Optimal Differential Privacy Composition for Exponential Mechanisms: Supplementary Materials

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

A. Omitted details in Section 2

Recall that we defined

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

and stated that

Lemma A.1. *The following facts hold*

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

Proof of Lemma A.1. 1. Let $u(x,y) = \frac{1}{\varepsilon}\ln P[M(x)=y]$.

$$\begin{aligned} M \text{ is } \varepsilon\text{-DP} &\Leftrightarrow \left|\ln \tfrac{P[M(x')=y]}{P[M(x)=y]}\right| \leqslant \varepsilon \text{ for all neighboring } x, x' \\ &\Leftrightarrow |\varepsilon u(x',y) - \varepsilon u(x,y)| \leqslant \varepsilon \text{ for all neighboring } x, x' \\ &\Leftrightarrow |u(x',y) - u(x,y)| \leqslant 1 \text{ for all neighboring } x, x' \\ &\Leftrightarrow \Delta u \leqslant 1. \end{aligned}$$

2. $\mathbb{M}_1^\varepsilon \subsetneqq \mathbb{M}_2^\varepsilon \subsetneqq \mathbb{M}_1^{2\varepsilon}$.

   $\mathbb{M}_1^\varepsilon \subseteq \mathbb{M}_2^\varepsilon$ is straightforward from definition. $\mathbb{M}_2^\varepsilon \subseteq \mathbb{M}_1^{2\varepsilon}$ follows from the well-known Theorem 1.

   Next we show the strict inequalities. For simplicity we assume $X = Y = \{0,1\}$. Consider $u(x,y) = x - 2xy$. Obviously $\Delta u = 1$, and hence the resulting mechanism $M_{u,\varepsilon} \in \mathbb{M}_2^\varepsilon$. However, if we compute its normalized log probability function

$$\tilde{u}(x,y) = \tfrac{1}{\varepsilon}\ln P[M_{u,\varepsilon}(x)=y] = \tfrac{1}{\varepsilon}\ln\frac{\mathrm{e}^{\varepsilon u(x,y)}}{\sum_y \mathrm{e}^{\varepsilon u(x,y)}} = u(x,y) - \tfrac{1}{\varepsilon}\ln\underbrace{\textstyle\sum_y \mathrm{e}^{\varepsilon u(x,y)}}_{f(x)}.$$

   Easy calculation shows $f(0) = 2, f(1) = \mathrm{e}^\varepsilon + \mathrm{e}^{-\varepsilon} > f(0)$. We have

$$\tilde{u}(1,y) - \tilde{u}(0,y) = 1 - 2y - \big(\underbrace{\tfrac{1}{\varepsilon}\ln f(1) - \tfrac{1}{\varepsilon}\ln f(0)}_{\gamma}\big).$$

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

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

Since $f(1) > f(0)$, we have $\gamma > 0$. Taking maximum over y, we have

$$\Delta\tilde{u} = \max_{y\in\{0,1\}} |\tilde{u}(1,y) - \tilde{u}(0,y)| = \max\{|-1-\gamma|, |1-\gamma|\} = \max\{|1+\gamma|, |1-\gamma|\}.$$

As long as $\gamma \neq 0$, we have $\Delta\tilde{u} > 1$, which is to say, $M_{u,\varepsilon} \notin \mathbb{M}_1^\varepsilon$. This proves the first strict inequality.

For the second strict inequality, consider the following randomized response: $M(0) = \mathrm{Bern}(p), M(1) = \mathrm{Bern}(1-p)$ where $p = \frac{\mathrm{e}^{2\varepsilon}}{1+\mathrm{e}^{2\varepsilon}}$. It's easy to verify that $M \in \mathbb{M}_1^{2\varepsilon}$. We want to argue that it is not in $\mathbb{M}_2^\varepsilon$. Suppose it was, i.e. there is a quality score u with $\Delta u \leqslant 1, P[M(x) = y] \sim \mathrm{e}^{\varepsilon u(x,y)}$. Then u must look like the following for some numbers a, b

$u(x,y)$	$y=0$	$y=1$
$x=0$	$a + \frac{1}{\varepsilon}\ln(1-p)$	$a + \frac{1}{\varepsilon}\ln p$
$x=1$	$b + \frac{1}{\varepsilon}\ln p$	$b + \frac{1}{\varepsilon}\ln(1-p)$

We can compute sensitivity as follows

$$u(0,1) - u(1,1) = a - b + \tfrac{1}{\varepsilon}\ln\tfrac{p}{1-p} = a - b + 2$$
$$u(0,0) - u(1,0) = a - b + \tfrac{1}{\varepsilon}\ln\tfrac{1-p}{p} = a - b - 2$$

$\Delta u \leqslant 1$ requires that both of them are in $[-1,1]$, but $a - b + 2 \leqslant 1$ and $a - b - 2 \geqslant -1$ obviously contradict each other. This shows the randomized response is in $\mathbb{M}_1^{2\varepsilon}$ but not in $\mathbb{M}_2^\varepsilon$, hence the second strict inequality.

□

Proof of Proposition 1. (a) By definition,

$$\begin{aligned}
\widetilde{\Delta}u' &= \sup_{x\sim x'} \left\{ (\max_y - \min_y) \left\{ u'(x',y) - u'(x,y) \right\} \right\} \\
&= \sup_{x\sim x'} \left\{ (\max_y - \min_y) \left\{ u(x',y) - u(x,y) + f(x') - f(x) \right\} \right\} \\
&= \sup_{x\sim x'} \left\{ (\max_y - \min_y) \left\{ u(x',y) - u(x,y) \right\} \right\} = \widetilde{\Delta}u
\end{aligned}$$

(b)

$$\begin{aligned}
\widetilde{\Delta}u &= \sup_{x\sim x'} \left\{ (\max_y - \min_y) \left\{ u(x',y) - u(x,y) \right\} \right\} \\
&\leqslant \sup_{x\sim x'} \left\{ \max_y \left\{ u(x',y) - u(x,y) \right\} \right\} + \sup_{x\sim x'} \left\{ -\min_y \left\{ u(x',y) - u(x,y) \right\} \right\} \\
&\leqslant \Delta u + \Delta u = 2\Delta u.
\end{aligned}$$

(c) Obviously, $\widetilde{\mathbb{M}}_1^\varepsilon \subseteq \widetilde{\mathbb{M}}_2^\varepsilon$. To see the reverse direction, notice that if a mechanism can be realized by two scores u, u', then $u - u'$ is independent of y. By (a) it implies $\widetilde{\Delta}u = \widetilde{\Delta}u'$. Hence $\widetilde{\mathbb{M}}_2^\varepsilon \subseteq \widetilde{\mathbb{M}}_1^\varepsilon$.

□

Proof of Proposition 2. The proof first appears in (Durfee and Rogers, 2019). We repeat here for completeness. Let $p(x,y) = \ln P[M_{u,\varepsilon}(x) = y]$. Since $\frac{1}{\varepsilon}p$ and u realize the same mechanism, it follows from (a) of Proposition 2 that $\widetilde{\Delta}\frac{1}{\varepsilon}p = \widetilde{\Delta}u$. Therefore, $\widetilde{\Delta}p = \widetilde{\Delta}u \cdot \varepsilon$. It suffices to show that $M_{u,\varepsilon}$ is $\widetilde{\Delta}p$-DP, i.e. for any neighboring x, x' we have

$$|p(x,y) - p(x',y)| \leqslant \widetilde{\Delta}p, \forall y.$$

Let

$$t = \max_y \{p(x,y) - p(x',y)\},$$
$$s = \min_y \{p(x,y) - p(x',y)\}.$$

Since $p(x,y)$ and $p(x',y)$ both satisfy a normalizing condition, i.e. $\sum_y \mathrm{e}^{p(x,y)} = \sum_y \mathrm{e}^{p(x',y)} = 1$, it's impossible that $p(x,y)$ is uniformly larger or smaller than $p(x',y)$. So we have $s \leqslant 0 \leqslant t$. Therefore for any y,

$$|p(x,y) - p(x',y)| \leqslant \max\{|t|,|s|\} \leqslant t - s \leqslant \widetilde{\Delta} p.$$

The proof is now complete. □

Proof of Proposition 3. The equivalence between (1) and (2) follows directly from the definition of $\widetilde{\mathbb{M}}_1^\varepsilon$. The equivalence of (1) and (3) follows essentially from the argument for Proposition 2. Let $p(x,y) = \ln P[M(x) = y]$ and

$$t = \max_y \{p(x,y) - p(x',y)\},$$
$$s = \min_y \{p(x,y) - p(x',y)\}.$$

We have seen that $s \leqslant 0 \leqslant t$. ε-BR property is equivalent to that $t - s \leqslant \varepsilon$, i.e. $s \geqslant t - \varepsilon$. Therefore for each pair of x, x' we have

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

The reverse argument also holds. The proof is now complete. □

B. Proof of Theorem 2 and 3 via reduction

In this section we prove the main results of this paper, Theorem 2 and 3. In Appendix B.1 we point out an issue that is not mentioned in the main body and explain why it does not affect the result. In Appendix B.2 we prove the basic tools – Lemma 4.1 and 4.2. In particular, Lemma 4.1 reduces bounded range mechanisms to Bernoulli distributions. In Appendix B.3 we use the tools to prove Theorem 2. The heavy calculation are relegated to Appendix C. In Appendix B.3 we prove Theorem 3.

We remark that the reduction technique in Appendix B.2 is first introduced to differential privacy by Kairouz et al. (2017), relying on the magical Blackwell's theorem. Later on it is greatly developed by Dong et al. (2019). Using their language and techniques, it is possible to get simplified proofs, and simultaneously avoid the use of Blackwell's theorem. However, we follow the approach in Kairouz et al. (2017) to minimize machinery and make it accessible to the broadest audience.

B.1. Handling randomization

In Section 2 we claimed to have justified that we should focus on the follwing class

$$\mathbb{M}^\varepsilon = \{M : X \to Y \mid M \text{ is } \varepsilon\text{-BR}\}.$$

It models the scenario where a data analyst is allowed to choose from a collection of queries that lead to quality scores with bounded range. However, it does not take into consideration that the data analyst can randomize over this collection of queries. That being said, what really needs to be modeled, is the convex hull of this class, namely, $\mathrm{conv}(\mathbb{M}^\varepsilon)$. This issue is rarely raised in the existing literature because previously considered classes of mechanisms are all *convex*. For example, a convex combination of (ε,δ)-DP mechanisms is still (ε,δ)-DP. However, there are simple examples showing that $\mathrm{Conv}(\mathbb{M}^\varepsilon) \neq \mathbb{M}^\varepsilon$. So what we should really consider, is the adaptive/non-adaptive composition with in the class $\mathrm{conv}(\mathbb{M}^\varepsilon)$. For example, in Section 2.3 we explained what we mean by "non-adaptive composition of ε-BR mechanisms". Taking randomization into consideration, we should consider "non-adaptive composition of *randomized* ε-BR mechanisms". More specifically, a mechanism $M : X \to Y^k$ is a k-fold non-adaptive composition of *randomized* ε-BR mechanisms if there are $M_i : X \to Y, i = 1, 2 \ldots, k$, each in $\mathrm{conv}(\mathbb{M}^\varepsilon)$, such that $M(x) = \big(M_1(x), \ldots, M_k(x)\big)$. The adaptive version can be similarly defined.

However, we argue here that randomization is not a concern: any (ε,δ)-DP guarantee that holds for composition of BR mechanisms also holds for composition of randomized BR mechanisms. The reason is simple: composition of randomized BR mechanisms must also be a randomization over composition of BR mechanisms. To see this, let M be a composition of randomized BR mechanisms. In the process of composition, although randomization can be introduced in every step,

we can condition on all the randomness introduced in the process, and end up with a composition of pure (in contrast to randomized) BR mechanisms. We can recover M by adding back all the randomness extracted, all at once. This argument explains why Theorem 2 and 3 also hold with randomization.

For the ease of proofs, we will continue to assume that the component mechanisms are "pure" BR mechanisms.

B.2. Proof of Lemma 4.1 and 4.2

In order to use Blackwell's theorem ((Blackwell, 1950), Theorem 10), we will need to first establish some notation. For a pair of probability distributions P and Q on a common probability space Ω, its trade-off function Dong et al. (2019) describes the hardness of the hypothesis testing problem $H_0 : P$ vs $H_1 : Q$. Let $E \subseteq \Omega$ be an arbitrary rejection region and $\alpha_E = P[E], \beta_E = 1 - Q[E]$ be the type I and type II errors of the test E respectively. Fix a level α_0 and let E run over all test with type I error at most α_0, the minimal type II error is

$$\inf\{\beta_E : E \text{ is a rejection region s.t. } \alpha_E \leqslant \alpha_0\}.$$

This correspondence of α_0 to the minimal type II error defines a function from $[0, 1]$ to $[0, 1]$. We will call this function $T(P, Q)$. Formally,

$$\begin{aligned} T(P, Q) : [0, 1] &\to [0, 1] \\ \alpha_0 &\mapsto \inf\{\beta_E : \alpha_E \leqslant \alpha_0\} \end{aligned}$$

The following form of Blackwell's theorem is taken from (Dong et al., 2019).

Theorem 1. *Let P, Q be probability distributions on Y and P', Q' be probability distributions on Z. The following two statements are equivalent:*

(a) $T(P, Q) \leqslant T(P', Q')$.

(b) There exists a randomized algorithm $\mathrm{Proc} : Y \to Z$ *such that* $\mathrm{Proc}(P) = P', \mathrm{Proc}(Q) = Q'$.

We now prove that we can post-process a pair of Bernoulli distributions to simulate any BR mechanism on neighboring inputs.

Proof of Lemma 4.1. Let P be the outcome distribution of $M(x^0)$ and Q be the outcome distribution of $M(x^1)$. By Proposition 3, we know there exists some $t \in [0, \varepsilon]$ such that

$$t - \varepsilon \leqslant \ln \frac{Q(y)}{P(y)} \leqslant t.$$

Equivalently, for any event $E \subseteq \mathcal{Y}$,

$$\mathrm{e}^{t-\varepsilon} P[E] \leqslant Q[E] \leqslant \mathrm{e}^t P[E]. \tag{1}$$

Applying the same rule for the complement event E^c, we have

$$\mathrm{e}^{t-\varepsilon} P[E^c] \leqslant Q[E^c] \leqslant \mathrm{e}^t P[E^c]. \tag{2}$$

The second inequality of (1) and the first inequality of (2) imply

$$1 - \beta_E \leqslant \mathrm{e}^t \alpha_E, \quad \mathrm{e}^{t-\varepsilon}(1 - \alpha_E) \leqslant \beta_E. \tag{3}$$

Let the piece-wise linear function $l_{t,\varepsilon} : [0, 1] \to [0, 1]$ be defined as

$$l_{t,\varepsilon}(x) = \max\{1 - \mathrm{e}^t x, \mathrm{e}^{t-\varepsilon}(1 - x)\}.$$

It's easy to see that (3) implies $T(P, Q) \geqslant l_{t,\varepsilon}$ pointwise in $[0, 1]$. Furthermore, it is straightforward to verify that $l_{t,\varepsilon} \equiv T\left(\mathrm{Bern}(p_t), \mathrm{Bern}(q_t)\right)$. Therefore, there must be a $t = t(M, x, x')$ such that

$$T\left(\mathrm{Bern}(p_t), \mathrm{Bern}(q_t)\right) \leqslant T\big(M(x), M(x')\big).$$

Applying Theorem 1 then gives our desired claim. □

Proof of Lemma 4.2. Starting from definition,

$$\begin{aligned}\delta_{\text{opt}}(M,\varepsilon) &= \inf\left\{\delta : M \text{ is } (\varepsilon,\delta)\text{-DP}\right\}\\ &= \inf_{x\sim x'} \min\{\delta : P[M(x')\in E] \leqslant \mathrm{e}^{\varepsilon} P[M(x)\in E] + \delta, \forall E\}\\ &= \sup_{x\sim x'} \max_{E} \left\{P[M(x')\in E] - \mathrm{e}^{\varepsilon} P[M(x)\in E]\right\}.\end{aligned}$$

Let $p_x(y) = P[M(x) = y]$ be the density function. Then for a fixed pair x, x'

$$\max_{E} \left\{P[M(x')\in E] - \mathrm{e}^{\varepsilon} P[M(x)\in E]\right\} = \max_{E} \int_{E} \left[p_{x'}(y) - \mathrm{e}^{\varepsilon} p_x(y)\right] \mathrm{d}y$$

Obviously, the maximum is attained at the event that the integrand being non-negative. That is, $E = \{y : p_{x'}(y) - \mathrm{e}^{\varepsilon} p_x(y) \geqslant 0\}$. Therefore,

$$\begin{aligned}\max_{E} \left\{P[M(x')\in E] - \mathrm{e}^{\varepsilon} P[M(x)\in E]\right\} &= \int \left[p_{x'}(y) - \mathrm{e}^{\varepsilon} p_x(y)\right]_+ \mathrm{d}y \qquad (*)\\ &= \int p_{x'}(y)\left[1 - \mathrm{e}^{\varepsilon}\cdot \tfrac{p_x(y)}{p_{x'}(y)}\right]_+ \mathrm{d}y\\ &= \int p_{x'}(y)\left[1 - \mathrm{e}^{\varepsilon - L(y;x,x')}\right]_+ \mathrm{d}y\\ &= \mathbb{E}_{y\sim M(x')}\left[1 - \mathrm{e}^{\varepsilon - L(y;x,x')}\right]_+ \mathrm{d}y\end{aligned}$$

Taking supremum over x, x' yields the desired result. □

B.3. Proof of Theorem 2

Let $M : X \to Y_1 \times \cdots \times Y_k$ be a k-fold non-adaptive composition of ε-BR mechanisms, i.e. $M(x) = \big(M_1(x), \ldots, M_k(x)\big)$ such that $M_i : X \to Y_i, i = 1, 2 \ldots, k$ are ε-BR mechanisms. Fix neighboring datasets x, x', Lemma 4.1 implies that there are randomized mappings $K_i : \{0,1\} \to Y_i, i = 1, 2, \ldots, k$ such that

$$\begin{aligned}&\text{If } b \sim \mathrm{Bern}(p_t), \text{ then } K_i(b) \sim M_i(x)\\ &\text{If } b \sim \mathrm{Bern}(q_t), \text{ then } K_i(b) \sim M_i(x')\end{aligned}$$

As a consequence, we can construct $K : \{0,1\}^k \to Y_1 \times \cdots \times Y_k$ such that

$$\begin{aligned}&\text{If } (b_1,\ldots,b_k) \sim \mathrm{Bern}(p_{t_1}) \times \cdots \times \mathrm{Bern}(p_{t_k}), \text{ then } K(b_1,\ldots,b_k) \sim M(x)\\ &\text{If } (b_1,\ldots,b_k) \sim \mathrm{Bern}(q_{t_1}) \times \cdots \times \mathrm{Bern}(q_{t_k}), \text{ then } K(b_1,\ldots,b_k) \sim M(x')\end{aligned}$$

Basically the distinction between the two Bernoulli products is larger than that between the neighboring distributions $M(x)$ and $M(x')$. In other words, for k-fold non-adaptive composition of ε-BR mechanisms, the worst case neighbors are Bernoulli products. Together with Lemma 4.2, we can compute $\delta_k^{\text{optNA}}(\varepsilon_g)$, the optimal δ such that all k-fold non-adaptive composition of ε-BR mechanisms are (ε_g, δ)-DP.

It turns out the intermediate formula $(*)$ is more useful, where we replace $p_x(y)$ and $p_{x'}(y)$ by the probability functions of $\mathrm{Bern}(p_{t_1}) \times \cdots \times \mathrm{Bern}(p_{t_k})$ and $\mathrm{Bern}(q_{t_1}) \times \cdots \times \mathrm{Bern}(q_{t_k})$ respectively. For example, the probability function of $\mathrm{Bern}(p_{t_1}) \times \cdots \times \mathrm{Bern}(p_{t_k})$ is

$$f(b_1,\ldots,b_k) = \prod_{i=1}^{k} p_{t_i}^{b_i}(1-p_{t_i})^{1-b_i}.$$

Integral in $(*)$ is replaced by summation over $b_1 b_2 \ldots b_k \in \{0,1\}^k$. Note that we need to take supremum over $t_1, \ldots, t_k$ because they depend on the neighboring datasets x, x' and hence should be maximized over as in Lemma 4.2.

$$\delta_k^{\text{optNA}}(\varepsilon_g) = \sup_{t_1,\ldots,t_k} \sum_{b_1 b_2 \ldots b_k \in \{0,1\}^k} \left[\prod_{i=1}^{k} q_{t_i}^{b_i}(1-q_{t_i})^{1-b_i} - \mathrm{e}^{\varepsilon_g} \prod_{i=1}^{k} p_{t_i}^{b_i}(1-p_{t_i})^{1-b_i}\right]_+.$$

Let S be the subset of indices such that $b_i = 0$. This change of dummy variable yields an equivalent formula for $\delta_k^{\text{optNA}}(\varepsilon_g)$.

$$\delta_k^{\text{optNA}}(\varepsilon_g) = \sup_{\mathbf{t}\in[0,\varepsilon]^k} \sum_{S\subseteq\{1,\dots,k\}} \left[\prod_{i\notin S} q_{t_i} \prod_{i\in S}(1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i\notin S} p_{t_i} \prod_{i\in S}(1-p_{t_i})\right]_+ .$$

In order to prove Theorem 2, it suffices to show $\delta_k^{\text{optNA}}(\varepsilon_g)$ above agrees with the expression $\delta_k^{\text{NA}}(\varepsilon_g)$ defined in the statement. This is highly non-trivial and involves identifying the symmetry of the maximizers of a high dimensional non-convex and non-smooth optimization problem. We state the lemma here and will accomplish it in Section C.

Lemma B.1. *$\delta_k^{\text{optNA}}(\varepsilon_g) = \delta_k^{\text{NA}}(\varepsilon_g)$. That is,*

$$\sup_{\mathbf{t}\in[0,\varepsilon]^k} \sum_{S\subseteq\{1,\dots,k\}} \left[\prod_{i\notin S} q_{t_i} \prod_{i\in S}(1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i\notin S} p_{t_i} \prod_{i\in S}(1-p_{t_i})\right]_+ = \max_{0\leqslant\ell\leqslant k} \sum_{i=0}^{k} \binom{k}{i} p_{t_\ell^*}^{k-i}(1-p_{t_\ell^*})^i (\mathrm{e}^{kt_\ell^* - i\varepsilon} - \mathrm{e}^{\varepsilon_g})_+,$$

where $t_\ell^* = \frac{\varepsilon_g + (\ell+1)\varepsilon}{k+1}$. If $t_\ell^* \notin [0,\varepsilon]$, then we round it to the closest point in $[0,\varepsilon]$.

B.4. Proof of Theorem 3

First we prove a useful lemma. Let $M_1 : X \to Y, M_2 : X \times Y \to Z$ be two randomized algorithms on databases and $M : X \to Y \times Z$ be their adaptive composition as explained in Section 2.1. Let $p_x(y) = P[M_1(x) = y]$ be the density function / probability function of the first mechanism and $L_{x,x'}(y) = \ln \frac{p_{x'}(y)}{p_x(y)}$ be the log likelihood ratio. Then we have

Lemma B.2. *If M_2 is $\big(\varepsilon, \delta_2(\varepsilon)\big)$-DP for all ε, then the composition M is $\big(\varepsilon, \delta_{\text{compo}}(\varepsilon)\big)$-DP for all ε with*

$$\delta_{\text{compo}}(\varepsilon) = \sup_{x,x'} \int \delta_2(\varepsilon - L_{x,x'}(y)) \cdot p_{x'}(y)\,\mathrm{d}y.$$

Proof of Lemma B.2. Using similar notation and Lemma 4.2, we have that the composition M is (ε,δ)-DP with

$$\begin{aligned}
\delta &= \sup_{x,x'} \iint \big[p_{x'}(y,z) - \mathrm{e}^{\varepsilon} p_x(y,z)\big]_+ \,\mathrm{d}y\,\mathrm{d}z \\
&= \sup_{x,x'} \iint \big[p_{x'}(z|y)p_{x'}(y) - \mathrm{e}^{\varepsilon} p_x(z|y)p_x(y)\big]_+ \,\mathrm{d}y\,\mathrm{d}z \\
&= \sup_{x,x'} \iint p_{x'}(y)\big[p_{x'}(z|y) - \mathrm{e}^{\varepsilon} p_x(z|y)\tfrac{p_x(y)}{p_{x'}(y)}\big]_+ \,\mathrm{d}y\,\mathrm{d}z \\
&= \sup_{x,x'} \int p_{x'}(y) \left(\int \big[p_{x'}(z|y) - \mathrm{e}^{\varepsilon - L_{x,x'}(y)} p_x(z|y)\big]_+ \,\mathrm{d}z\right) \mathrm{d}y \\
&\leqslant \sup_{x,x'} \int p_{x'}(y)\delta_2(\varepsilon - L_{x,x'}(y))\,\mathrm{d}y.
\end{aligned}$$

So we can pick $\delta = \delta_{\text{compo}}(\varepsilon) = \sup_{x,x'} \int \delta_2(\varepsilon - L_{x,x'}(y)) \cdot p_{x'}(y)\,\mathrm{d}y$. □

To prove Theorem 3, we do two inductions on k for k-fold adaptive composition of ε-BR mechanisms, one for validity of the privacy guarantee and one for its optimality. Now we start from the induction for validity.

For the base case where $k = 0$, we apply no mechanism, which amounts to set all inputs the same, i.e. $x = x'$. In that case $L(y; x, x') = 0$ in Lemma 4.2, and $\mathbb{E}_{y\sim M(x')}\big[1 - \mathrm{e}^{\varepsilon_g - L(y;x,x')}\big]_+ \,\mathrm{d}y \equiv \max\{1 - \mathrm{e}^{\varepsilon_g}, 0\}$.

Assuming all k-fold adaptive composition of ε-BR mechanisms are $\big(\varepsilon_g, \delta_k^{\text{A}}(\varepsilon_g)\big)$-DP for any ε_g. We proceed to consider a $k+1$-fold adaptive composition $M : X \to Y_1 \times \cdots \times Y_{k+1}$. It can be decomposed as the adaptive composition of an ε-BR mechanism $M_1 : X \to Y_1$ and a k-fold composition of ε-BR mechanism $M_2 : X \times Y_1 \to Y_2 \times \cdots \times Y_{k+1}$. Fix a pair of neighboring datasets x, x'. Using Lemma 4.1, we know that the first step can be assumed to be Bernoulli distributions $\mathrm{Bern}(p_t)$ and $\mathrm{Bern}(q_t)$ where $t = t(x, x')$, and then the second step remains to be $\big(\varepsilon_g, \delta_k^{\text{A}}(\varepsilon_g)\big)$-DP for any ε_g. Now we

can use Lemma B.2 and conclude that any $k+1$-fold adaptive composition of ε-BR mechanisms is $\big(\varepsilon_g, \delta^{\mathrm{A}}_{k+1}(\varepsilon_g)\big)$-DP for any ε_g, where

$$\delta^{\mathrm{A}}_{k+1}(\varepsilon_g) = \sup_{x,x'} q_{t(x,x')}\delta^{\mathrm{A}}_k(\varepsilon_g - L_{x,x'}(1)) + (1-q_{t(x,x')})\delta^{\mathrm{A}}_k(\varepsilon_g - L_{x,x'}(0))$$

$L_{x,x'}$ is the log likelihood ratio between $\mathrm{Bern}(p_t)$ and $\mathrm{Bern}(q_t)$. So using Equation (1) at the beginnig of Section 3, we have

$$L_{x,x'}(1) = \ln\frac{q_t}{p_t} = t, \quad L_{x,x'}(0) = \ln\frac{1-q_t}{1-p_t} = t-\varepsilon.$$

Therefore,

$$\begin{aligned}\delta^{\mathrm{A}}_{k+1}(\varepsilon_g) &= \sup_{x,x'} q_{t(x,x')}\delta^{\mathrm{A}}_k(\varepsilon_g - t) + (1-q_{t(x,x')})\delta^{\mathrm{A}}_k(\varepsilon_g - t(x,x') + \varepsilon)\\ &= \sup_{t\in[0,\varepsilon]} q_t\delta^{\mathrm{A}}_k(\varepsilon_g - t) + (1-q_t)\delta^{\mathrm{A}}_k(\varepsilon_g - t + \varepsilon).\end{aligned}$$

This finishes the induction and proves the validity of the privacy guarantee in Theorem 3.

Now we start the induction for optimality. Obviously the base case $k=0$ is optimal. Assuming the recursive formula is optimal for k, i.e. for any given ε_g there is a k-fold adaptive composition of ε-BR mechanisms that is not (ε_g, δ)-DP for any $\delta < \delta^{\mathrm{A}}_k(\varepsilon_g)$. We claim it is also true for $k+1$-fold composition. In fact, for a fixed ε_g, let t^* be a maximizer in the formula for $\delta^{\mathrm{A}}_{k+1}(\varepsilon_g)$. By the induction hypothesis, we can construct M_1^* and M_2^*, both of which are k-fold adaptive compositions of ε-BR mechanisms, such that M_1^* achieves $\big(\varepsilon_g - t^*, \delta^{\mathrm{A}}_k(\varepsilon_g - t^*)\big)$-DP and M_2^* achieves $\big(\varepsilon_g - t^* + \varepsilon, \delta^{\mathrm{A}}_k(\varepsilon_g - t^* + \varepsilon)\big)$-DP. One can verify that the following mechanism achieves $\big(\varepsilon_g, \delta^{\mathrm{A}}_{k+1}(\varepsilon_g)\big)$-DP:

On the input x, sample a bit b from $\mathrm{Bern}(p_{t^*})$. Run $M_1^*(x)$ if $b=1$ and $M_2^*(x)$ otherwise.
On the input x', sample a bit b from $\mathrm{Bern}(q_{t^*})$. Run $M_1^*(x')$ if $b=1$ and $M_2^*(x')$ otherwise.

For completeness, the mechanism can run as if it had x' when it is fed with an input other than x or x'. Since the first step is ε-BR, it is by construction a $k+1$-fold adaptive compositions of ε-BR mechanisms. This finishes the induction for optimality, and hence finishes the proof of Theorem 3.

C. Proof of Lemma B.1

The goal of this section is to prove Lemma B.1 and hence finish the proof of Theorem 2. We need to simplify the expression

$$\delta^{\mathrm{optNA}}_k(\varepsilon_g) = \max_{\mathbf{t}\in[0,\varepsilon]^k} \underbrace{\sum_{S\subseteq\{1,\dots,k\}} \left[\prod_{i\notin S} q_{t_i}\prod_{i\in S}(1-q_{t_i}) - \mathrm{e}^{\varepsilon_g}\prod_{i\notin S} p_{t_i}\prod_{i\in S}(1-p_{t_i})\right]_+}_{\delta(\mathbf{t},\varepsilon_g)}$$

so that it agrees with $\delta^{\mathrm{NA}}_k(\varepsilon_g)$. Here $\mathbf{t} = (t_1,\cdots,t_k)\in[0,\varepsilon]^k$ and $\delta : [0,\varepsilon]^k\times\mathbb{R}\to[0,1]$ is our objective function (with a slight abuse of notation). Written in this way, we have $\delta^{\mathrm{optNA}}_k(\varepsilon_g) = \max_{\mathbf{t}\in[0,\varepsilon]^k}\delta(\mathbf{t},\varepsilon_g)$. We have been using $\max$ instead of $\sup$ and will continue to use it because obviously $\delta(\mathbf{t},\varepsilon_g)$ is continuous in $\mathbf{t}$ and $\mathbf{t}$ belongs to a compact domain $[0,\varepsilon]^k$.

As we mentioned, the proof essentially involves solving a high dimensional, non-convex and non-smooth optimization problem. The complete solution is relatively long, so we divide the section into four subsections: Appendix C.1 reduces the dimension to one and Appendix C.2 solves the one-dimensional problem. Technical arguments for dimension reduction is in Appendix C.3, and the heavy calculation involved in the one-dimensional optimization is in Appendix C.4.

C.1. Strong symmetry of maximizers

We first show that when $\varepsilon_g\notin(-k\varepsilon, k\varepsilon)$, then the choice of $\delta(\mathbf{t},\varepsilon_g)$ does not depend on $\mathbf{t}\in[0,\varepsilon]^k$. However, this region for ε_g is not typically interesting in most DP applications.

Lemma C.1. *For any $\mathbf{t}\in[0,\varepsilon]^k$, if $\varepsilon_g\leqslant -k\varepsilon$ then $\delta(\mathbf{t},\varepsilon_g) = 1-\mathrm{e}^{\varepsilon_g}$, and if $\varepsilon_g\geqslant k\varepsilon$ then $\delta(\mathbf{t},\varepsilon_g) = 0$.*

Proof. Using the fact that $q_t = \mathrm{e}^t p_t$ and $(1-q_t) = \mathrm{e}^{t-\varepsilon}(1-p_t)$, we equivalently have

$$\delta(\mathbf{t}, \varepsilon_g) = \sum_{S \subseteq \{1,\dots,k\}} \prod_{i \notin S} p_{t_i} \prod_{i \in S} (1-p_{t_i}) \max\left\{\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\}$$

If $\varepsilon_g \geqslant k\varepsilon$ then $\max\{\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = 0$ for any $S \subseteq \{1,\dots,k\}$. Similarly, if $\varepsilon_g \leqslant -k\varepsilon$ then $\max\{\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = \mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}$ for any $S \subseteq \{1,\dots,k\}$ and we get

$$\delta(\mathbf{t}, \varepsilon_g) = \sum_{S \subseteq \{1,\dots,k\}} \left(\prod_{i \notin S} q_{t_i} \prod_{i \in S} (1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i \notin S} p_{t_i} \prod_{i \in S} (1-p_{t_i}) \right) = 1 - \mathrm{e}^{\varepsilon_g}$$

□

For the remainder of our analysis, we will focus on the interesting setting where $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$. Despite the large domain $[0,\varepsilon]^k$ of values to choose from in the $\max_{\mathbf{t}}$ for δ_{opt}, we show that the maximizer must have strong symmetry, i.e. $t_i = t^*$ for some t^* for all $i \in [k]$. This result is crucial in simplifing δ_k^{optNA} to δ_k^{NA}. We first give an easy condition on what the t_i must satisfy to optimize the δ parameter which will be important for proving a strict inequality in the subsequent claim.

Lemma C.2. *If $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$ then for any $\mathbf{t} \in [0,\varepsilon]^k$ such that $\delta(\mathbf{t}, \varepsilon_g) = \max_{\mathbf{t} \in [0,\varepsilon]^k} \delta(\mathbf{t}, \varepsilon_g)$, we must have*

$$\varepsilon_g < \sum_{i=1}^{k} t_i < \varepsilon_g + k\varepsilon$$

Proof. Using the fact that $q_t = e^t p_t$ and $(1-q_t) = \mathrm{e}^{t-\varepsilon}(1-p_t)$, we equivalently have

$$\delta(\mathbf{t}, \varepsilon_g) = \sum_{S \subseteq \{1,\dots,k\}} \prod_{i \notin S} p_{t_i} \prod_{i \in S} (1-p_{t_i}) \max\left\{\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\}$$

It then follows that if $\sum t_i \leqslant \varepsilon_g$ we must have

$$\max\left\{\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\} = 0$$

for any S and so $\delta(\mathbf{t}, \varepsilon_g) = 0$. However, if $\varepsilon_g < k\varepsilon$, then there must exist $\mathbf{t}$ such that $t_i < \varepsilon$ for each i and $\sum t_i > \varepsilon_g$. Setting $S = \emptyset$ we must have $p_{t_i} > 0$ for all i and $\max\{\mathrm{e}^{\sum t_i} - \mathrm{e}^{\varepsilon_g}, 0\} > 0$. Therefore, $\delta_k^{\text{optNA}}(\varepsilon_g) > 0$ and if $\sum t_i \leqslant \varepsilon_g$ we must have $\delta(\mathbf{t}, \varepsilon_g) < \delta_k^{\text{optNA}}(\varepsilon_g)$.

Similarly, if $\sum t_i \geqslant \varepsilon_g + k\varepsilon$ we must have the following for any subset S

$$\max\left\{\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\} = \mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}$$

We then have the following,

$$\begin{aligned}
\delta(\mathbf{t}, \varepsilon_g) &= \sum_{S \subseteq \{1,\dots,k\}} \prod_{i \notin S} p_{t_i} \prod_{i \in S} (1-p_{t_i}) \left(\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}\right) \\
&= \sum_{S \subseteq \{1,\dots,k\}} \prod_{i \notin S} q_{t_i} \prod_{i \in S} (1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i \notin S} p_{t_i} \prod_{i \in S} (1-p_{t_i}) = 1 - \mathrm{e}^{\varepsilon_g}
\end{aligned}$$

By the same reasoning, we have $\delta(\mathbf{t}, \varepsilon_g) > 1 - \mathrm{e}^{\varepsilon_g}$ if $\mathrm{e}^{\sum t_i - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g} < 0$ for some $S \subseteq \{1, \cdots, k\}$ and all $t_i \in (0, \varepsilon)$, which implies $p_{t_i} \in (0,1)$ for all i. Accordingly, we have $\delta(\mathbf{t}, \varepsilon_g) > 1 - \mathrm{e}^{\varepsilon_g}$ if $\sum t_i < \varepsilon_g + k\varepsilon$, and if $\varepsilon_g > -k\varepsilon$, there must exist positive t_i such that $\sum t_i < \varepsilon_g + k\varepsilon$. Therefore if $\sum t_i \geqslant \varepsilon_g + k\varepsilon$, we must have $\delta(\mathbf{t}, \varepsilon_g) < \max_{\mathbf{t} \in [0,\varepsilon]^k} \delta(\mathbf{t}, \varepsilon_g)$. □

The next lemma shows that taking the average of some t_i, t_j can only increase the value of $\delta(\mathbf{t}, \varepsilon_g)$. Further, this will strictly increase the δ when the t_i satisfy the condition of the lemma above. We will be able to easily conclude from this that δ cannot be optimal if $t_i \neq t_j$ for some i, j

Lemma C.3. *For any $\varepsilon_g \in \mathbb{R}$ and $\mathbf{t} \in [0, \varepsilon]^k$,*

$$\delta(\mathbf{t}, \varepsilon_g) \leqslant \delta\left(\left(\frac{t_1 + t_2}{2}, \frac{t_1 + t_2}{2}, t_3, ..., t_k\right), \varepsilon_g\right)$$

Further, the inequality is strict whenever $\varepsilon_g < \sum t_i < \varepsilon_g + k\varepsilon$ and $t_1 \neq t_2$.

The proof of this lemma will require quite a bit of technical detail which we relegate to Appendix C.3. We then have the immediate corollary.

Corollary C.1. *For any $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$ we must have the following for any $\mathbf{t} \in [0, \varepsilon]^k$ such that there exists some $t_i \neq t_j$*

$$\delta(\mathbf{t}, \varepsilon_g) < \delta_k^{\text{optNA}}(\varepsilon_g).$$

Proof. We will prove by contradiction and suppose $\delta(\mathbf{t}, \varepsilon_g) = \delta_k^{\text{optNA}}(\varepsilon_g)$ and $t_i \neq t_j$ for some pair of indices. Note that $\delta(\mathbf{t}, \varepsilon_g)$ is equal under permutation of the indices in $\mathbf{t}$, so without loss of generality, we let $t_1 \neq t_2$. From Lemma C.2, we must have $\varepsilon_g < \sum t_i < \varepsilon_g + k\varepsilon$. We then apply Lemma C.3 to get our contradiction

$$\delta(\mathbf{t}, \varepsilon_g) < \delta\left(\frac{t_1 + t_2}{2}, \frac{t_1 + t_2}{2}, t_3, ..., t_k\right) \leqslant \delta_k^{\text{optNA}}(\varepsilon_g)$$

□

Using the strong symmetry of maximizers, we can reduce the high dimensional optimization to a one-dimensional one.

Lemma C.4. *For any $\varepsilon_g \in \mathbb{R}$ and $\varepsilon \geqslant 0$*

$$\delta_k^{\text{optNA}}(\varepsilon_g) = \max_{t \in [0, \varepsilon]} \sum_{i=0}^{k} \binom{k}{i} p_t^{k-i} (1 - p_t)^i \max\left\{\left(\mathrm{e}^{kt - i\varepsilon} - \mathrm{e}^{\varepsilon_g}\right), 0\right\} \tag{4}$$

Proof. From Corollary C.1 we know that for $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$,

$$\delta_k^{\text{optNA}}(\varepsilon_g) = \max_{t \in [0, \varepsilon]} \delta(t, \ldots, t, \varepsilon_g).$$

Furthermore, we know if $\varepsilon_g \geqslant k\varepsilon$ then $\delta(\mathbf{t}, \varepsilon_g) = 0$ for any $\mathbf{t} \in [0, \varepsilon]^k$, and also if $\varepsilon_g \leqslant -k\varepsilon$ then $\delta(\mathbf{t}, \varepsilon_g) = 1 - \mathrm{e}^{\varepsilon_g}$ for any $\mathbf{t} \in [0, \varepsilon]^k$. Therefore,

$$\begin{aligned}
\delta_k^{\text{optNA}}(\varepsilon_g) &= \max_{t \in [0, \varepsilon]} \sum_{S \subseteq \{1, \ldots, k\}} \prod_{i \notin S} p_t \prod_{i \in S} (1 - p_t) \max\left\{\mathrm{e}^{kt - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\} \\
&= \max_{t \in [0, \varepsilon]} \sum_{S \subseteq \{1, \ldots, k\}} p_t^{k - |S|} (1 - p_t)^{|S|} \max\left\{\mathrm{e}^{kt - |S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\}
\end{aligned}$$

For each $i \in \{0, 1, \cdots, k\}$ there are $\binom{k}{i}$ subsets $S \subseteq \{1, \ldots, k\}$ such that $|S| = i$, and grouping these together gives our desired equality. □

C.2. Solving the one-dimensional optimization

Now that we have a much simpler one-dimensional optimization problem (4), it's possible to explicitly solve for the maximizer. Ultimately, we will show that there are only k different candidate values of t that maximizes $\delta((t, t, \cdots, t), \varepsilon_g)$, and give explicit expressions for these candidate values of t. These explicit expressions will also be necessary in later sections when we show that there is a difference between the adaptive and nonadaptive setting.

Since we no longer need to consider any $\mathbf{t} \in [0,\varepsilon]^k$ where $\mathbf{t}$ is not a scalar times the all ones vector, we will simplify our notation to be

$$\delta^k(t,\varepsilon_g) := \sum_{i=0}^{k} \binom{k}{i} p_t^{k-i}(1-p_t)^i \max\left\{\left(\mathrm{e}^{kt-i\varepsilon} - \mathrm{e}^{\varepsilon_g}\right), 0\right\}. \tag{5}$$

Given that we want to find the t which maximizes this expression, our goal will be to take the partial derivative of this function with respect to t. The maximization within the expression will make this more difficult, however, because the maximization is over a variable term and zero, we will always be able to write $\delta^k(t,\varepsilon_g)$ in terms of the following function F_ℓ for some $\ell \in \{0,\cdots,k\}$ that will depend on t.

$$F_\ell(t,\varepsilon_g) := \sum_{i=0}^{\ell} \binom{k}{i} p_t^{k-i}(1-p_t)^i \left(\mathrm{e}^{kt-i\varepsilon} - \mathrm{e}^{\varepsilon_g}\right). \tag{6}$$

This function is differentiable and we show its relation to $\delta^k(t,\varepsilon_g)$.

Lemma C.5. *For any $\varepsilon_g \in \mathbb{R}$, $\varepsilon \geqslant 0$, and $t \in [0,\varepsilon]$, there must exist some $\ell \in [k]$ such that*

$$\delta^k(t,\varepsilon_g) = F_\ell(t,\varepsilon_g).$$

Proof. Note that $\mathrm{e}^{kt-i\varepsilon} - \mathrm{e}^{\varepsilon_g}$ decreases as i increases, which implies that for any $t \in [0,\varepsilon]$ there must exist some ℓ such that $\max\{\mathrm{e}^{kt-i\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = \mathrm{e}^{kt-i\varepsilon} - \mathrm{e}^{\varepsilon_g}$ for all $i \leqslant \ell$ and $\max\{\mathrm{e}^{kt-i\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = 0$ for all $i > \ell$. Therefore, because p_t and $(1-p_t)$ are non-negative we have

$$\delta^k(t,\varepsilon_g) = F_\ell(t,\varepsilon_g).$$

□

It then follows that optimizing over $t \in [0,\varepsilon]$ for $\delta^k(t,\varepsilon_g)$ can be reduced to optimizing over $t \in [0,\varepsilon]$ for each $F_\ell(t,\varepsilon_g)$.

Corollary C.2. *For any $\varepsilon_g \in \mathbb{R}$ and $\varepsilon \geqslant 0$,*

$$\delta_k^{\text{optNA}}(\varepsilon_g) = \max_{0\leqslant\ell\leqslant k}\{\max_{t\in[0,\varepsilon]} F_\ell(t,\varepsilon_g)\}.$$

Proof. Follows immediately from Lemma C.5 and because for any ε_g and $t \in [0,\varepsilon]$, by definition $F_\ell(t,\varepsilon_g) \geqslant \delta^k(t,\varepsilon_g)$ for all ℓ.

□

We will now individually solve each $\max_{t\in[0,\varepsilon]} F_\ell(t,\varepsilon_g)$, which does not contain a maximization term and is differentiable. Our ultimate goal will be to solve $\frac{\partial F_\ell(t,\varepsilon_g)}{\partial t} = 0$, and we want explicit expressions for t, which will require a simple formulation of the partial derivate with respect to t. These explicit expressions will also be necessary for proving that there is a gap between the nonadaptive and adaptive settings. The proof for this will become quite involved with some surprisingly nice cancellation, and we relegate the details to Appendix C.4.

Lemma C.6. *For $\varepsilon_g \in \mathbb{R}$, $\varepsilon \geqslant 0$, and $0 \leqslant \ell \leqslant k$*

$$\frac{\partial F_\ell(t,\varepsilon_g)}{\partial t} = (k-\ell)\binom{k}{\ell} p_t^{k-1-\ell}(1-p_t)^\ell \frac{1}{1-\mathrm{e}^{-\varepsilon}}\left(\mathrm{e}^{\varepsilon_g - t} - \mathrm{e}^{kt-(\ell+1)\varepsilon}\right).$$

In order to prove that there is a gap between composition of adaptive and nonadaptive BR mechanisms, we will further utilize this exact characterization of the partial derivative to give a strict interpretation of the set of t that can achieve a maximization of our full expression. However, for giving an efficiently computable expression for optimal composition, the following simple corollary will suffice.

Corollary C.3. *For $\varepsilon_g \in \mathbb{R}$, $\varepsilon \geqslant 0$, and $0 \leqslant \ell \leqslant k$*

$$\arg\max_{t\in[0,\varepsilon]} F_\ell(t,\varepsilon_g) \in \left\{0, \varepsilon, \frac{\varepsilon_g + (\ell+1)\varepsilon}{k+1}\right\}.$$

Proof. Note that $p_t = 1$ when $t = 0$ and $p_t = 0$ when $t = \varepsilon$. Therefore $\frac{\partial F_\ell(t,\varepsilon_g)}{\partial t} = 0$ when $t \in \{0,\varepsilon\}$ or when $\varepsilon_g - t = kt - (\ell+1)\varepsilon$ which evaluates to $t = \frac{\varepsilon_g+(\ell+1)\varepsilon}{k+1}$. □

We can now prove Lemma B.1, which gives an efficient computation of optimal composition in the non-adaptive setting.

Proof. From Lemma C.4 we have

$$\delta_k^{\text{optNA}}(\varepsilon_g) = \max_{t\in[0,\varepsilon]} \delta^k(t,\varepsilon_g)$$

From Lemma C.5 and Corollary C.2 we can restrict our consideration to values of $t \in [0,\varepsilon]$ that maximize $F_\ell(t,\varepsilon_g)$ for some $\ell \in [k]$. Applying Corollary C.3 we can then restrict our consideration to t_ℓ for all $\ell \in [k]$, along with 0 and ε. Note that $p_t = 1$ when $t = 0$ and $p_t = 0$ when $t = \varepsilon$, so it is straightforward to verify that $\delta^k(0,\varepsilon_g) = \delta^k(\varepsilon,\varepsilon_g) = \max\{1 - \mathrm{e}^{\varepsilon_g}, 0\}$ for any ε_g. In the proof of Lemma C.2, we showed that $\delta_k^{\text{optNA}}(\varepsilon_g) > 0$ and $\delta_k^{\text{optNA}}(\varepsilon_g) > 1 - \mathrm{e}^{\varepsilon_g}$ when $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$, so it is irrelevant whether we include $0, \varepsilon$ in this setting. Finally, if $\varepsilon_g \notin (-k\varepsilon, k\varepsilon)$, then from Lemma C.1 we have $\delta^k(t,\varepsilon_g) = \delta_k^{\text{optNA}}(\varepsilon_g) = \max\{1 - \mathrm{e}^{\varepsilon_g}, 0\}$ for any t.

For the running time, first note that for any t we can compute $p_t^k(\mathrm{e}^{kt} - \mathrm{e}^{\varepsilon_g})$ in $O(k)$ time. Further, for any t, if we are given the values $\binom{k}{i}p_t^{k-i}(1-p_t)^i$ and $\mathrm{e}^{kt-i\varepsilon}$, then we can compute $\binom{k}{i+1}p_t^{k-(i+1)}(1-p_t)^{i+1}$ and $\mathrm{e}^{kt-(i+1)\varepsilon}$ in $O(1)$ time. Our running time of $O(k^2)$ then immediately follows. □

C.3. Proof of Lemma C.3

This lemma will be proven in two main sublemmas. First, we show that it holds for $k = 2$, then we show how we can reduce the general case to $k = 2$ by conditioning outcomes other than the first and second terms.

Lemma C.7. *For any $\varepsilon_g \in \mathbb{R}$ and $t_1, t_2 \in [0,\varepsilon]$*

$$\delta((t_1,t_2),\varepsilon_g) \leqslant \delta\left(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}\right), \varepsilon_g\right)$$

Further, the inequality is strict whenever $\varepsilon_g < t_1 + t_2 < \varepsilon_g + 2\varepsilon$ and $t_1 \neq t_2$.

Proof. Using the fact that $q_t = \mathrm{e}^t p_t$ and $(1-q_t) = \mathrm{e}^{t-\varepsilon}(1-p_t)$, we rewrite

$$\delta((t_1,t_2),\varepsilon_g) = \sum_{S\subseteq\{1,2\}} \prod_{i\notin S} p_{t_i} \prod_{i\in S}(1-p_{t_i}) \max\left\{\mathrm{e}^{t_1+t_2-|S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\}$$

We will then prove our desired inequality by considering four cases.

Case I ($t_1 + t_2 \leqslant \varepsilon_g$): This implies that $\max\{\mathrm{e}^{t_1+t_2-|S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = 0$ for any subset S and

$$\delta((t_1,t_2),\varepsilon_g) = \delta\left(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}\right), \varepsilon_g\right) = 0.$$

Case II ($t_1 + t_2 \geqslant \varepsilon_g + 2\varepsilon$): This implies $\max\{\mathrm{e}^{t_1+t_2-|S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = \mathrm{e}^{t_1+t_2-|S|\varepsilon} - \mathrm{e}^{\varepsilon_g}$ for any S, which gives

$$\delta((t_1,t_2),\varepsilon_g) = \sum_{S\subseteq\{1,2\}} \left(\prod_{i\notin S} q_{t_i} \prod_{i\in S}(1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i\notin S} p_{t_i} \prod_{i\in S}(1-p_{t_i})\right) = 1 - \mathrm{e}^{\varepsilon_g}$$

and equivalently holds for $\delta(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}\right), \varepsilon_g)$.

Case III ($\varepsilon_g < t_1 + t_2 \leqslant \varepsilon_g + \varepsilon$): This implies that $\max\{\mathrm{e}^{t_1+t_2-|S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = 0$ for any S such that $|S| > 0$. Therefore,

$$\delta((t_1, t_2), \varepsilon_g) = p_{t_1} p_{t_2} \left(\mathrm{e}^{t_1+t_2} - \mathrm{e}^{\varepsilon_g}\right)$$

Equivalently, we have

$$\delta\left(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}\right), \varepsilon_g\right) = p^2_{\frac{t_1+t_2}{2}} \left(\mathrm{e}^{t_1+t_2} - \mathrm{e}^{\varepsilon_g}\right)$$

We want strict inequality for this case, so it suffices to show $p_{t_1} p_{t_2} < p^2_{\frac{t_1+t_2}{2}}$. Plugging in the explicit formula for each p_t and performing some simple algebraic manipulations gives that this is equivalent to

$$2\mathrm{e}^{-\frac{t_1+t_2}{2}} < \mathrm{e}^{-t_1} + \mathrm{e}^{-t_2}$$

which holds due to the strict-convexity of the exponential function.

Case IV ($\varepsilon_g + \varepsilon \leqslant t_1 + t_2 < \varepsilon_g + 2\varepsilon$): This implies that $\max\{\mathrm{e}^{t_1+t_2-|S|\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\} = 0$ when $|S| = 2$. Therefore,

$$\delta((t_1, t_2), \varepsilon_g) = p_{t_1} p_{t_2} \left(\mathrm{e}^{t_1+t_2} - \mathrm{e}^{\varepsilon_g}\right) + \left(p_{t_1}(1-p_{t_2}) + p_{t_2}(1-p_{t_1})\right) \left(\mathrm{e}^{t_1+t_2-\varepsilon} - \mathrm{e}^{\varepsilon_g}\right)$$

From Case II, we know

$$\sum_{S \subseteq \{1,2\}} \prod_{i \notin S} p_{t_i} \prod_{i \in S} (1-p_{t_i}) \left(\mathrm{e}^{t_1+t_2-|S|\varepsilon} - \mathrm{e}^{\varepsilon_g}\right) = 1 - \mathrm{e}^{\varepsilon_g}$$

which yields

$$\delta((t_1, t_2), \varepsilon_g) = 1 - \mathrm{e}^{\varepsilon_g} - (1-p_{t_1})(1-p_{t_2}) \left(\mathrm{e}^{t_1+t_2-2\varepsilon} - \mathrm{e}^{\varepsilon_g}\right).$$

This equivalently holds for $\delta(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}\right), \varepsilon_g)$ and because $\mathrm{e}^{t_1+t_2-2\varepsilon} - \mathrm{e}^{\varepsilon_g} < 0$, we have

$$\delta((t_1, t_2), \varepsilon_g) < \delta\left(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}\right), \varepsilon_g\right) \qquad \Leftrightarrow \qquad (1-p_{t_1})(1-p_{t_2}) < \left(1 - p_{\frac{t_1+t_2}{2}}\right)^2.$$

Once again, we plug in the explicit formula for each p_t and perform some simple algebraic manipulations to see that this is also equivalent to

$$2\mathrm{e}^{-\frac{t_1+t_2}{2}} < \mathrm{e}^{-t_1} + \mathrm{e}^{-t_2}$$

and this again holds due to the strict-convexity of the exponential function.

□

We now want to extend this to $k > 2$, which will be done by fixing an arbitrary subset of $\{3, \cdots, k\}$ and show that the inequality holds when we restrict the summation to subsets of $\{1, \cdots, k\}$ that must contain that subset of $\{3, \cdots, k\}$. This will allow for easy cancellation. We will denote $\delta_U(\mathbf{t}, \varepsilon_g, S)$ for a set $U \subseteq [k]$ and $S \subseteq U$ as

$$\delta_U(\mathbf{t}, \varepsilon_g, S) := \prod_{i \in U \setminus S} p_{t_i} \prod_{i \in S} (1-p_{t_i}) \cdot \sum_{S' \subseteq [k] \setminus U} \max\left\{\mathrm{e}^{\sum_{j \in U} t_j - |S|\varepsilon} \prod_{i \notin U \cup S'} q_{t_i} \prod_{i \in S'} (1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i \notin U \cup S'} p_{t_i} \prod_{i \in S'} (1-p_{t_i}), 0\right\}.$$

Claim C.1. *Let $\varepsilon_g \in \mathbb{R}$. Then for any $\mathbf{t} \in [0, \varepsilon]^k$, we have for $U = \{3, \cdots, k\}$*

$$\delta(\mathbf{t}, \varepsilon_g) = \sum_{S \subseteq U} \delta_U(\mathbf{t}, \varepsilon_g, S)$$

Proof. We fix a set $S \subseteq \{3, \cdots, k\} = U$. Using the fact that $q_t = \mathrm{e}^t p_t$ and $(1-q_t) = \mathrm{e}^{t-\varepsilon}(1-p_t)$, we have

$$\prod_{i \in U \setminus S} q_{t_i} \prod_{i \in S} (1-q_{t_i}) = \mathrm{e}^{t_3 + \cdots + t_k - |S|\varepsilon} \prod_{i \in U \setminus S} p_{t_i} \prod_{i \in S} (1-p_{t_i})$$

Therefore, we also have

$$\delta_U(\mathbf{t}, \varepsilon_g, S) = \sum_{S' \subseteq \{1,2\}} \max\left\{ \prod_{i \notin S' \cup S} q_{t_i} \prod_{i \in S' \cup S} (1-q_{t_i}) - \mathrm{e}^{\varepsilon_g} \prod_{i \notin S' \cup S} p_{t_i} \prod_{i \in S' \cup S} (1-p_{t_i}), 0 \right\}$$

Summing over all S we can simply rewrite this summation over all subsets of $\{1, \cdots, k\}$, giving our desired equality. □

Lemma C.8. *For any $S \subseteq \{3, ..., k\} = U$, we have the following inequality*

$$\delta_U(\mathbf{t}, \varepsilon_g, S) \leqslant \delta_U\left(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}, t_3, ..., t_k\right), \varepsilon_g, S\right)$$

Further, the inequality is strict if $\varepsilon_g < \sum_{i=1}^k t_i - |S|\varepsilon < \varepsilon_g + 2\varepsilon$ and $t_1 \neq t_2$.

Proof. We fix $S \subseteq \{3, \cdots, k\}$. Let $\varepsilon_g' = \varepsilon_g + |S| - t_3 - \cdots - t_k$, and then by cancelling non-negative like terms it suffices to show

$$\sum_{S' \subseteq \{1,2\}} \max\left\{ \prod_{i \notin S'} q_{t_i} \prod_{i \in S'} (1-q_{t_i}) - \mathrm{e}^{\varepsilon_g'} \prod_{i \notin S'} p_{t_i} \prod_{i \in S'} (1-p_{t_i}), 0 \right\}$$
$$\leqslant \sum_{S' \subseteq \{1,2\}} \max\left\{ \prod_{i \notin S'} q_{t'} \prod_{i \in S'} (1-q_{t'}) - \mathrm{e}^{\varepsilon_g'} \prod_{i \notin S'} p_{t'} \prod_{i \in S'} (1-p_{t'}), 0 \right\}$$

where $t' = \frac{t_1+t_2}{2}$. By definition, this is then equivalent to showing

$$\delta((t_1, t_2), \varepsilon_g') \leqslant \delta\left(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}\right), \varepsilon_g'\right)$$

which follows from Lemma C.7, and the strictness follows from the fact that $\varepsilon_g' = \varepsilon_g + |S| - \sum_{j>2} t_j$. □

With these we can now prove our main convexity lemma.

Proof of Lemma C.3. It immediately follows from Claim C.1 and Lemma C.8 that for any $\mathbf{t} \in [0, \varepsilon]^k$

$$\delta(\mathbf{t}, \varepsilon_g) \leqslant \delta\left(\left(\frac{t_1+t_2}{2}, \frac{t_1+t_2}{2}, t_3, ..., t_k\right), \varepsilon_g\right)$$

Additionally, if we assume that $t_1 \neq t_2$ and $\varepsilon_g < \sum t_i < \varepsilon_g + k\varepsilon$, then there must exist some $\ell \in [0, k-2]$ such that $\varepsilon_g + \ell\varepsilon < \sum t_i < \varepsilon_g + (\ell+2)\varepsilon$, which implies that $\varepsilon_g < \sum t_i - \ell\varepsilon < \varepsilon_g + 2\varepsilon$. Further, we know that for any $\ell \in [0, k-2]$ there exists $S \subseteq \{3, \cdots, k\}$ such that $|S| = \ell$. Therefore, for one of these subsets the inequality is strict and the sum must be a strict inequality as well. □

C.4. Proof of Lemma C.6

Recall that we had the following definition, for which we wanted to compute the partial derivate with respect to t.

$$F_\ell(t, \varepsilon_g) := \sum_{i=0}^{\ell} \binom{k}{i} p_t^{k-i} (1-p_t)^i \left(\mathrm{e}^{kt - i\varepsilon} - \mathrm{e}^{\varepsilon_g}\right) \tag{7}$$

We further split each $F_\ell(t, \varepsilon_g)$ into the individual terms to more easily differentiate the full summation with respect to t.

$$f_\ell(t, \varepsilon_g) := \binom{k}{\ell} p_t^{k-\ell} (1 - p_t)^\ell \left(\mathrm{e}^{kt - \ell\varepsilon} - \mathrm{e}^{\varepsilon_g} \right)$$

In particular, giving a much simpler formulation for the partial derivative will rely upon an inductive proof, so this definition will allow an even easier comparison between $F_\ell(t, \varepsilon_g)$ and $F_{\ell+1}(t, \varepsilon_g)$ that follows immediately from the definition.

Corollary C.4. *For any* $\ell \in [1, k]$

$$F_\ell(t, \varepsilon_g) = F_{\ell-1}(t, \varepsilon_g) + f_\ell(t, \varepsilon_g)$$

We first differentiate the simplest of these expressions $F_0(t, \varepsilon_g)$, and then we will ultimately use this as the base case for proving a simplified formulation of derivative for the general case.

Lemma C.9.

$$\frac{\partial F_0(t, \varepsilon_g)}{\partial t} = k p_t^{k-1} \frac{1}{1 - \mathrm{e}^{-\varepsilon}} \left(\mathrm{e}^{\varepsilon_g - t} - \mathrm{e}^{kt - \varepsilon} \right)$$

Proof. By definition

$$F_0(t, \varepsilon_g) = p_t^k \left(\mathrm{e}^{kt} - \mathrm{e}^{\varepsilon_g} \right) = \left(\frac{\mathrm{e}^{-t} - \mathrm{e}^{-\varepsilon}}{1 - \mathrm{e}^{-\varepsilon}} \right)^k \left(\mathrm{e}^{kt} - \mathrm{e}^{\varepsilon_g} \right)$$

Therefore, by basic differentiation rules

$$\begin{aligned} \frac{\partial F_0(t, \varepsilon_g)}{\partial t} &= \left(-k \frac{\mathrm{e}^{-t}}{1 - \mathrm{e}^{-\varepsilon}} \left(\frac{\mathrm{e}^{-t} - \mathrm{e}^{-\varepsilon}}{1 - \mathrm{e}^{-\varepsilon}} \right)^{k-1} \left(\mathrm{e}^{kt} - \mathrm{e}^{\varepsilon_g} \right) \right) + \left(\frac{\mathrm{e}^{-t} - \mathrm{e}^{-\varepsilon}}{1 - \mathrm{e}^{-\varepsilon}} \right)^k k \mathrm{e}^{kt} \\ &= k p_t^{k-1} \frac{1}{1 - \mathrm{e}^{-\varepsilon}} \left(-\mathrm{e}^{-t} \left(\mathrm{e}^{kt} - \mathrm{e}^{\varepsilon_g} \right) + \left(\mathrm{e}^{-t} - \mathrm{e}^{\varepsilon} \right) \mathrm{e}^{kt} \right) \end{aligned}$$

which easily reduces to our desired term.

□

To apply an inductive claim to the general case, we will also need to evaluate the partial derivative of the last term for each sum.

Lemma C.10. *For* $1 \leqslant \ell \leqslant k$

$$\begin{aligned} \frac{\partial f_\ell(t, \varepsilon_g)}{\partial t} = \binom{k}{\ell} p_t^{k-1-\ell} (1 - p_t)^{\ell-1} \left(\frac{1}{1 - \mathrm{e}^{-\varepsilon}} \right)^2 \Big(&(k - \ell) \left(\mathrm{e}^{\varepsilon_g - t} + \mathrm{e}^{(k-1)t - (\ell+1)\varepsilon} \right) \\ &+ \ell \left(\mathrm{e}^{(k-1)t - \ell\varepsilon} + \mathrm{e}^{\varepsilon_g - \varepsilon - t} \right) - k \left(\mathrm{e}^{\varepsilon_g - 2t} + \mathrm{e}^{kt - (\ell+1)\varepsilon} \right) \Big) \end{aligned}$$

Proof. By definition

$$f_\ell(t, \varepsilon_g) = \binom{k}{\ell} p_t^{k-\ell} (1 - p_t)^\ell \left(\mathrm{e}^{kt - \ell\varepsilon} - \mathrm{e}^{\varepsilon_g} \right)$$

We can consider this then to instead be $f_\ell(t, \varepsilon_g) = \binom{k}{\ell} f(t) \cdot g(t) \cdot h(t)$ with $f(t) = p_t^{k-\ell}$, $g(t) = (1 - p_t)^\ell$, and $h(t) = \mathrm{e}^{kt - \ell\varepsilon} - \mathrm{e}^{\varepsilon_g}$. Applying basic differentiation rules and using the fact that $p_t = \frac{\mathrm{e}^{-t} - \mathrm{e}^{-\varepsilon}}{1 - \mathrm{e}^{-\varepsilon}}$, we obtain

$$\frac{\partial f_\ell(t,\varepsilon_g)}{\partial t} = \binom{k}{\ell}(k-\ell)\left(\frac{-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\varepsilon}}\right)p_t^{k-1-\ell}(1-p_t)^\ell\left(\mathrm{e}^{kt-\ell\varepsilon}-\mathrm{e}^{\varepsilon_g}\right) + \binom{k}{\ell}\ell\left(\frac{\mathrm{e}^{-t}}{1-\mathrm{e}^{-\varepsilon}}\right)p_t^{k-\ell}(1-p_t)^{\ell-1}\left(\mathrm{e}^{kt-\ell\varepsilon}-\mathrm{e}^{\varepsilon_g}\right) + \binom{k}{\ell}k\mathrm{e}^{kt-\ell\varepsilon}p_t^{k-\ell}(1-p_t)^\ell$$

We can pull out similar terms from each expression to achieve

$$\frac{\partial f_\ell(t,\varepsilon_g)}{\partial t} = \binom{k}{\ell}p_t^{k-1-\ell}(1-p_t)^{\ell-1}\left(\frac{1}{1-\mathrm{e}^{-\varepsilon}}\right)^2\Big(-(k-\ell)\mathrm{e}^{-t}(1-\mathrm{e}^{-t})\left(\mathrm{e}^{kt-\ell\varepsilon}-\mathrm{e}^{\varepsilon_g}\right) + \ell\mathrm{e}^{-t}(\mathrm{e}^{-t}-\mathrm{e}^{-\varepsilon})\left(\mathrm{e}^{kt-\ell\varepsilon}-\mathrm{e}^{\varepsilon_g}\right) + k\mathrm{e}^{kt-\ell\varepsilon}\left(\mathrm{e}^{-t}-\mathrm{e}^{-\varepsilon}\right)\left(1-\mathrm{e}^{-t}\right)\Big)$$

Further examination of the inner term by expanding each expression and cancelling like terms gives

$$\begin{aligned}&-(k-\ell)\mathrm{e}^{-t}(1-\mathrm{e}^{-t})\left(\mathrm{e}^{kt-\ell\varepsilon}-\mathrm{e}^{\varepsilon_g}\right) + \ell\mathrm{e}^{-t}(\mathrm{e}^{-t}-\mathrm{e}^{-\varepsilon})\left(\mathrm{e}^{kt-\ell\varepsilon}-\mathrm{e}^{\varepsilon_g}\right) + k\mathrm{e}^{kt-\ell\varepsilon}\left(\mathrm{e}^{-t}-\mathrm{e}^{-\varepsilon}\right)\left(1-\mathrm{e}^{-t}\right)\\ &\quad= (k-\ell)\left(\mathrm{e}^{\varepsilon_g-t}+\mathrm{e}^{(k-1)t-(\ell+1)\varepsilon}\right) + \ell\left(\mathrm{e}^{(k-1)t-\ell\varepsilon}+\mathrm{e}^{\varepsilon_g-\varepsilon-t}\right) - k\left(\mathrm{e}^{\varepsilon_g-2t}+\mathrm{e}^{kt-(\ell+1)\varepsilon}\right)\end{aligned}$$

This then implies our desired expression.

□

We now have the pieces to give a simpler evaluation of the partial derivative for the general case using an inductive argument. Surprisingly, with a bit of combinatorial and algebraic massaging, the full partial derivative will reduce to a rather simple expression.

Proof of Lemma C.6. The base case of $\ell = 0$ is true from Lemma C.9. We then assume the claim for $\ell - 1$, and by Corollary C.4 we know $F_\ell(t,\varepsilon_g) = F_{\ell-1}(t,\varepsilon_g) + f_\ell(t,\varepsilon_g)$, which implies

$$\frac{\partial F_\ell(t,\varepsilon_g)}{\partial t} = \frac{\partial F_{\ell-1}(t,\varepsilon_g)}{\partial t} + \frac{\partial f_\ell(t,\varepsilon_g)}{\partial t}$$

Applying our inductive claim and Lemma C.10 we then have

$$\begin{aligned}\frac{\partial F_\ell(t,\varepsilon_g)}{\partial t} = (k-(\ell-1))\binom{k}{\ell-1}p_t^{k-1-(\ell-1)}(1-p_t)^{\ell-1}\frac{1}{1-\mathrm{e}^{-\varepsilon}}\left(\mathrm{e}^{\varepsilon_g-t}-\mathrm{e}^{kt-\ell\varepsilon}\right) +\\ \binom{k}{\ell}p_t^{k-1-\ell}(1-p_t)^{\ell-1}\left(\frac{1}{1-\mathrm{e}^{-\varepsilon}}\right)^2\Big((k-\ell)\left(\mathrm{e}^{\varepsilon_g-t}+\mathrm{e}^{(k-1)t-(\ell+1)\varepsilon}\right)\\ +\ell\left(\mathrm{e}^{(k-1)t-\ell\varepsilon}+\mathrm{e}^{\varepsilon_g-\varepsilon-t}\right) - k\left(\mathrm{e}^{\varepsilon_g-2t}+\mathrm{e}^{kt-(\ell+1)\varepsilon}\right)\Big)\end{aligned}$$

We use the fact that $(k-(\ell-1))\binom{k}{\ell-1} = \ell\binom{k}{\ell}$ and this reduces to

$$\begin{aligned}\frac{\partial F_\ell(t,\varepsilon_g)}{\partial t} = \binom{k}{\ell}p_t^{k-1-\ell}(1-p_t)^{\ell-1}\left(\frac{1}{1-\mathrm{e}^{-\varepsilon}}\right)^2\Big(\ell\left(\mathrm{e}^{-t}-\mathrm{e}^{-\varepsilon}\right)\left(\mathrm{e}^{\varepsilon_g-t}-\mathrm{e}^{kt-\ell\varepsilon}\right) +\\ (k-\ell)\left(\mathrm{e}^{\varepsilon_g-t}+\mathrm{e}^{(k-1)t-(\ell+1)\varepsilon}\right) + \ell\left(\mathrm{e}^{(k-1)t-\ell\varepsilon}+\mathrm{e}^{\varepsilon_g-\varepsilon-t}\right) - k\left(\mathrm{e}^{\varepsilon_g-2t}+\mathrm{e}^{kt-(\ell+1)\varepsilon}\right)\Big)\end{aligned}$$

Further examination of the inner term by expanding each expression and cancelling like terms gives

$$\begin{aligned}\ell\left(\mathrm{e}^{-t}-\mathrm{e}^{-\varepsilon}\right)\left(\mathrm{e}^{\varepsilon_g-t}-\mathrm{e}^{kt-\ell\varepsilon}\right)+(k-\ell)\left(\mathrm{e}^{\varepsilon_g-t}+\mathrm{e}^{(k-1)t-(\ell+1)\varepsilon}\right)&\\ +\ell\left(\mathrm{e}^{(k-1)t-\ell\varepsilon}+\mathrm{e}^{\varepsilon_g-\varepsilon-t}\right)-k\left(\mathrm{e}^{\varepsilon_g-2t}+\mathrm{e}^{kt-(\ell+1)\varepsilon}\right)&\\ =(k-\ell)\left(\mathrm{e}^{\varepsilon_g-t}-\mathrm{e}^{\varepsilon_g-2t}+\mathrm{e}^{(k-1)t-(\ell+1)\varepsilon}-\mathrm{e}^{kt-(\ell+1)\varepsilon}\right)&\\ =(k-\ell)(1-\mathrm{e}^{-t})\left(\mathrm{e}^{\varepsilon_g-t}-\mathrm{e}^{kt-(\ell+1)\varepsilon}\right)&\end{aligned}$$

Substituting for this simplified expression and using the fact that $1-p_t=\frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\varepsilon}}$ then gives our desired result. □

D. Proof of Theorem 4

The goal of this section is to prove Theorem 4, which states that there is a gap between adaptive and non-adaptive compositions of BR mechanisms. In fact, the gap exists for all $k\geqslant 2$ and most ε_g. By optimality statements of Theorem 2 and 3, it suffices to show that

$$\delta_k^{\mathrm{NA}}(\varepsilon_g)<\delta_k^{\mathrm{A}}(\varepsilon_g)$$

for the claimed parameter regime. The argument still requires a lot of careful work. We will execute main steps in Appendix D.1 and then provide missing proofs in Appendix D.2.

D.1. Main steps

The general idea for proving the gap will be to also give the recursive definition for the non-adaptive optimal composition that must fix t for each recursive call. The goal will then be to show that at some point within this recursion the summation will strictly increase if the value for t is changed. This will require that we first fully characterize the possible values of t for the non-adaptive optimal composition. Fortunately, most of the heavy lifting in this regard was done in the previous section. With this characterization, we show that there is a gap when $k=2$, and then further show that we can apply this gap for $k\geqslant 2$.

Notations are inherited from the previous section. Recall $\delta(\mathbf{t},\varepsilon_g)$ is the optimization objective in non-adaptive composition, i.e. $\delta_k^{\mathrm{NA}}(\varepsilon_g)=\max_{\mathbf{t}\in[0,\varepsilon]^k}\delta(\mathbf{t},\varepsilon_g)$. From Corollary C.1, we know that any maximizer must look like $(t,t,\ldots,t)$ for some $t\in[0,\varepsilon]$. So we introduced $\delta^k(t,\varepsilon_g)=\delta\big((t,t,\ldots,t),\varepsilon_g\big)$. Now we have $\delta_k^{\mathrm{NA}}(\varepsilon_g)=\max_{t\in[0,\varepsilon]}\delta^k(t,\varepsilon_g)$. We are interested in all possible maximizers, i.e. the set

$$t_{\mathrm{opt}}(k,\varepsilon_g)=\{t\in[0,\varepsilon]:\delta^k(t,\varepsilon_g)=\delta_k^{\mathrm{NA}}(\varepsilon_g)\}$$

Lemma D.1. *Let $\varepsilon\geqslant 0$. If $\varepsilon_g\in(-k\varepsilon,k\varepsilon)$, then*

$$t_{\mathrm{opt}}(k,\varepsilon_g)\subseteq\left\{\frac{\varepsilon_g+(\ell+1)\varepsilon}{k+1}:\ell\in\{0,\cdots,k-1\}\right\}\cap(0,\varepsilon).$$

Proof. From Lemma C.5 and Corollary C.2 we can restrict our consideration to values of $t\in[0,\varepsilon]$ that maximize $F_\ell(t,\varepsilon_g)$ for some $\ell\in[k]$. Furthermore, $F_\ell(t,\varepsilon_g)$ can only maximized at the endpoints of the interval or whenever $\frac{\partial F_\ell(t,\varepsilon_g)}{\partial t}=0$. Thus, from Corollary C.3 we have

$$t_{\mathrm{opt}}(k,\varepsilon_g)\subseteq\Big\{\underbrace{\frac{\varepsilon_g+(\ell+1)\varepsilon}{k+1}}_{t_\ell^*}:\ell\in\{0,k\}\Big\}\cup\{0,\varepsilon\}.$$

By definition, we can remove all values outside of $[0,\varepsilon]$, so it then suffices to show that we can also remove $\{0,\varepsilon,t_k^*\}$. Note that $p_t=1$ when $t=0$ and $p_t=0$ when $t=\varepsilon$ and recall $\delta^k(t,\varepsilon_g)$ from (5), so it is straightforward to verify

that $\delta^k(0,\varepsilon_g) = \delta^k(\varepsilon,\varepsilon_g) = \max\{1-\mathrm{e}^{\varepsilon_g}, 0\}$ for any ε_g. In the proof of Lemma C.2, we showed that $\delta_k^{\mathrm{NA}}(\varepsilon_g) > 0$ and $\delta_k^{\mathrm{NA}}(\varepsilon_g) > 1-\mathrm{e}^{\varepsilon_g}$ when $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$, which implies $0, \varepsilon \notin t_{\mathrm{opt}}(k,\varepsilon_g)$.

It then suffices to show $t_k^* \notin t_{\mathrm{opt}}(k,\varepsilon_g)$. If $\varepsilon_g > 0$, then $t_k^* > \varepsilon$, so we only need to consider $\varepsilon_g \leqslant 0$. Note that $kt_k^* = k(\frac{\varepsilon_g}{k+1} + \varepsilon)$, so for any $i \leqslant k$ we have $kt_k^* - i\varepsilon \geqslant \frac{k}{k+1}\varepsilon_g$ which implies

$$\max\left\{\mathrm{e}^{kt_k^* - i\varepsilon} - \mathrm{e}^{\varepsilon_g}, 0\right\} = \mathrm{e}^{kt_k^* - i\varepsilon} - \mathrm{e}^{\varepsilon_g}.$$

Therefore, $\delta^k(t_k^*, \varepsilon_g) = 1 - \mathrm{e}^{\varepsilon_g}$ and from above we know $\delta_k^{\mathrm{NA}}(\varepsilon_g) > 1 - \mathrm{e}^{\varepsilon_g}$ when $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$, which implies $t_k^* \notin t_{\mathrm{opt}}(k,\varepsilon_g)$ as desired. □

We now want to show that we can write the optimal non-adaptive composition in a similar form as the adaptive composition. This recursive formulation will then fix a value t throughout the recursion and $\delta_k^{\mathrm{NA}}(\varepsilon_g)$ is then just the maximum value of this recursion over all $t \in [0,\varepsilon]$.

Corollary D.1. *For $k \geqslant 1$ and for $\delta^k(t,\varepsilon_g)$ from (5), we have $\delta^0(t,\varepsilon_g) = \max\{1-\mathrm{e}^{\varepsilon_g}, 0\}$ and*

$$\delta^k(t,\varepsilon_g) = q_t\delta^{k-1}(t, \varepsilon_g - t) + (1-q_t)\delta^{k-1}(t, \varepsilon_g + \varepsilon - t).$$

We relegate the proof of this corollary to Appendix D.2. Now that the formulations are similar, we show the intuitive fact that if at any point in the recursion either it is the case that either 1) switching the value of t, or 2) switching to the adaptive setting, will strictly increase that δ_{opt} then there must be a gap between the non-adaptive and adaptive setting.

Lemma D.2. *Fix the individual privacy parameter $\varepsilon > 0$, some global privacy parameter $\varepsilon_g \in \mathbb{R}$ and $k \geqslant 2$, along with some $t \in t_{\mathrm{opt}}(k,\varepsilon_g)$, if there exist $0 \leqslant \ell' \leqslant \ell < k$ such that either $\delta_{k-\ell}^{\mathrm{NA}}(\varepsilon_g - \ell t + \ell'\varepsilon) < \delta_{k-\ell}^{\mathrm{A}}(\varepsilon_g - \ell t + \ell'\varepsilon)$ or $t \notin t_{\mathrm{opt}}(k-\ell, \varepsilon_g - \ell t + \ell'\varepsilon)$, then we must have*

$$\delta_k^{\mathrm{NA}}(\varepsilon_g) < \delta_k^{\mathrm{A}}(\varepsilon_g).$$

This lemma will actually require quite a bit of technical detail, so we instead give a proof in Appendix D.2. With this property and our characterization of $t_{\mathrm{opt}}(k,\varepsilon_g)$, we now show that there is a gap for the base case of $k = 2$.

Lemma D.3. *For any $\varepsilon_g \in (-\varepsilon/2, \varepsilon/2)$ we have*

$$\delta_2^{\mathrm{NA}}(\varepsilon_g) < \delta_2^{\mathrm{A}}(\varepsilon_g)$$

Proof. From Lemma D.1, we know that there exists and $\ell \in \{0,1\}$ such that $t_\ell = \frac{\varepsilon_g + (\ell+1)\varepsilon}{3} \in t_{\mathrm{opt}}(2,\varepsilon_g)$. Furthermore, if both $\varepsilon_g - t_\ell$ and $\varepsilon_g - t_\ell + \varepsilon$ are in $(-\varepsilon,\varepsilon)$, then we also must have $t_{\mathrm{opt}}(1, \varepsilon_g - t_\ell) = \frac{\varepsilon_g - t_\ell + \varepsilon}{2}$ and $t_{\mathrm{opt}}(1, \varepsilon_g - t_\ell + \varepsilon) = \frac{\varepsilon_g - t_\ell + 2\varepsilon}{2}$ which implies $t_{\mathrm{opt}}(1, \varepsilon_g - t_\ell) \neq t_{\mathrm{opt}}(1, \varepsilon_g - t_\ell + \varepsilon)$.

Therefore, by Lemma D.2 it suffices to show that both $\varepsilon_g - t_\ell$ and $\varepsilon_g - t_\ell + \varepsilon$ are in $(-\varepsilon,\varepsilon)$, which is equivalent to showing $\varepsilon_g - t_\ell \in (-\varepsilon, 0)$. Plugging in for t_ℓ we then have

$$\varepsilon_g - \frac{\varepsilon_g + (\ell+1)\varepsilon}{3} \in (-\varepsilon, 0) \qquad \Leftrightarrow \qquad \varepsilon_g \in \left(\frac{(\ell-2)\varepsilon}{2}, \frac{(\ell+1)\varepsilon}{2}\right)$$

which holds for $\ell \in \{0,1\}$ by our assumption that $\varepsilon_g \in (-\varepsilon/2, \varepsilon/2)$. □

We will then apply this base case to the more general case for certain conditions by applying Lemma D.2.

Lemma D.4. *Given some $\varepsilon_g \in (-k\varepsilon, k\varepsilon)$ and $t \in t_{\mathrm{opt}}(k,\varepsilon_g)$. For $k \geqslant 4$, if $\varepsilon_g - (k-2)t < \varepsilon/2$ and $\varepsilon_g - (k-2)t + (k-2)\varepsilon > -\varepsilon/2$, then*

$$\delta_k^{\mathrm{NA}}(\varepsilon_g) < \delta_k^{\mathrm{A}}(\varepsilon_g).$$

We relegate the proof of this lemma to Appendix D.2 and will use this to show our desired result.

Lemma D.5. *For any $\varepsilon_g \in [-(k-3)\varepsilon, (k-3)\varepsilon]$ and $k \geqslant 4$ we have*

$$\delta_k^{\mathrm{NA}}(\varepsilon_g) < \delta_k^{\mathrm{A}}(\varepsilon_g).$$

Proof. We will prove for $\varepsilon_g \in [0, (k-3)\varepsilon]$ and the case of $\varepsilon_g \in [-(k-3)\varepsilon, 0]$ follows symmetrically. From Lemma D.1 we know that for any $t \in t_{\mathrm{opt}}(k, \varepsilon_g)$ we must have $t = \frac{\varepsilon_g + (\ell+1)\varepsilon}{k+1}$ for some $0 \leqslant \ell \leqslant k-1$. The general idea will then be to show that for any $t_\ell = \frac{\varepsilon_g + (\ell+1)\varepsilon}{k+1}$, if $t_\ell \in t_{\mathrm{opt}}(k, \varepsilon_g)$, then $\delta_k^{\mathrm{A}}(\varepsilon_g) > \delta_k^{\mathrm{NA}}(\varepsilon_g)$. We will split this into three cases.

Case I: ($\ell \geqslant 2$) For this setting, we want to show that we can apply Lemma D.4 where we know $\varepsilon_g + (k-2)(\varepsilon - t) \geqslant -\varepsilon/2$ for any t because we are assuming $\varepsilon_g \geqslant 0$. It then suffices to show that $\varepsilon_g - (k-2)t_\ell < \varepsilon/2$. Plugging in for t_ℓ we have

$$\varepsilon_g - (k-2)t_\ell < \varepsilon/2 \qquad \Leftrightarrow \qquad 6\varepsilon_g < (2(k-2)(\ell+1) + (k+1))\,\varepsilon.$$

By assumption, we know $\varepsilon_g \leqslant (k-3)\varepsilon$, so for $\ell \geqslant 2$, we have

$$6\varepsilon_g \leqslant 6(k-3)\varepsilon < (7k-11)\varepsilon \leqslant (2(k-2)(\ell+1) + (k+1))\,\varepsilon.$$

and therefore $\delta_k^{\mathrm{A}}(\varepsilon_g) > \delta_k^{\mathrm{NA}}(\varepsilon_g)$ by Lemma D.4.

Case II: ($\ell = 0$) For this setting we have $t_0 = \frac{\varepsilon_g + \varepsilon}{k+1}$. By our assumption that $\varepsilon_g \in [-(k-3)\varepsilon, (k-3)\varepsilon]$, we must have $\varepsilon_g + \varepsilon - t_0 \in (-(k-1)\varepsilon, (k-1)\varepsilon)$. From Lemma D.1 we then know $t_{\mathrm{opt}}(k-1, \varepsilon_g + \varepsilon - t_0) \subseteq \left\{ \frac{\varepsilon_g + \varepsilon - t_0 + (\ell'+1)\varepsilon}{k} : \ell' \in \{0, k-2\} \right\}$. We further see that for any $\ell' \geqslant 0$,

$$\frac{\varepsilon_g + \varepsilon}{k+1} < \frac{\varepsilon_g + \varepsilon}{k} \leqslant \frac{\varepsilon_g + \varepsilon - t_0 + (\ell'+1)\varepsilon}{k}.$$

This implies $t_0 \notin t_{\mathrm{opt}}(k-1, \varepsilon_g + \varepsilon - t_0)$ and so $\delta_k^{\mathrm{A}}(\varepsilon_g) > \delta_k^{\mathrm{NA}}(\varepsilon_g)$ by Lemma D.2.

Case III: ($\ell = 1$) This will follow from the same argument as the previous case. For this setting we have $t_1 = \frac{\varepsilon_g + 2\varepsilon}{k+1}$. Once again, we use our more restrictive assumption that $\varepsilon_g \in [0, (k-3)\varepsilon]$, and therefore $\varepsilon_g + 2(\varepsilon - t_1) \geqslant -(k-2)\varepsilon$. Furthermore, we have

$$\varepsilon_g + 2\left(\varepsilon - \frac{\varepsilon_g + 2\varepsilon}{k+1}\right) = \frac{k-1}{k+1}(\varepsilon_g + 2\varepsilon) \leqslant \frac{(k-1)^2}{k+1}\varepsilon < (k-2)\varepsilon$$

where the last step follows because $(k-1)^2 < (k+1)(k-2)$ for $k > 1$. Thus $\varepsilon_g + 2(\varepsilon - t_1) \in (-(k-2)\varepsilon, (k-2)\varepsilon)$ and by Lemma D.1, $t_{\mathrm{opt}}(k-2, \varepsilon_g + 2(\varepsilon - t_1)) \subseteq \left\{ \frac{\varepsilon_g + 2(\varepsilon - t_1) + (\ell''+1)\varepsilon}{k-1} : \ell'' \in \{0, k-3\} \right\}$. It then follows that

$$\frac{\varepsilon_g + 2\varepsilon}{k+1} = \frac{\varepsilon_g + 2(\varepsilon - t_1)}{k-1} < \frac{\varepsilon_g + 2(\varepsilon - t_1) + (\ell''+1)\varepsilon}{k-1}$$

for any $\ell'' \geqslant 0$. This implies $t_1 \notin t_{\mathrm{opt}}(k-2, \varepsilon_g + 2(\varepsilon - t_1))$ and so $\delta_k^{\mathrm{A}}(\varepsilon_g) > \delta_k^{\mathrm{NA}}(\varepsilon_g)$ by Lemma D.2.

□

D.2. Proofs of lemmas

In this section we collect proofs of lemmas in the previous section.

Proof of Corollary D.1. Note that by our definition, $q_t = \mathrm{e}^t p_t$ and $1 - q_t = \mathrm{e}^{t-\varepsilon}(1 - p_t)$, so we can equivalently write

$$\delta^k(t, \varepsilon_g) = \sum_{i=0}^{k} \binom{k}{i} q_t^{k-i} (1 - q_t)^i \max\left\{ \left(1 - \mathrm{e}^{\varepsilon_g - kt + i\varepsilon}\right), 0 \right\}.$$

We then prove by induction. For $k = 1$, the base case,

$$\delta^1(t, \varepsilon_g) = q_t \max\{1 - \mathrm{e}^{\varepsilon_g - t}, 0\} + (1 - q_t) \max\{1 - \mathrm{e}^{\varepsilon_g - t + \varepsilon}, 0\},$$

and the claim follows by definition of $\delta^0(t, \varepsilon_g)$. We can then apply our inductive hypothesis to get both

$$\begin{aligned}
q_t \cdot \delta^{k-1}(\varepsilon_g - t) &= \sum_{i=0}^{k-1} \binom{k-1}{i} q_t^{k-i} (1 - q_t)^i \max\left\{\left(1 - \mathrm{e}^{\varepsilon_g - kt + i\varepsilon}\right), 0\right\}, \\
(1 - q_t) \cdot \delta^{k-1}(\varepsilon_g - t + \varepsilon) &= \sum_{i=0}^{k-1} \binom{k-1}{i} q_t^{k-1-i} (1 - q_t)^{i+1} \max\left\{\left(1 - \mathrm{e}^{\varepsilon_g - kt + (i+1)\varepsilon}\right), 0\right\} \\
&= \sum_{i=1}^{k} \binom{k-1}{i-1} q_t^{k-i} (1 - q_t)^i \max\left\{\left(1 - \mathrm{e}^{\varepsilon_g - kt + i\varepsilon}\right), 0\right\}.
\end{aligned}$$

Our claim then follows from the fact that for any $i \in [1, k-1]$, we must have $\binom{k-1}{i-1} + \binom{k-1}{i} = \binom{k}{i}$. □

Proof of Lemma D.2. We prove this inductively. For the base case $k = 2$, from Corollary D.1 and our definition of $t_{\text{opt}}(2, \varepsilon_g)$ we have

$$\delta_2^{\text{NA}}(\varepsilon_g) = q_t \delta^1(t, \varepsilon_g - t) + (1 - q_t)\delta^1(t, \varepsilon_g + \varepsilon - t)$$

If $t \notin t_{\text{opt}}(1, \varepsilon_g - t + \ell'\varepsilon)$ for some $\ell' \in \{0, 1\}$, then $\delta_1^{\text{NA}}(\varepsilon_g - t + \ell'\varepsilon) > \delta^1(t, \varepsilon_g - t + \ell'\varepsilon)$. Applying Theorem 3,

$$\begin{aligned}
\delta_2^{\text{A}}(\varepsilon_g) &\geqslant q_t \delta_1^{\text{A}}(\varepsilon_g - t) + (1 - q_t)\delta_1^{\text{A}}(\varepsilon_g - t + \varepsilon) \\
&> q_t \delta^1(t, \varepsilon_g - t) + (1 - q_t)\delta^1(t, \varepsilon_g + \varepsilon - t) = \delta_2^{\text{NA}}(\varepsilon_g)
\end{aligned}$$

This equivalently follows if $\delta_1^{\text{NA}}(\varepsilon_g - t + \ell'\varepsilon) < \delta_1^{\text{A}}(\varepsilon_g - t + \ell'\varepsilon)$ for either $\ell' \in \{0, 1\}$.

The inductive step will then follow equivalently. Once again, we have

$$\delta_k^{\text{NA}}(\varepsilon_g) = q_t \delta^{k-1}(t, \varepsilon_g - t) + (1 - q_t)\delta^{k-1}(t, \varepsilon_g + \varepsilon - t)$$

which similarly implies

$$\begin{aligned}
\delta_k^{\text{A}}(\varepsilon_g) &\geqslant q_t \delta_{k-1}^{\text{A}}(\varepsilon_q - t) + (1 - q_t)\delta_{k-1}^{\text{A}}(\varepsilon_g - t + \varepsilon) \\
&\geqslant q_t \delta_{k-1}^{\text{NA}}(\varepsilon_q - t) + (1 - q_t)\delta_{k-1}^{\text{NA}}(\varepsilon_g - t + \varepsilon) \\
&\geqslant q_t \delta^{k-1}(t, \varepsilon_g - t) + (1 - q_t)\delta^{k-1}(t, \varepsilon_g + \varepsilon - t) = \delta_k^{\text{NA}}(\varepsilon_g)
\end{aligned}$$

The goal will then be to show that this inequality becomes strict if one of the conditions in the statement holds. First, suppose $t \notin t_{\text{opt}}(k-1, \varepsilon_g - t + \ell'\varepsilon)$ for either $\ell' \in \{0, 1\}$, then $\delta_{k-1}^{\text{NA}}(\varepsilon_g - t + \ell'\varepsilon) > \delta_g^{k-1}(t, \varepsilon_g - t + \ell'\varepsilon)$ and the inequality must be strict. On the other hand, if $t \in t_{\text{opt}}(k-1, \varepsilon_g - t + \ell'\varepsilon)$ for both $\ell' \in \{0, 1\}$, then this fits the condition of our inductive hypothesis, and we will then use this to prove our claim for the remaining cases.

Let $0 \leqslant \ell' \leqslant \ell < k$ be such that $\delta_{k-\ell}^{\text{A}}(\varepsilon_g - \ell t + \ell'\varepsilon) > \delta_{k-\ell}^{\text{NA}}(\varepsilon_g - \ell t + \ell'\varepsilon)$, and if $\ell = 0$, then the inequality holds trivially. If $\ell \geqslant 1$, then rewriting the inequality, we equivalently have both of the following inequalities,

$$\begin{aligned}
&\delta_{k-1-(\ell-1)}^{\text{A}}(\varepsilon_g - t - (\ell-1)t + \ell'\varepsilon) > \delta_{k-1-(\ell-1)}^{\text{NA}}(\varepsilon_g - t - (\ell-1)t + \ell'\varepsilon), \\
\text{and} \quad &\delta_{k-1-(\ell-1)}^{\text{A}}(\varepsilon_g - t + \varepsilon - (\ell-1)t + (\ell'-1)\varepsilon) \\
&\qquad > \delta_{k-1-(\ell-1)}^{\text{NA}}(\varepsilon_g - t + \varepsilon - (\ell-1)t + (\ell'-1)\varepsilon).
\end{aligned}$$

If $\ell \geqslant 1$, then we must have either $0 \leqslant \ell' \leqslant (\ell-1) < k-1$ or $0 \leqslant (\ell'-1) \leqslant (\ell-1) < k-1$. We can then apply our inductive hypothesis to achieve $\delta^{\mathrm{A}}_{k-1}(\varepsilon_g - t) > \delta^{\mathrm{NA}}_{k-1}(\varepsilon_g - t)$, or $\delta^{\mathrm{A}}_{k-1}(\varepsilon_g - t + \varepsilon) > \delta^{\mathrm{NA}}_{k-1}(\varepsilon_g - t + \varepsilon)$, respectively, which implies that our inequality is strict.

Similarly, let $0 \leqslant \ell' \leqslant \ell < k$ be such that $t \notin t_{\text{opt}}(k-\ell, \varepsilon_g - \ell t + \ell'\varepsilon)$ By definition we cannot have $\ell = 0$, and we previously considered $\ell = 1$, so we assume $\ell > 1$ in order to apply our inductive claim. Rewriting the set t_{opt}, we must then have both hold

$$\begin{aligned} & t \notin t_{\text{opt}}(k-1-(\ell-1), \varepsilon_g - t - (\ell-1)t + \ell'\varepsilon) \\ \text{and} \quad & t \notin t_{\text{opt}}(k-1-(\ell-1), \varepsilon_g - t + \varepsilon - (\ell-1)t + (\ell'-1)\varepsilon). \end{aligned}$$

If $\ell > 1$, then $\ell - 1 > 0$ and either $0 \leqslant \ell' \leqslant (\ell-1) < k-1$ or $0 \leqslant (\ell'-1) \leqslant (\ell-1) < k-1$. Applying our inductive hypothesis, we have either case hold, respectively

$$\begin{aligned} & \delta^{\mathrm{A}}_{k-1}(\varepsilon_g - t) > \delta^{\mathrm{NA}}_{k-1}(\varepsilon_g - t), \\ \text{or } & \delta^{\mathrm{A}}_{k-1}(\varepsilon_g - t + \varepsilon) > \delta^{\mathrm{NA}}_{k-1}(\varepsilon_g - t + \varepsilon). \end{aligned}$$

This implies our inequality is strict. □

In order to prove Lemma D.4, we will also need the following edge case.

Lemma D.6. *$t_{\text{opt}}(2, -3\varepsilon/2) \cap t_{\text{opt}}(2, \varepsilon/2) = \emptyset$ and $t_{\text{opt}}(2, -\varepsilon/2) \cap t_{\text{opt}}(2, \varepsilon/2) = \emptyset$*

Proof. For any $\varepsilon_g \in \{-3\varepsilon/2, -\varepsilon/2, \varepsilon/2, 3\varepsilon/2\}$, from Lemma D.1 that $t_{\text{opt}}(2, \varepsilon_g) \subseteq \{\frac{\varepsilon_g + (\ell+1)\varepsilon}{3}\} \cap (0, \varepsilon)$ for $\ell \in \{0, 1\}$. This then implies that $t_{\text{opt}}(2, -3\varepsilon/2) = \varepsilon/6$, $t_{\text{opt}}(2, -\varepsilon/2) \subseteq \{\varepsilon/6, \varepsilon/2\}$, $t_{\text{opt}}(2, \varepsilon/2) \subseteq \{\varepsilon/2, 5\varepsilon/6\}$, and $t_{\text{opt}}(2, 3\varepsilon/2) = 5\varepsilon/6$. The claim then follows immediately.

□

Proof of Lemma D.4. By our assumptions, it immediately follows that either there exists $0 \leqslant j \leqslant k-2$ such that $\varepsilon_g - (k-2)t + j\varepsilon \in (-\varepsilon/2, \varepsilon/2)$, or we are in the edge case where there exists $0 \leqslant j < k-2$ such that $\varepsilon_g - (k-2)t + j\varepsilon = -\varepsilon/2$. In first case, we know that $\delta^{\mathrm{NA}}_2(\varepsilon_g - (k-2)t_\ell + j\varepsilon) < \delta^{\mathrm{A}}_2(\varepsilon_g - (k-2)t + j\varepsilon)$ from Lemma D.3. In the second case (the edge case), if $j = 0$ then we know $j + 2 \leqslant k - 2$ because $k \geqslant 4$, and from Lemma D.6 we must either have $t \notin t_{\text{opt}}(2, \varepsilon_g - (k-2)t_\ell + j\varepsilon)$ or $t \notin t_{\text{opt}}(2, \varepsilon_g - (k-2)t_\ell + (j+2)\varepsilon)$. Otherwise, if $j > 0$, then we again have from Lemma D.6 that either $t \notin t_{\text{opt}}(2, \varepsilon_g - (k-2)t_\ell + (j-1)\varepsilon)$ or $t \notin t_{\text{opt}}(2, \varepsilon_g - (k-2)t_\ell + (j+1)\varepsilon)$.

In either case, we can immediately apply Lemma D.2 to achieve our desired inequality. □

E. Proof of Proposition 4 and Theorem 5

In this section we prove Proposition 4 and Theorem 5, which are sub-optimal, but involves less computation. Because we give up optimality, we can now deal with inhomogeneous composition, i.e. each step is ε_i-BR with different ε_i. Both apply to adaptive composition, and hence also non-adaptive composition. In Appendix E.1 we summarize a general approach for this kind of results, all of which employs tricks from standard concentration inequality. Appendix E.2 treats BR mechanisms, prove Proposition 4 and Theorem 5 and compare with previous results. Appendix E.3 provides missing proofs from Appendix E.2.

E.1. General approach from concentration inequality

Consider the adaptive composition of $M_i : X \times Y_1 \times \cdots \times Y_{i-1} \to Y_i, i = 1, 2, \ldots, k$. The output $(y_1, y_2, \ldots, y_k)$ of the composition $M : X \to Y_1 \times \cdots \times Y_k$ satisfies

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

Fix a pair of neighboring datasets x, x'. Let L_i be log likelihood ratio for the i-th mechanism, i.e.

$$L_i(y_1, \dots, y_i) = \ln \frac{P[M(x', y_1, \cdots, y_{i-1}) = y_i]}{P[M(x, y_1, \cdots, y_{i-1}) = y_i]}$$

Note that all these quantities depend on x and x'. Let $\delta_{\text{opt}}(M, \varepsilon)$ be the optimal δ such that M is (ε, δ)-DP. We have the following lemma that bounds $\delta_{\text{opt}}(M, \varepsilon)$.

Lemma E.1. *If we can find functions $U_i : (0, +\infty) \to \mathbb{R}$ such that for all x, x' the following holds*

$$\ln \mathbb{E}_Q[\mathrm{e}^{\lambda L_i} \mid y_1, \dots, y_{i-1}] \leqslant U_i(\lambda), \quad \forall \lambda > 0.$$

Then

$$\delta_{\text{opt}}(M, \varepsilon) \leqslant \mathrm{e}^{-\sup_{\lambda>0}\{\lambda\varepsilon - \sum_{i=1}^k U_i(\lambda)\}}.$$

Proof. For a pair of neighboring datasets x, x' let $P = M(x)$ and $Q = M(x')$. Let L be the log likelihood ratio of P and Q. With these notations we have $L = \sum_{i=1}^k L_i$. By Lemma 4.2 we have

$$\begin{aligned}
\delta_{\text{opt}}(M, \varepsilon) &= \sup_{x,x'} Q[L > \varepsilon] - \mathrm{e}^{\varepsilon} P[L > \varepsilon] \\
&\leqslant \sup_{x,x'} Q[L > \varepsilon] \\
&= \sup_{x,x'} Q\big[\mathrm{e}^{\lambda \sum_{i=1}^k L_i} > \mathrm{e}^{\lambda\varepsilon}\big] \\
&\leqslant \sup_{x,x'} \mathrm{e}^{-\lambda\varepsilon} \cdot \mathbb{E}_Q\big[\mathrm{e}^{\lambda \sum_{i=1}^k L_i}\big]
\end{aligned}$$

We can use the tower rule of expectation to show that

$$\mathbb{E}\big[\mathrm{e}^{\lambda \sum_{i=1}^k L_i}\big] = \mathbb{E}\big[\mathrm{e}^{\lambda \sum_{i=1}^{k-1} L_i} \cdot \mathrm{e}^{\lambda L_k}\big] = \mathbb{E}\big[\mathrm{e}^{\lambda \sum_{i=1}^{k-1} L_i} \cdot \mathbb{E}[\mathrm{e}^{\lambda L_k} \mid y_1, \dots, y_{k-1}]\big] \leqslant \mathrm{e}^{U_k(\lambda)} \cdot \mathbb{E}\big[\mathrm{e}^{\lambda \sum_{i=1}^k L_i}\big].$$

Continue doing this, we have

$$\mathbb{E}_Q\big[\mathrm{e}^{\lambda \sum L_i}\big] \leqslant \mathrm{e}^{\sum_{i=1}^k U_i(\lambda)}.$$

Furthermore, it holds for all neighboring x, x', so

$$\delta_{\text{opt}}(M, \varepsilon) \leqslant \mathrm{e}^{-\lambda\varepsilon} \cdot \mathrm{e}^{\sum_{i=1}^k U_i(\lambda)}.$$

It also holds for arbitrary $\lambda > 0$, so we can optimize over λ and get

$$\delta_{\text{opt}}(M, \varepsilon) \leqslant \mathrm{e}^{-\sup_{\lambda>0}\{\lambda\varepsilon - \sum_{i=1}^k U_i(\lambda)\}}.$$

□

E.2. Bounding U_i's for bounded range mechanisms

For now on we assume the component mechanisms M_i is ε_i-BR, $i = 1, 2, \dots, k$. More precisely, we assume $M_i(\cdot, y_1, \dots, y_{i-1})$ is ε_i-BR for all $y_1 \in Y_1, \dots, y_{i-1} \in Y_{i-1}$.

Lemma E.1 yields bounds on $\delta_k^{\mathsf{A}}(\varepsilon_g)$. In fact, Different choices of $U_i(\lambda)$ in Lemma E.1 result in different bounds on $\delta_k^{\mathsf{A}}(\varepsilon_g)$. For example, both Dwork et al. (2010) and Durfee and Rogers (2019) utilize the following lemma:

Lemma E.2 (Hoeffding's lemma)**.** *If a random variable $X \in [a, b]$ then* $\ln \mathbb{E}[\mathrm{e}^{\lambda X}] \leqslant \frac{1}{8}(b-a)^2\lambda^2 + \lambda \mathbb{E}X$.

We now walk through the following comparisons with previous work to highlight our improvement. Dwork et al. (2010) only uses the fact that $L_i \in [-\varepsilon_i, \varepsilon_i]$ (which is weaker than ε-BR). It implies

(a) $\ln \mathbb{E}_Q[\mathrm{e}^{\lambda L_i} \mid y_1, \dots, y_{i-1}] \leqslant \frac{1}{2}\varepsilon_i^2\lambda^2 + \lambda \mathbb{E}_Q[L_i \mid y_1, \dots, y_{i-1}]$

(b) $\mathbb{E}_Q[L_i \mid y_1, \ldots, y_{i-1}] \leqslant \varepsilon_i \tanh \frac{\varepsilon_i}{2} \leqslant \frac{1}{2}\varepsilon_i^2$.

For part (b), Dwork et al. (2010) used a much rougher estimate. The $\frac{1}{2}\varepsilon_i^2$ upper bound appears in (Bun and Steinke, 2016). For the most refined bound in terms of hyperbolic tangent function, readers can refer to Lemma D.8 in (Dong et al., 2019).

Combining both (a) and (b), we have $U_i(\lambda) = \frac{1}{2}\varepsilon_i^2(\lambda^2 + \lambda)$, which we refer to as "Improved DRV10" in Figure 1.

Using the bounded range property from Durfee and Rogers (2019), we know that for ε-BR there is a $t_i \in [0, \varepsilon_i]$ such that $a = t_i - \varepsilon_i, b = t_i$ in Hoeffding's lemma. A similar argument yields $U_i(\lambda) = \frac{1}{2}\varepsilon_i^2(\frac{1}{4}\lambda^2 + \lambda)$, which we label as "DR19" in Figure 1.

A straightforward improvement could come from a finer treatment of (b). In fact,

Lemma E.3. *Let* $\operatorname{maxkl}(\varepsilon) := \frac{\varepsilon}{\mathrm{e}^\varepsilon - 1} - 1 - \ln \frac{\varepsilon}{\mathrm{e}^\varepsilon - 1}$. *Then we have*

(b′) $\mathbb{E}_Q[L_i \mid y_1, \ldots, y_{i-1}] \leqslant \operatorname{maxkl}(\varepsilon_i)$.

Combining (a) and (b′), we can use $U_i(\lambda) = \frac{1}{8}\varepsilon_i^2\lambda^2 + \lambda \cdot \operatorname{maxkl}(\varepsilon_i)$, which we label as "KL-improved DR19" in Figure 1. This observation on the expectation together with the Durfee and Rogers (2019) bound that uses Azuma-Hoeffding, but with a weaker bound on the expectation term allows us to derive Proposition 4. We skip the algebra.

The finest analysis is stated in the following lemma, which basically says $U_i(\lambda)$ can be taken as $h_{\varepsilon_i}(\lambda)$. Recall that $h(\lambda; \varepsilon)$ is defined to be $\sup_{t \in [0,\varepsilon]} \lambda(\varepsilon - t) + \ln\left(1 + p_{\varepsilon,t}(\mathrm{e}^{-\lambda\varepsilon} - 1)\right)$. We label it as "General MGF" in Figure 1.

Lemma E.4.

$$\ln \mathbb{E}_Q[\mathrm{e}^{\lambda L_i} \mid y_1, \ldots, y_{i-1}] \leqslant h_{\varepsilon_i}(\lambda).$$

Theorem 5 directly follows from this lemma and Lemma E.1.

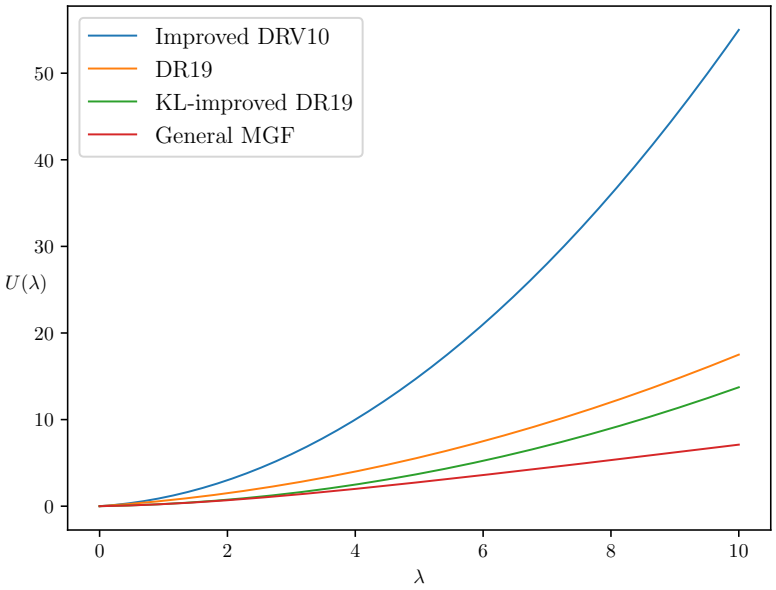

Figure 1. A unified view and comparison of composition theorems involving concentration inequalities. The figure shows graphs of different U functions (see Lemma E.1) used in different results, such as from Dwork et al. (2010) (labeled "Improved DRV10") and Durfee and Rogers (2019) (labeled "DR19"). According to Lemma E.1, smaller function U yields tighter privacy result. Theorem 5 uses the smallest U (labeled "General MGF") among all and is hence the tightest. All curves use $\varepsilon = 1$.

Numerical Issue We now point out a potential numeric issue in computing the function $h_\varepsilon(\lambda)$. Note that it can be simplified differently as

$$h_\varepsilon(\lambda) = \sup_{t\in[0,\varepsilon]} -\lambda t + \ln\big(p_{\varepsilon,t} + \mathrm{e}^{\varepsilon\lambda}(1-p_{\varepsilon,t})\big).$$

For comparision, the expression we use in the definition is

$$h_\varepsilon(\lambda) = \sup_{t\in[0,\varepsilon]} \lambda(\varepsilon - t) + \ln\big(1 + p_{\varepsilon,t}(\mathrm{e}^{-\lambda\varepsilon} - 1)\big).$$

At first glance it may appear that the above two expressions are equal. However, the one used in the theorem is far more robust numerically, as in the optimization step, $\varepsilon\lambda$ can be large, which could make $\mathrm{e}^{\varepsilon\lambda}$ beyond the range of floating point numbers.

E.3. Proof of Lemmas E.3 and E.4

We do Lemma E.4 first.

Proof of Lemma E.4. In order to simplify notations (especially the subscripts), we assume $k = 2$. It will be clear that this is without loss of generality. That is, $M_1 : X \to Y$ is ε_1-DP and $M_2 : X \times Y \to Z$ satisfy that $M_2(\cdot, y) : X \to Z$ is ε_2-DP for all $y \in Y$. Let M be their composition. For a fixed pair of x, x', let P, Q be the distributions of $M(x)$ and $M(x')$ and $L(y,z) = \ln\frac{q(y,z)}{p(y,z)}$ be the log likelihood ratio. We are interested in

$$\ln\mathbb{E}\Big[\Big(\frac{q(z|y)}{p(z|y)}\Big)^\lambda\Big|y\Big].$$

By Lemma 4.1, there is a $t_2 = t_2(x, x', y)$ such that a common randomized function $\mathrm{Proc} : \{0,1\} \to Z$ turns $\mathrm{Bern}(p_{\varepsilon_2,t_2})$ to $p(\cdot|y)$ and $\mathrm{Bern}(q_{\varepsilon_2,t_2})$ to $q(\cdot|y)$. This means we can use a data processing inequality on the expectations. In fact, recall that Rényi divergence of Q and P of order $\alpha > 1$ is defined as

$$D_\alpha(Q\|P) = \tfrac{1}{\alpha-1}\ln\mathbb{E}_Q[(\tfrac{Q}{P})^{\alpha-1}].$$

We see that $\ln\mathbb{E}\Big[\Big(\frac{q(z|y)}{p(z|y)}\Big)^\lambda\Big|y\Big] = \lambda D_{\lambda+1}\big(q(\cdot|y)\|p(\cdot|y)\big)$. Data processing inequality of Rényi divergence (see (Van Erven and Harremos, 2014)) implies

$$\lambda D_{\lambda+1}\big(q(\cdot|y)\|p(\cdot|y)\big) \leqslant \lambda D_{\lambda+1}\big(\mathrm{Bern}(q_{\varepsilon_2,t_2})\|\mathrm{Bern}(p_{\varepsilon_2,t_2})\big).$$

We need to compute Rényi divergence of two Bernoullis. Making use of the following facts

$$\begin{aligned} q_t &= \mathrm{e}^t p_t, \quad 1 - q_t = \mathrm{e}^{t-\varepsilon}(1-p_t),\\ q_{\varepsilon-t} &= 1 - p_t, \quad p_{\varepsilon-t} = 1 - q_t, \end{aligned}$$

we have

$$\begin{aligned} \lambda D_{\lambda+1}\big(\mathrm{Bern}(q_t)\|\mathrm{Bern}(p_t)\big) &= \ln\big(q_t^{\lambda+1}p_t^{-\lambda} + (1-q_t)^{\lambda+1}(1-p_t)^{-\lambda}\big)\\ &= \ln\big(q_t\mathrm{e}^{\lambda t} + (1-q_t)\mathrm{e}^{\lambda(t-\varepsilon)}\big)\\ &= \lambda t + \ln\big(q_t + (1-q_t)\mathrm{e}^{-\lambda\varepsilon}\big)\\ &= \lambda t + \ln\big(1 - p_{\varepsilon-t} + p_{\varepsilon-t}\mathrm{e}^{-\lambda\varepsilon}\big)\\ &= \underbrace{\lambda t + \ln\big(1 + p_{\varepsilon-t}(\mathrm{e}^{-\lambda\varepsilon} - 1)\big)}_{\psi(\lambda,\varepsilon,t)}. \end{aligned}$$

Recall that $h(\lambda;\varepsilon)$ is defined to be $\sup_{t\in[0,\varepsilon]} \lambda(\varepsilon-t)+\ln\big(1+p_{\varepsilon,t}(\mathrm{e}^{-\lambda\varepsilon}-1)\big)$. With this notation, we can do a change of variable $t'=\varepsilon-t$ and have

$$\begin{aligned} h(\lambda;\varepsilon) &= \sup_{t\in[0,\varepsilon]} \lambda(\varepsilon-t)+\ln\big(1+p_{\varepsilon,t}(\mathrm{e}^{-\lambda\varepsilon}-1)\big) \\ &= \sup_{t'\in[0,\varepsilon]} \lambda t'+\ln\big(1+p_{\varepsilon,\varepsilon-t'}(\mathrm{e}^{-\lambda\varepsilon}-1)\big) \\ &= \sup_{t\in[0,\varepsilon]} \psi(\lambda,\varepsilon,t). \end{aligned}$$

This explains why $h(\lambda;\varepsilon)$ is defined this way. In summary,

$$\ln\mathbb{E}\Big[\Big(\frac{q(z|y)}{p(z|y)}\Big)^{\lambda}\Big|y\Big] = \lambda D_{\lambda+1}\big(q(\cdot|y)\|p(\cdot|y)\big) \leqslant \lambda D_{\lambda+1}\big(\mathrm{Bern}(q_{\varepsilon_2,t_2})\|\mathrm{Bern}(p_{\varepsilon_2,t_2})\big) = \psi(\lambda,\varepsilon_2,t_2) \leqslant h(\lambda;\varepsilon_2).$$

Clearly the argument carries over to general k. □

Proof of Lemma E.3. Using the same "Lemma 4.1 + data processing inequality" argument, we have

$$\begin{aligned} \mathbb{E}_Q[L_i \mid y_1,\dots,y_{i-1}] &\leqslant \mathrm{KL}\big(\mathrm{Bern}(q_{\varepsilon_i,t_i})\|\mathrm{Bern}(p_{\varepsilon_i,t_i})\big) \\ &= q_{\varepsilon_i,t_i}\cdot\ln\frac{q_{\varepsilon_i,t_i}}{p_{\varepsilon_i,t_i}} + (1-q_{\varepsilon_i,t_i})\cdot\ln\frac{1-q_{\varepsilon_i,t_i}}{1-p_{\varepsilon_i,t_i}} \\ &= t_i q_{\varepsilon_i,t_i} + (t_i-\varepsilon_i)(1-q_{\varepsilon_i,t_i}) \\ &= t_i - \frac{\varepsilon_i}{\mathrm{e}^{\varepsilon_i}-1}(\mathrm{e}^{t_i}-1). \end{aligned}$$

A bit of calculus shows the above expression is maximized at $t_i=\ln\frac{\mathrm{e}^{\varepsilon_i}-1}{\varepsilon_i}$, and the value is

$$\frac{\varepsilon_i}{\mathrm{e}^{\varepsilon}_i-1} - 1 - \ln\frac{\varepsilon_i}{\mathrm{e}^{\varepsilon}_i-1} = \mathrm{maxkl}(\varepsilon_i).$$

□

References

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

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

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

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

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

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.

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