right. \\
&\qquad \left. + (1 - q_{\varepsilon,t_{1,1}}) \left(q_{2\varepsilon,\varepsilon}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - t_{1,1}) + (1 - q_{2\varepsilon,\varepsilon}) \underbrace{\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g - t_{1,1} + 2\varepsilon)}_{=0} \right) \right\} \\
&= \sup_{t_{1,1},t_{1,3},t_{2,3},t_{3,3}} \left\{q_{\varepsilon,t_{1,1}}q_{\varepsilon,t_{1,3}}q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g - t_{1,1} - \varepsilon - t_{1,3}}]_+ + q_{\varepsilon,t_{1,1}}q_{\varepsilon,t_{2,3}}(1 - q_{2\varepsilon,\varepsilon})[1 - e^{\varepsilon_g - t_{1,1} + \varepsilon - t_{2,3}}]_+ \right. \\
&\qquad + q_{\varepsilon,t_{1,1}}(1 - q_{\varepsilon,t_{1,3}})q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g - t_{1,1} - t_{1,3}}]_+ + (1 - q_{\varepsilon,t_{1,1}})q_{\varepsilon,t_{3,3}}q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g - t_{1,1} - t_{3,3}}]_+ \\
&\qquad \left. + (1 - q_{\varepsilon,t_{1,1}})(1 - q_{\varepsilon,t_{3,3}})q_{2\varepsilon,\varepsilon}[1 - e^{\varepsilon_g + \varepsilon - t_{1,1} - t_{3,3}}]_+\right\}
\end{aligned}
$$

In the case when $\varepsilon_g \geq \varepsilon$, all the terms with $[1 - e^{\varepsilon_g + \varepsilon - t - t'}]_+ = 0$, so all three become equal, with a sup over three terms. Hence, in the case when $\varepsilon_g \geq \varepsilon$ we have

$$\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g) = \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g)$$

We then assume for the rest of the proof that $\varepsilon_g < \varepsilon$. There are many similar terms in each of the expressions. We start by focusing on the following term when $t < \varepsilon_g$,

$$\sup_{t' \in [0,\varepsilon]} \left\{ q_{\varepsilon,t'}[1 - e^{\varepsilon_g - \varepsilon - t - t'}]_+ + (1 - q_{\varepsilon,t'})[1 - e^{\varepsilon_g - t - t'}]_+ \right\}$$
$$= \max \left\{ q^2_{\varepsilon, \frac{\varepsilon_g - t}{2}} \left(1 - e^{-\varepsilon}\right), 1 - e^{\varepsilon_g - \varepsilon - t} \right\}.$$

To determine which term attains the maximum, we consider each term,

$$q^2_{\varepsilon, \frac{\varepsilon_g - t}{2}} \left(1 - e^{-\varepsilon}\right) = \frac{\left(1 - e^{\frac{\varepsilon_g - t}{2} - \varepsilon}\right)^2}{1 - e^{-\varepsilon}} = \frac{1 - 2e^{\frac{\varepsilon_g - t}{2} - \varepsilon} + e^{\varepsilon_g - t - 2\varepsilon}}{1 - e^{-\varepsilon}}$$
$$1 - e^{\varepsilon_g - \varepsilon - t} = \frac{1 - e^{-\varepsilon} - e^{\varepsilon_g - t - \varepsilon} + e^{\varepsilon_g - t - 2\varepsilon}}{1 - e^{-\varepsilon}}$$

Hence, for $t < \varepsilon_g$, we have

$$\max \left\{ q^2_{\varepsilon, \frac{\varepsilon_g - t}{2}} \left(1 - e^{-\varepsilon}\right), 1 - e^{\varepsilon_g - \varepsilon - t} \right\} = q^2_{\varepsilon, \frac{\varepsilon_g - t}{2}} \left(1 - e^{-\varepsilon}\right)$$
$$\iff 0 \leq e^{\varepsilon_g - t} - 2e^{\frac{\varepsilon_g - t}{2}} + 1 = \left(1 - e^{\frac{\varepsilon_g - t}{2}}\right)^2.$$

In the case where $t \in [\varepsilon_g, \varepsilon]$, we have

$$\sup_{t' \in [0,\varepsilon]} \left\{ q_{\varepsilon,t'}[1 - e^{\varepsilon_g - \varepsilon - t - t'}]_+ + (1 - q_{\varepsilon,t'})[1 - e^{\varepsilon_g - t - t'}]_+ \right\} = 1 - e^{\varepsilon_g - \varepsilon - t}.$$

We next focus on another common term in each expression, considering first when $t < \varepsilon_g$.

$$\sup_{t' \in [0,\varepsilon]} \left\{ q_{\varepsilon,t'}[1 - e^{\varepsilon_g - t - t'}]_+ + (1 - q_{\varepsilon,t'})[1 - e^{\varepsilon_g + \varepsilon - t - t'}]_+ \right\} = q^2_{\varepsilon, \frac{\varepsilon_g + \varepsilon - t}{2}} \left(1 - e^{-\varepsilon}\right)$$

On the other hand, if $t \in [\varepsilon_g, \varepsilon]$, we have

$$\sup_{t' \in [0,\varepsilon]} \left\{ q_{\varepsilon,t'}[1 - e^{\varepsilon_g - t - t'}]_+ + (1 - q_{\varepsilon,t'})[1 - e^{\varepsilon_g + \varepsilon - t - t'}]_+ \right\}$$
$$= \max \left\{ q^2_{\varepsilon, \frac{\varepsilon_g + \varepsilon - t}{2}} \left(1 - e^{-\varepsilon}\right), 1 - e^{\varepsilon_g - t} \right\}$$

Once again, to determine which term attains the maximum, we consider each term,

$$q^2_{\varepsilon, \frac{\varepsilon_g + \varepsilon - t}{2}} \left(1 - e^{-\varepsilon}\right) = \frac{\left(1 - e^{\frac{\varepsilon_g - \varepsilon - t}{2}}\right)^2}{1 - e^{-\varepsilon}} = \frac{1 - 2e^{\frac{\varepsilon_g - \varepsilon - t}{2}} + e^{\varepsilon_g - \varepsilon - t}}{1 - e^{-\varepsilon}}$$
$$1 - e^{\varepsilon_g - t} = \frac{1 - e^{-\varepsilon} - e^{\varepsilon_g - t} + e^{\varepsilon_g - \varepsilon - t}}{1 - e^{-\varepsilon}}$$

Hence, for $t \in [\varepsilon_g, \varepsilon]$, we have

$$\max\left\{q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t}{2}}\left(1-e^{-\varepsilon}\right), 1-e^{\varepsilon_g-t}\right\} = q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t}{2}}\left(1-e^{-\varepsilon}\right)$$
$$\iff 0 \le e^{\varepsilon_g-\varepsilon-t} - 2e^{\frac{\varepsilon_g-\varepsilon-t}{2}} + 1 = \left(1-e^{\frac{\varepsilon_g-\varepsilon-t}{2}}\right)^2$$

Putting this together, we have

$$\begin{aligned}
&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g)\\
&= q_{2\varepsilon,\varepsilon}\max\left\{\sup_{t_{1,2}\in[0,\varepsilon_g)}\left\{q_{\varepsilon,t_{1,2}}q^2_{\varepsilon,\frac{\varepsilon_g-t_{1,2}}{2}}\left(1-e^{-\varepsilon}\right) + (1-q_{\varepsilon,t_{1,2}})q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t_{1,2}}{2}}\left(1-e^{-\varepsilon}\right)\right\},\right.\\
&\qquad\left.\sup_{t_{1,2}\in[\varepsilon_g,\varepsilon]}\left\{q_{\varepsilon,t_{1,2}}(1-e^{\varepsilon_g-\varepsilon-t_{1,2}}) + (1-q_{\varepsilon,t_{1,2}})q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t_{1,2}}{2}}\left(1-e^{-\varepsilon}\right)\right\}\right\}\\
&\qquad + (1-q_{2\varepsilon,\varepsilon})\sup_{t_{2,2}\in[\varepsilon_g,\varepsilon]}\left\{q_{\varepsilon,t_{2,2}}q^2_{\varepsilon,\varepsilon+\frac{\varepsilon_g-t_{2,2}}{2}}\left(1-e^{-\varepsilon}\right)\right\}\\
&= \max\left\{\sup_{t_{1,2}\in[0,\varepsilon_g)} x(t_{1.2}), \sup_{t_{1,2}\in[\varepsilon_g,\varepsilon]} y(t_{1,2})\right\} + \sup_{t_{2,2}\in[\varepsilon_g,\varepsilon]} z(t_{2,2})
\end{aligned}$$

We now focus on $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g)$,

$$\begin{aligned}
&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g)\\
&= \max\left\{q_{2\varepsilon,\varepsilon}\sup_{t_{1,1}\in[0,\varepsilon_g)}\left\{q_{\varepsilon,t_{1,1}}q^2_{\varepsilon,\frac{\varepsilon_g-t_{1,1}}{2}}\left(1-e^{-\varepsilon}\right) + (1-q_{\varepsilon,t_{1,1}})q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t_{1,1}}{2}}\left(1-e^{-\varepsilon}\right)\right\},\right.\\
&\qquad q_{2\varepsilon,\varepsilon}\sup_{t_{1,1}\in[\varepsilon_g,\varepsilon]}\left\{q_{\varepsilon,t_{1,1}}(1-e^{\varepsilon_g-\varepsilon-t_{1,1}}) + (1-q_{\varepsilon,t_{1,1}})q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t_{1,1}}{2}}\left(1-e^{-\varepsilon}\right)\right.\\
&\qquad\left.\left.+(1-q_{2\varepsilon,\varepsilon})q_{\varepsilon,t_{1,1}}q^2_{\varepsilon,\varepsilon+\frac{\varepsilon_g-t_{1,1}}{2}}\left(1-e^{-\varepsilon}\right)\right\}\right\}\\
&= \max\left\{\sup_{t_{1,1}\in[0,\varepsilon_g)} x(t_{1,1}), \sup_{t_{1,1}\in[\varepsilon_g,\varepsilon]} y(t_{1,1}) + z(t_{1,1})\right\}
\end{aligned}$$

Lastly, we focus on $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon); \varepsilon_g)$. Note that when $t_{1,1} < \varepsilon_g$, we always have $\varepsilon_g + \varepsilon - t_{1,1} - t_{1,2} \ge 0$ for all $t_{1,2} \in [0, \varepsilon]$, and for $t_{1,1} \in [\varepsilon_g, \varepsilon]$ we have $\varepsilon_g - t_{1,1} - t_{1,2} < 0$ for any $t_{1,2} > 0$.

$$\begin{aligned}
&\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g)\\
&= \sup_{t_{1,1},t_{1,2}}\left\{q_{\varepsilon,t_{1,1}}q_{\varepsilon,t_{1,2}}q_{2\varepsilon,\varepsilon}[1-e^{\varepsilon_g-\varepsilon-t_{1,1}-t_{1,2}}]_+ + q_{\varepsilon,t_{1,1}}q_{\varepsilon,t_{1,2}}(1-q_{2\varepsilon,\varepsilon})[1-e^{\varepsilon_g+\varepsilon-t_{1,1}-t_{1,2}}]_+\right.\\
&\qquad\left.+q_{\varepsilon,t_{1,1}}(1-q_{\varepsilon,t_{1,2}})q_{2\varepsilon,\varepsilon}[1-e^{\varepsilon_g-t_{1,1}-t_{1,2}}]_+ + (1-q_{\varepsilon,t_{1,1}})q_{2\varepsilon,\varepsilon}q^2_{\varepsilon,\frac{\varepsilon_g+\varepsilon-t_{1,1}}{2}}(1-e^{-\varepsilon)}\right\}\\
&= \max\left\{\sup_{t_{1,1}\in[0,\varepsilon_g)} x(t_{1,1}), \sup_{t_{1,1}\in[\varepsilon_g,\varepsilon]} y(t_{1,1}) + z(t_{1,1})\right\}
\end{aligned}$$

Summarizing this, we then have for the nonnegative functions $x(t)$, $y(t)$, and $z(t)$ defined above,

$$\begin{aligned}\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g) &= \max\left\{\sup_{t\in[0,\varepsilon_g)} x(t), \sup_{t\in[\varepsilon_g,\varepsilon]} y(t)\right\} + \sup_{t'\in[\varepsilon_g,\varepsilon]} z(t') \\ \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g) &= \max\left\{\sup_{t\in[0,\varepsilon_g)} x(t), \sup_{t\in[\varepsilon_g,\varepsilon]} y(t) + z(t)\right\} \\ &= \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon); \varepsilon_g)\end{aligned}$$

■

From this, we can now proof Lemma 13, which we restate here and explicitly give the optimal privacy bounds for the different orderings.

Lemma 27 *Let $0 \le \varepsilon_g < \varepsilon$. Then*

$$\delta_{OPT}(\mathcal{M}_{DP}, \mathcal{M}_{BR}, \mathcal{M}_{BR}; \varepsilon_g) > \delta_{OPT}(\mathcal{M}_{BR}, \mathcal{M}_{DP}, \mathcal{M}_{BR}; \varepsilon_g).$$

Furthermore, for $x(t), y(t), z(t)$ defined in Lemma 26, we have

$$\delta_{OPT}(\mathcal{M}_{DP}, \mathcal{M}_{BR}, \mathcal{M}_{BR}; \varepsilon_g) = \begin{cases} x(\varepsilon/2) + z\left(\frac{2\varepsilon+\varepsilon_g}{3}\right) & \text{if } \varepsilon_g \ge \varepsilon/2 \\ y\left(\frac{\varepsilon+\varepsilon_g}{3}\right) + z\left(\frac{2\varepsilon+\varepsilon_g}{3}\right) & \text{else} \end{cases}$$

$$\delta_{OPT}(\mathcal{M}_{BR}, \mathcal{M}_{DP}, \mathcal{M}_{BR}; \varepsilon_g) = \begin{cases} \max\{x(\varepsilon/2), y(\varepsilon_g) + z(\varepsilon_g)\} & \text{if } \varepsilon_g \ge \varepsilon/2 \\ \max\{x(\varepsilon_g), y(\varepsilon/2) + z(\varepsilon/2)\} & \text{else} \end{cases}$$

Proof We start by computing the derivatives of the necessary functions,

$$\begin{aligned} x'(t) &= \tfrac{1-e^{\varepsilon/2}}{(1-e^{-\varepsilon})^2}\left(e^t - e^{\varepsilon/2}\right)e^{\frac{\varepsilon_g - t}{2} - 2\varepsilon}. \\ y'(t) &= e^{-2\varepsilon - t}\left(e^{t/2} - e^{\frac{\varepsilon_g+\varepsilon}{2}}\right)\left(e^{3t/2} - e^{\frac{\varepsilon_g+\varepsilon}{2}}\right) \\ z'(t) &= -e^{-\varepsilon-t}(e^{t/2} - e^{\varepsilon_g/2})\left(e^{3t/2} - e^{\varepsilon+\varepsilon_g/2}\right) \\ (y+z)'(t) &= (e^{\varepsilon/2}-1)^2(e^{\varepsilon/2}+1)(e^{\varepsilon/2} - e^t)e^{\frac{\varepsilon_g - 6\varepsilon - t}{2}}. \end{aligned}$$

Note that for any $t \in [0, \varepsilon]$, we have

$$\begin{aligned} & q^2_{\varepsilon, \frac{\varepsilon_g - t}{2}}\left(1 - e^{-\varepsilon}\right) > (1 - e^{\varepsilon_g - \varepsilon - t}) \\ \iff & 0 < \left(1 - e^{\frac{\varepsilon_g - t}{2}}\right)^2 \end{aligned}$$

Hence, as long as $t \neq \varepsilon_g$, we have $x(t) > y(t)$. We break up the analysis into two cases, depending on whether ε_g is larger or smaller than $\varepsilon/2$.

Case 1: First assume that $\varepsilon/2 < \varepsilon_g$. In this case, we have,

$$\sup_{t\in[\varepsilon_g,\varepsilon]} y(t) < x(\varepsilon/2) = \sup_{t\in[0,\varepsilon_g)} x(t).$$

Furthermore, we have for any $t \in [\varepsilon_g, \varepsilon]$

$$y(t) + z(t) < x(\varepsilon/2) + z(t) < x(\varepsilon/2) + z(t) \leq x(\varepsilon/2) + \sup_{t'\in[\varepsilon_g,\varepsilon]} z(t')$$

Case 2: When $\varepsilon/2 = \varepsilon_g$, we have for any $t \in (\varepsilon_g, \varepsilon]$

$$y(t) + z(t) < x(t) + z(t) < x(\varepsilon_g) + z(t) \leq \sup_{t'\in[0,\varepsilon_g)} x(t') + \sup_{t'\in[\varepsilon_g,\varepsilon]} z(t')$$

$$\text{and } y(\varepsilon_g) + \underbrace{z(\varepsilon_g)}_{=0} = x(\varepsilon_g) < x(\varepsilon_g) + \sup_{t'\in[\varepsilon_g,\varepsilon]} z(t') = \sup_{t'\in[0,\varepsilon_g)} x(t') + \sup_{t'\in[\varepsilon_g,\varepsilon]} z(t')$$

Note that we are taking sup over compact sets, including the element ε_g. Thus, we have, in the case of $\varepsilon/2 \leq \varepsilon_g$ that $\delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g) < \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon), \mathcal{M}_{\texttt{BR}}(\varepsilon); \varepsilon_g)$.

Case 3: Now we consider $\varepsilon/2 > \varepsilon_g$. In this case,

$$\sup_{t\in[0,\varepsilon_g)} x(t) = x(\varepsilon_g) = y(\varepsilon_g) < y\left(\tfrac{\varepsilon_g+\varepsilon}{3}\right) = \sup_{t'\in[\varepsilon_g,\varepsilon]} y(t').$$

Looking at the derivatives of $z(t)$ and $y(t)$, computed above, the term that maximizes $y(t)$ is not the same as $z(t)$, so that

$$\sup_{t\in[\varepsilon_g,\varepsilon]} y(t) + z(t) < \sup_{t\in[\varepsilon_g,\varepsilon]} y(t) + \sup_{t'\in[\varepsilon_g,\varepsilon]} z(t')$$

This concludes the proof. ■

Appendix D. Additional Details for Section 6

In this section, we compare the Laplace mechanism and the Gaussian mechanism over multiple rounds of composition. We now discuss how we can actually bound the privacy loss of the Laplace mechanism by considering composition of Δ (ℓ_0-sensitivity) many DP mechanisms.

Lemma 28 *Consider a function $f : \mathcal{X} \to \mathbb{R}^d$ with ℓ_0-sensitivity Δ and ℓ_∞-sensitivity τ. Then the Laplace mechanism with parameter $\varepsilon > 0$*

$$f(x) + (Z_1, \cdots, Z_d), \qquad \{Z_i : i \in [d]\} \overset{i.i.d.}{\sim} \texttt{Lap}(\tau/\varepsilon).$$

is $(\varepsilon_g, \delta_{\texttt{OPT}}(\mathcal{M}_{\texttt{DP}}(\varepsilon_1), \cdots, \mathcal{M}_{\texttt{DP}}(\varepsilon_\Delta); \varepsilon_g))$-DP for $\varepsilon_g \geq 0$ and $\varepsilon_i = \varepsilon$.

Proof Note that we fix neighboring datasets, x, x' which induces two function values $f(x)$, $f(x')$ that differ in at most Δ positions, and in each position they differ by at most τ. Since we fix the neighboring datasets, we also know the Δ positions that have changed. Hence, we need to only consider the contributions to the overall privacy loss in these Δ positions, while the other positions contribute zero to the overall privacy loss and can be dropped. ■

Note that a similar argument can be made with the Gaussian mechanisms, given the ℓ_0-sensitivity. For this, we will analyze the mechanism using zCDP from Definition 5.

Using a similar argument to Lemma 28 and using the composition property of zCDP from Bun and Steinke (2016), we have the following result, which can be optimized over $\alpha > 1$.

Lemma 29 *Consider a function $f : \mathcal{X} \to \mathbb{R}^d$ with ℓ_0-sensitivity Δ and ℓ_∞-sensitivity τ. Then the Gaussian mechanism*

$$f(x) + (Z_1, \cdots, Z_d), \qquad \{Z_i : i \in [d]\} \overset{i.i.d.}{\sim} N\left(0, \tau^2\sigma^2\right).$$

is $(\frac{\Delta}{2\sigma^2} + \frac{1}{\sigma}\sqrt{2\Delta \ln(1/\delta)}, \delta)$-DP for any $\delta > 0$.

We can also translate our ℓ_0-sensitivity bound Δ and ℓ_∞-sensitivity bound τ into an ℓ_2-sensitivity bound and use the best scale of noise for the Gaussian mechanism. The following result from Balle and Wang (2018) gives this this bound.

Lemma 30 (Analytic Gauss Balle and Wang (2018)) *Let $f : \mathcal{X} \to \mathbb{R}^d$ have ℓ_2-sensitivity Δ_2, then for any $\varepsilon > 0$ and $\delta \in (0, 1]$ we have $M(x) = N\left(f(x), \Delta_2^2\sigma^2 I_d\right)$ is (ε, δ)-DP if and only if*

$$\Phi\left(\frac{1}{2\sigma} - \varepsilon\sigma\right) - e^{\varepsilon}\Phi\left(-\frac{1}{2\sigma} - \varepsilon\sigma\right) \leq \delta$$

D.1. Comparison Results Between Laplace and Gaussian Mechanisms

Given an ℓ_0-sensitivity bound, we then want to compare the various bounds of the Gaussian mechanism from Lemma 30 and Lemma 29 with our Laplace mechanism bound from Lemma 28. In order to compare the utility of the Laplace and Gaussian mechanisms, we will fix the variances to be the same between them and set $\delta > 0$ to be the same across.[2] We present the comparisons in Figure 3. Note that for relatively small ℓ_0-sensitivities, we get an improvement in the overall privacy guarantee with the Laplace mechanism, but then Gaussian noise seems to win out as the ℓ_0-sensitivity increases.

We are not just concerned with the overall DP parameters for a one-shot mechanism; we also want to consider composition. In order to apply composition for Lemma 30, we will use the optimal DP composition bound from Murtagh and Vadhan (2016) over k different mechanisms with ℓ_0-sensitivity Δ and find the smallest ε_g value for the given σ and overall δ. We present the results in Figure 4.

Note the bounds for "Laplace Noise" and "Gaussian Noise with zCDP" do not change between Figures 3 and 4, since the bounds consider composing $k \cdot \Delta$-mechanisms which varies from 20 to 100 in Figure 4. We see that the best one-shot mechanism then does not provide a small privacy loss when composing several mechanisms. From the empirical results, we see that the Gaussian mechanisms typically have smaller accumulated privacy loss after a large enough k, however for a reasonable number of compositions, Laplace outperforms Gaussian.

2. Note that the variance of $\texttt{Lap}(b)$ is $2b^2$.

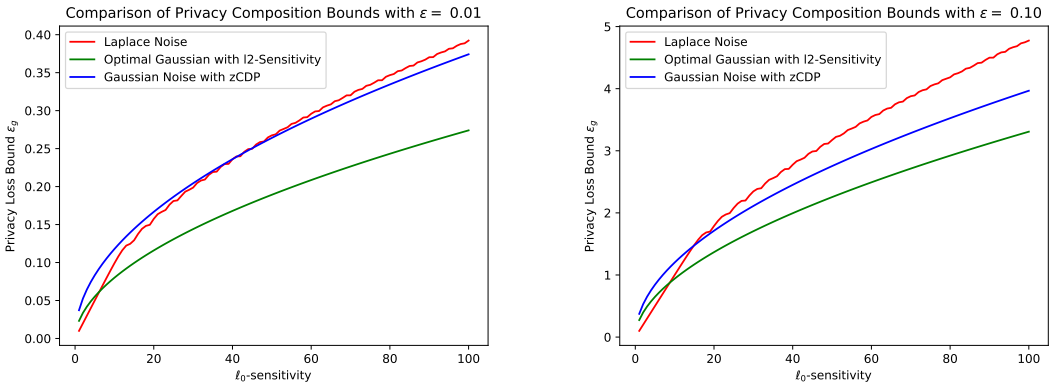

Figure 3: Comparison of the overall DP guarantee with $\delta = 10^{-6}$, with "Laplace Noise" being the bound in Lemma 28, "Optimal Gaussian with ℓ_2-sensitivity" being the bound in Lemma 30, and "Gaussian Noise with zCDP" being the bound in Lemma 29. Note that we equalize the standard deviations between the Laplace and Gaussian noise.

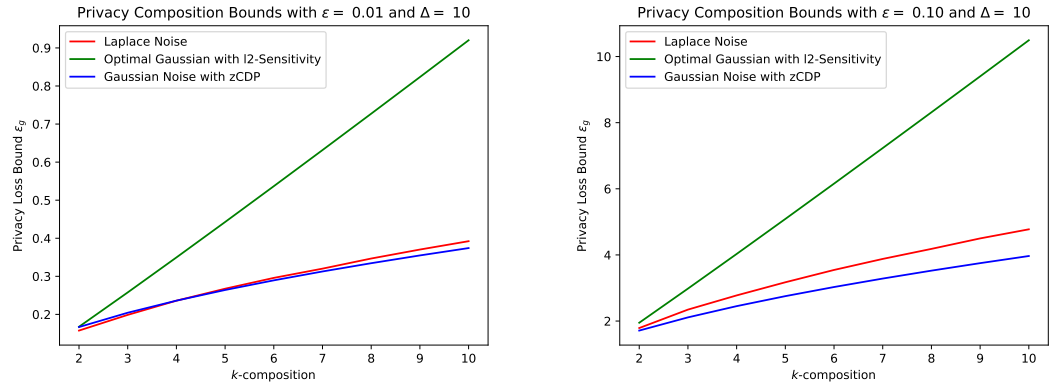

Figure 4: Similar to Figure 3, we now consider composition with k mechanisms, each adding noise to a ℓ_0-sensitivity $\Delta = 10$ function.

D.2. Gaussian Based Private Top-k Mechanisms

Given the results above, we then propose variants of existing mechanisms that use Gaussian noise rather than Laplace noise. From Rogers et al. (2020), we have the following table of mechanisms for data analytics based on histogram data. In the Δ-restricted sensitivity setting we assume that the number of categories a single user can impact is at most Δ (i.e.the ℓ_0-sensitivity) and the unrestricted sensitivity setting has no such restriction, but requires limiting the output to the top-k. Furthermore, the known domain setting is where the algorithms are given the set of categories over the histogram (the labels for the x-axis), because they cannot be given by the data. Whereas the unknown domain has no such restriction and must have a parameter $\bar{d}$ for an upper bound on the number of distinct elements the histogram can have. Each setting requires a ℓ_∞-sensitivity bound τ.

Rather than present each algorithm here, we summarize each one and present a variant of it using Gaussian noise. The $\texttt{KnownLap}^{\Delta,\tau}$ mechanism can easily be replaced with a Gaussian mechanism

	Δ-restricted sensitivity	unrestricted sensitivity
Known Domain	$\texttt{KnownLap}^{\Delta,\tau}$ Dwork et al. (2006b)	$\texttt{KnownGumb}^{k,\tau}$ McSherry and Talwar (2007)
Unknown Domain	$\texttt{UnkLap}^{\Delta,\bar{d},\tau}$ Rogers et al. (2020)	$\texttt{UnkGumb}^{k,\bar{d},\tau}$ Rogers et al. (2020)

Table 1: DP algorithms for various data analytics tasks

that adds Gaussian noise to each count with $\tau\sigma$ standard deviation to ensure $(\frac{\Delta}{2\sigma^2}, \frac{\sqrt{\Delta}}{\sigma})$-CDP Dwork and Rothblum (2016) and $\frac{\Delta}{2\sigma^2}$-zCDP Bun and Steinke (2016) since at most Δ bins in neighboring histograms can change.

The unrestricted sensitivity algorithms are based on the exponential mechanism (using Gumbel noise) to first discover the elements in the top-k and then use the Laplace mechanism to release noisy counts on the discovered elements. A simple modification of this algorithm is to still use exponential mechanisms to discover the elements but then use Gaussian noise on the resulting counts. Hence, adding Gaussian noise to the count of each discovered element with $\tau\sigma$ standard deviation will guarantee $(\frac{k}{2\sigma^2}, \frac{\sqrt{k}}{\sigma})$-CDP and $\frac{k}{2\sigma^2}$-zCDP. However, this ignores a useful detail that the first phase of exponential mechanisms is giving, a ranked list of elements. Hence, we propose a simple post processing function of the Gaussian mechanism that will respect the order given by the first phase of domain discovery. We solve a constrained least squares problem to return the maximum likelihood estimator for the true counts given an ordering. We present the $\texttt{CountMLE}^{\sigma}$ procedure in Algorithm 2. Because this is a post-processing function of the Gaussian mechanism, the privacy

Algorithm 2 $\texttt{CountMLE}^{\sigma}$; Return noisy counts subject to a fixed ordering

Input: Histogram $\mathbf{h} = h_1, \cdots, h_k$ of k elements with ℓ_∞-sensitivity τ and an ordering $i_1, i_2, \cdots i_k$.
Output: Noisy counts $\tilde{h}_{i_1}, \tilde{h}_{i_2}, \cdots, \tilde{h}_{i_k}$.
Add noise $\hat{\mathbf{h}} = (h_1, \cdots, h_k) + (Z_1, \cdots, Z_k)$ where $Z_i \sim \mathrm{N}\left(0, \tau^2\sigma^2\right)$.
Solve the following and let $\tilde{\mathbf{h}} = (\tilde{h}_1, \cdots, \tilde{h}_k)$ be the solution:

$$\begin{aligned} \min_{\mathbf{x}} \quad & ||\hat{\mathbf{h}} - \mathbf{x}||_2 \\ \text{s.t.} \quad & x_{i_1} \geq \cdots \geq x_{i_k} \geq 0 \end{aligned} \tag{11}$$

Return $(\tilde{h}_{i_1}, \cdots, \tilde{h}_{i_k})$.

parameters remain the same given the outcome of $\texttt{CountMLE}^{\sigma}$. Note that we could use an ℓ_1 loss to find the MLE when using Laplace noise, however this would not guarantee a unique solution. An additional advantage of using the ℓ_2 loss is that we can use standard constrained least squares numerical methods, e.g. 'nnls' in scipy.

We give the comparison between $\texttt{CountMLE}^{\sigma}$ and simply adding Laplace noise to the counts with the same overall privacy budget in Figure 5 and ℓ_∞-sensitivity of $\tau = 1$. In particular we consider the top-25 words in Macbeth and fix $\varepsilon = 0.1$ in $\texttt{KnownLap}^{25,1}$. We then compute the overall privacy parameter with $\delta = 10^{-6}$ using the optimal composition bound to get $\varepsilon'(\delta) = 2.08$.

Solving for σ in Lemma 29 with the given total $\epsilon'(\delta)$ and $\delta = 10^{-6}$, we get $\sigma = 13.1$. We then ran the two privatized count algorithms over 100 trials and plot 1 standard deviation from the empirical average. From the figure, we can see that $\texttt{CountMLE}^{\sigma}$ has smaller error for smaller counts than the basic Laplace mechanism with the same privacy guarantee.

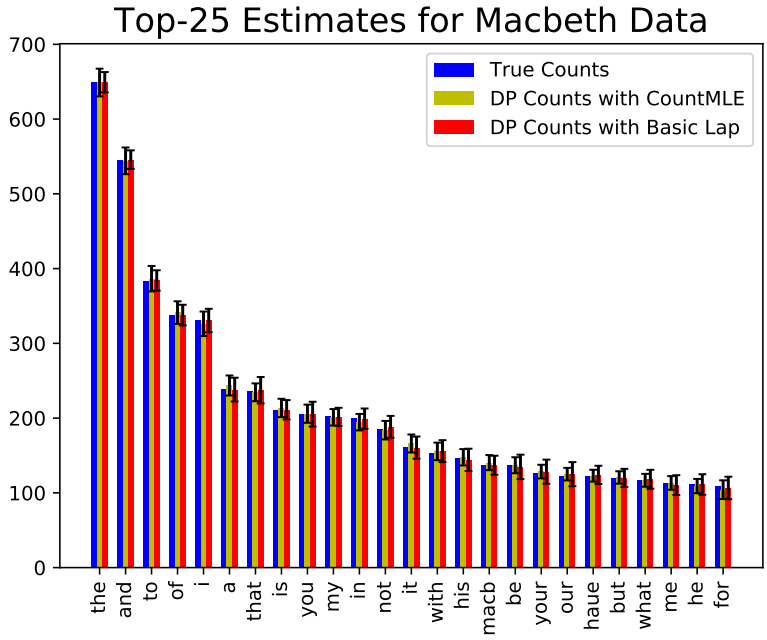

Figure 5: Comparison of $(2.08, 10^{-6})$-DP noise addition mechanisms given a fixed ordering.

Note that in repeated calls to $\texttt{TruncGauss}^{\Delta_i,\bar{d}_i,\tau_i}$, where each round i may have different parameter values, we can apply zCDP composition to get an overall privacy guarantee over the entire interaction.

See Figure 6 for a plot of the truncation level T from (12) when compared to the truncation level in $\texttt{UnkLap}^{\Delta,\bar{d},\tau}$ for various ℓ_0-sensitivities Δ and a given (ε, δ)-DP guarantee.

$$\delta = \Delta\left(1 - \frac{\Phi\left(\frac{T}{\tau\sigma}\right) - \Phi\left(\frac{\tau - T}{\tau\sigma}\right)}{\Phi\left(\frac{T}{\tau\sigma}\right) - \Phi\left(\frac{-T}{\tau\sigma}\right)}\right) \tag{12}$$

Note that the truncation level in $\texttt{UnkLap}^{\Delta,\bar{d},\tau}$ has Laplace noise added to it, whereas the truncation level in $\texttt{TruncGauss}^{\Delta,\bar{d},\tau}$ is fixed.

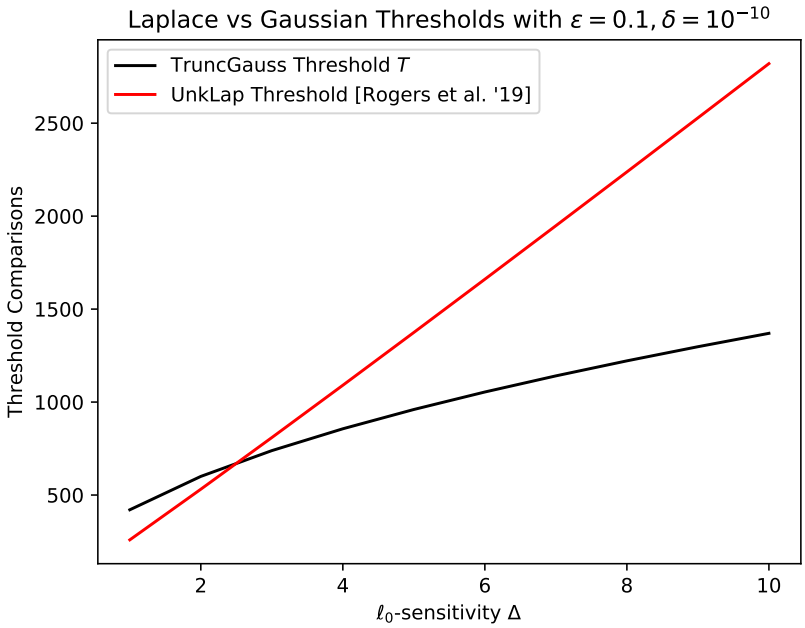

Figure 6: Comparing the truncation level T computed in (12) with that given in $\texttt{UnkLap}^{\Delta,\bar{d},\tau}$. We fix ℓ_∞-sensitivity $\tau = 1$ as well as the overall privacy parameters $(\varepsilon = 0.1, \delta = 10^{-10})$.