Lemmas of Differential Privacy

Yiyang Huang
University of New South Wales
sophia.huang1@student.unsw.edu.au

Clément L. Canonne
University of Sydney
clement.canonne@sydney.edu.au

November 22, 2022

Abstract
We aim to collect buried lemmas that are useful for proofs. In particular, we try to provide self-contained proofs for those lemmas and categorise them according to their usage.

Contents

1 Introduction 1
2 Preliminaries 2

2.1 Central Differential Privacy . 2
2.2 Local Differential Privacy . 2
2.3 Shuffle Differential Privacy . 2
2.4 Private- and public-coin protocols . 3

3 Central Differential Privacy 3

4 Local Differential Privacy 5

4.1 Replacement and Removal LDP . 5
4.2 Replacement LDPs . 6

5 Relation between Shuffle DP and LDP 9

6 Privacy Amplification by Subsampling 11

7 Appendix 13

7.1 Proof of the weaker result in Corollary 4.1 13

1 Introduction
Differential privacy [DMNS06] ensures that the result of an algorithm is statistically insensitive to changes in its input. The core differential privacy framework has been expanded with additional constraints based on the different stakeholders involved, such as central-, local-, shuffle-, and many others. Nonetheless, due to the sheer amount of work in this field, some useful results
are “buried” and remain somewhat unknown, and some are repeatedly proved. We hope this survey can serve as a helpful compendium, ideally curating some valuable facts for various differential privacy settings.

2 Preliminaries

We first recall the notions of central, local, and shuffle privacy. In what follows, two datasets \(X, X' \in \mathcal{X}^n\) consisting of \(n\) entries are said to be neighboring (denoted \(X \sim X'\)) if they differ in exactly one entry – i.e., are at Hamming distance one.

2.1 Central Differential Privacy

Definition 2.1 ((Central) Differential Privacy). Fix \(\varepsilon > 0\) and \(\delta \in [0, 1]\). An algorithm \(M : \mathcal{X}^n \rightarrow \mathcal{Y}\) satisfies \((\varepsilon, \delta)\)-differential privacy (DP) if for every pair of neighboring datasets \(X, X'\), and every (measurable) subset \(S \subseteq \mathcal{Y}\):

\[
\Pr[M(X) \in S] \leq e^\varepsilon \Pr[M(X') \in S] + \delta.
\]

We further say that \(M\) satisfies pure differential privacy (\(\varepsilon\)-DP) if \(\delta = 0\), otherwise it is approximate differential privacy.

2.2 Local Differential Privacy

Definition 2.2 (Local Differential Privacy). An algorithm (local randomizer) \(R : \mathcal{X} \rightarrow \mathcal{Y}\) satisfies \((\varepsilon, \delta)\)-local differential privacy (LDP) if for any two data records, \(x, x' \in \mathcal{X}\), and every subset \(S \subseteq \mathcal{Y}\):

\[
\Pr[R(x) \in S] \leq e^\varepsilon \Pr[R(x') \in S] + \delta.
\]

We define pure and approximate LDP analogously to the (Central) DP case.

The definition of Local DP (LDP) above is sometimes referred to as Replacement Local Differential Privacy, to contrast it with the variant below, Deletion LDP, introduced in [EFMRSTT20].

Definition 2.3 (Deletion/Removal LDP). An algorithm \(R : \mathcal{X} \rightarrow \mathcal{Y}\) is a deletion \((\varepsilon, \delta)\)-differentially private local randomizer if there exists a random variable (also known as the reference distribution) \(R_0\) such that for all \(S \subseteq \mathcal{Y}\) and for all \(x \in \mathcal{X}\):

\[
e^{-\varepsilon}(\Pr[R_0 \in S] - \delta) \leq \Pr[R(x) \in S] \leq e^\varepsilon \Pr[R_0 \in S] + \delta.
\]

Unless explicitly specified otherwise, a local randomizer is always (replacement) LDP. One advantage of the notion of deletion LDP is that it allows in some case to obtain tighter results, e.g., by constant factors.

2.3 Shuffle Differential Privacy

Definition 2.4 (Shuffle Model). A protocol \(P = (R, S, A)\) in the shuffle model consists of three building blocks, and applies to the data from \(n\) users, where user \(i\) has data \(x_i \in \mathcal{X}\).

1. Each user applies the local randomizer \(R : \mathcal{X} \rightarrow \mathcal{Y}^*\) to their data and reports messages \((y_{i,1}, \ldots, y_{i,m}) \leftarrow R(x_i)\). (The number of messages \(m\) can itself be randomized.)

2. The tuple of all resulting messages \((y_{i,j})_{i,j}\) is sent to the shuffler \(S : \mathcal{Y}^* \rightarrow \mathcal{Y}^*\) that takes in these messages and outputs them in a uniformly random order.

3. The shuffled tuple of messages is then passed through some analyzer \(A : \mathcal{Y}^* \rightarrow \mathcal{Z}\) to estimate some function \(f(x_1, \ldots, x_n)\).
Lemma 3.1 3. Central Differential Privacy

Note that the definition of Shuffle Differential Privacy does not guarantees anything in the presence of malicious users, i.e., users which may deviate from the protocol (not use the randomizer $R$, but something else) in order to jeopardize the privacy of other (honest) users. To account for this, the variant of Robust Shuffle Differential Privacy was introduced, which ensures privacy even when only a fraction of the users do follow the protocol.

Definition 2.6 (Robust Shuffle Differential Privacy). Fix $\gamma \in (0, 1]$. A protocol $P = \langle R, S, A \rangle$ is $(\varepsilon, \delta, \gamma)$-robustly shuffle differentially private if, for all $n \in \mathbb{N}$ and $\gamma' \geq \gamma$, the algorithm $S \circ R$, while only taking $\gamma \cdot n$ users input, is $(\varepsilon, \delta)$-differentially private. I.e., $P$ guarantees $(\varepsilon, \delta)$-shuffle differential privacy when at least a $\gamma$ fraction of users follow the protocol.

While many shuffle private protocols which can be found in the literature do satisfy this robust version "out-of-the-box", the two notions are not equivalent and there exist shuffle private protocols which are not robust: see, e.g., [Che21, Appendix C].

2.4 Private- and public-coin protocols

In both the local and shuffle DP settings, all users run a local randomizer on their private data. This is captured by formally letting the randomizer $R$ be a deterministic mapping $R : \mathcal{X} \times \{0, 1\}^* \to \mathcal{Y}$, where the second input is the randomness, i.e., a string of uniformly random bits, assumed private ("private coins"). Then, user $i$, having data $x_i$ and "private" randomness $r_i$, computes $R(x_i, r_i)$, and the $(\varepsilon, \delta)$-DP guarantee must hold as the probability is taken over the random choice of $r_i$.

One can also assume that a source of public randomness ("public coins") is available to all parties (users, analyzer, and outside world alike), acting as a common (non-private) random seed. In this case, the users still have access to their own, private randomness (necessary to ensure differential privacy), but also to an additional input, the public random string (possibly useful to achieve better accuracy or utility). In this case, the "public-coin" setting, $R$ can be seen as a deterministic mapping of the form

$$R : \mathcal{X} \times \{0, 1\}^* \times \{0, 1\}^* \to \mathcal{Y}$$

and user $i$, having data $x_i$, "private" randomness $r_i$, and common (to all users) randomness $r_{pub}$, computes $R(x_i, r_i, r_{pub})$. Then, the $(\varepsilon, \delta)$-DP guarantee must hold when the probability is taken over the random choice of $r_i$, for every fixed setting of $r_{pub}$.

3 Central Differential Privacy

Lemma 3.1 (Pure DP Implies Approximate DP. [ASZ18, Lemma 5]). Any $(\varepsilon + \delta, 0)$-differentially private algorithm is also $(\varepsilon, \delta)$-differentially private.

Proof. Suppose $A$ is a $(\varepsilon + \delta, 0)$-DP algorithm, then for any neighboring dataset $X, X'$ and any $S \subseteq \text{range}(A)$, we have:

$$\Pr[A(X) \in S] \leq e^{\varepsilon + \delta} \Pr[A(X') \in S]$$

$$= e^\varepsilon \Pr[A(X') \in S] + (e^\delta - 1)e^\varepsilon \Pr[A(X') \in S].$$

We want to show that $\Pr[A(X) \in S] \leq e^\varepsilon \Pr[A(X') \in S] + \delta$. Since the inequality is trivially true if $e^\varepsilon \Pr[A(X') \in S] + \delta > 1$, we can assume $e^\varepsilon \Pr[A(X') \in S] \leq 1 - \delta$. The proof is complete if we show that $(e^\delta - 1)e^\varepsilon \Pr[A(X') \in S] \leq \delta$. In this case,

$$(e^\delta - 1)e^\varepsilon \Pr[A(X') \in S] \leq (e^\delta - 1)(1 - \delta) \leq (e^\delta - 1)e^{-\delta} = 1 - e^{-\delta} \leq \delta$$
where the last two inequalities use the fact that $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$.

The following is a “folklore” result, showing that one can convert approximate DP to pure DP for mechanisms with finite output space (and so, in particular, decision algorithms).

**Lemma 3.2 (Approximate DP Implies Pure DP for Finite Output Spaces).** Suppose $A: X \rightarrow Y$ is an $(\epsilon, \delta)$-DP mechanism, where $|Y| = k \in \mathbb{N}$. Then, for every $\eta \in (0, 1]$, there exists an $\varepsilon'$-DP mechanism $A': X \rightarrow Y$ such that, for every $x \in X$, $TV(A(x), A'(x)) \leq \eta$, where $\varepsilon' := \epsilon + \ln(1 + \frac{k \delta}{\eta} e^{-\epsilon})$. In particular, for $Y = \{0, 1\}$, $\varepsilon' \leq \epsilon + \frac{2\delta}{\eta}$.

**Proof.** Let $A$ be an $(\epsilon, \delta)$-DP mechanism as in the statement, and define $A'$ as the algorithm which, on input $x \in X$, outputs $A(x)$ with probability $1 - \eta$ and, otherwise, outputs an element of $Y$ uniformly at random. We have that, for every $x$, denoting by $u_Y$ the uniform distribution on $Y$,

$$TV(A(x), A'(x)) = \frac{1}{2} \sum_{y \in Y} |\Pr[A(x) = y] - \Pr'[(1 - \eta) A(x) = y + \frac{\eta}{k}]|$$

$$= \eta \frac{1}{2} \sum_{y \in Y} |\Pr[A(x) = y] - \frac{1}{k}| = \eta \cdot TV(A(x), u_Y)$$

$$\leq \eta .$$

Note that $A'$ is also $(\epsilon, \delta)$-DP by postprocessing; and further, for every $x$, we now have $Pr[A'(x) \in S] \geq \eta \cdot \frac{|S|}{k}$ for every $S \subseteq Y$. Together, the two imply that, for every $S$ and any two neighboring $x \sim x'$,

$$Pr[A'(x') \in S] \leq e^\varepsilon Pr[A'(x) \in S] + \delta$$

$$= e^\varepsilon Pr[A'(x) \in S] + \frac{k \delta}{|S|} \eta |S|$$

$$\leq e^\varepsilon Pr[A'(x) \in S] + \frac{k \delta}{|S|} \eta Pr[A'(x) \in S]$$

$$\leq \left(e^\varepsilon + \frac{k \delta}{\eta}\right) \Pr[A'(x) \in S] = e^{\varepsilon'} \Pr[A'(x) \in S]$$

where $\varepsilon' := \ln\left(\frac{e^\varepsilon + \frac{k \delta}{\eta}}{\eta}\right) = \epsilon + \ln\left(1 + \frac{k \delta}{\eta} e^{-\epsilon}\right).$
It is easy to see that Lemma 3.4 stated in [FMT21] follows as a corollary where $R$ is a $(\epsilon, 0)$-LDP randomizer with $x, x'$ of size 1.

**Corollary 3.1** (LDP Randomizer as Postprocessing of Binary Randomized Response [FMT21], Lemma 3.4). Let $R: \mathcal{X} \rightarrow \mathcal{Y}$ be an $\epsilon$-LDP local randomizer and for all $x, x' \in \mathcal{X}$. Then there exists a randomized algorithm $Q : \{0, 1\} \rightarrow \mathcal{Y}$ such that $R(x) = \frac{e^{\epsilon}}{e^{\epsilon} + 1}Q(0) + \frac{1}{e^{\epsilon} + 1}Q(1)$ and $R(x') = \frac{1}{e^{\epsilon} + 1}Q(0) + \frac{e^{\epsilon}}{e^{\epsilon} + 1}Q(1)$.

### 4 Local Differential Privacy

Readers should note that the LDP results we stated in this section are all non-interactive. The interactive case would be more technical, we did not include the most general case here since it’s not the aim of this survey.

#### 4.1 Replacement and Removal LDP

Here, we present some results involving replacement-deletion notion of local differential privacy.

**Lemma 4.1** (Replacement implies Deletion). Every $(\epsilon, \delta)$-replacement LDP randomizer is also an $(\epsilon, \delta)$-deletion LDP randomizer.

It is easy to see that a replacement LDP randomizer $R$ is also a deletion $\epsilon$-differentially private local randomizer by fixing a $x_0$ such that $R_0 = R(x_0)$.

**Lemma 4.2** (Deletion implies Replacement). Every $(\epsilon, \delta)$-deletion LDP randomizer is also an $(2\epsilon, (e^\epsilon + 1)\delta)$-replacement LDP randomizer. In particular, every $\epsilon$-deletion LDP randomizer is also an $2\epsilon$-replacement LDP randomizer.

**Proof.** For a reference distribution $R_0$ and $\forall x \in \mathcal{X}, S \subseteq \mathcal{Y}$ since $R: \mathcal{X} \rightarrow \mathcal{Y}$ is $(\epsilon, \delta)$-deletion LDP, we have:

$$\Pr[R(x) \in S] \leq e^\epsilon \Pr[R_0 \in S] + \delta$$

$$\leq e^\epsilon (e^\epsilon \Pr[R(x') \in S] + \delta) + \delta$$

$$= e^{2\epsilon} \Pr[R(x') \in S] + \delta(1 + e^\epsilon).$$

The second inequality follows from the fact that $R_0$ is the reference distribution for all $x \in \mathcal{X}$, thus it is also true for $x' \in \mathcal{X}$. \hfill \square

[EFMRSTT20] stated, in passing, that “every $(\epsilon, \delta)$-deletion LDP randomizer implies $(2\epsilon, 2\delta)$-replacement LDP”; however, this is not true in general. We provide an counterexample of this statement in the following:

**Counterexample:** Consider, for $\epsilon \in [0, 1/2]$ and $\delta \in [0, 1/5]$ (so that all probabilities below are indeed probabilities), the randomizer $R: \{0, 1\} \rightarrow \{1, 2, 3\}$ such that

$$R(0) = \begin{cases} 1 & \text{w.p. } \frac{e^\epsilon}{3} \\ 2 & \text{w.p. } e^{-\epsilon} \left(\frac{1}{3} - \delta\right) \\ 3 & \text{w.p. } 1 - \left(\frac{e^\epsilon}{3} + e^{-\epsilon} \left(\frac{1}{3} - \delta\right)\right) \end{cases}, \quad R(1) = \begin{cases} 1 & \text{w.p. } \frac{e^{-\epsilon}}{3} \\ 2 & \text{w.p. } \frac{e^\epsilon}{3} + \delta \\ 3 & \text{w.p. } 1 - \left(\frac{e^{-\epsilon}}{3} + \frac{e^\epsilon}{3} + \delta\right) \end{cases}$$

with reference distribution $R_0$ uniform on $\{1, 2, 3\}$. For instance, for $\epsilon = 1/4$ and $\delta = 1/6$, one can check that $R$ as above is indeed an $(\epsilon, \delta)$-deletion LDP randomizer with reference distribution $R_0$; however, we have

$$\Pr[R(1) = 2] = e^{2\epsilon} \Pr[R(0) = 2] + \delta(1 + e^\epsilon) > e^{2\epsilon} \Pr[R(0) = 2] + 2\delta$$
so R is not (2ε, 2δ)-replacement LDP.

Finally, we conclude with the following lemma, which states that every (ε, δ)-deletion LDP randomizer is δ-close to an ε-deletion LDP randomizer.

Lemma 4.3 ([FMT21], Lemma 3.7). Fix any ε > 0, δ ∈ [0, 1]. If R is an (ε, δ)-deletion LDP randomizer with reference distribution R_i, then there exists R’ such that (i) TV(R(x), R’(x)) ≤ δ for every x, x’, and (ii) R’ is an ε-deletion LDP randomizer with reference distribution R_i. In particular, R’ is a 2ε-replacement LDP randomizer by Lemma 4.2.

4.2 Replacement LDPs

Recall that symmetric protocol means all LDP randomizers are the same (not their randomness). The following lemma shows how to convert an asymmetric protocol (where each user might use a different local randomizer) into a symmetric one (where all users use the same local randomizer).

Lemma 4.4 (Asymmetric to Symmetric LDP. [ACFT19, Lemma 4]). Suppose there exists a private-coin (resp., public-coin) mechanism composed of LDP randomizers for some task T with n users and probability of success 5/6. Then, there exists a private-coin (resp., public-coin) mechanism with symmetric LDP randomizers for T with n’ = O(n log n) users and probability of success 2/3.

Proof. We follow the proof from [ACFT19]. Let (R_i)_{i=1}^n be the mechanism with R_i : X → Y being the local randomizer of the i-th user. We create a symmetric (randomized) mechanism R : X → [n] × Y defined as follows: On input x ∈ X: (i) use private randomness to generate j ∈ [n] uniformly at random, then (ii) output (j, R_j(x)).

(In public randomness is available, then R is defined as R : X → Y, where j is chosen uniformly at random using the public randomness, and the output is simply R_j(x) – since all parties have access to the public randomness, there is no need to send j as well.)

We simulate W if we have each j = 1..n once and report them. Further, by a standard coupon-collector argument, requiring n’ = O(n log n) we have that with probability 5/6, each j ∈ [n] will be drawn at least once.

Overall, the probability of failure is at most 1/6 + 1/6 = 1/3 by union bound (where the first 1/6 is the failure rate of original mechanism, and the other 1/6 is the failure rate of the coupon-collector).

Theorem 4.1 (Advanced Grouposition for Pure LDP. [BNS19, Theorem 4.2]). Let X ∈ X^n, X’ ∈ X^n differ in at most k entries for some 1 ≤ k ≤ n. Let A = (R_1, . . . , R_k) : X^n → Y, where each R_i is ε-LDP. Then for every δ > 0 and ε’ = kε^2/2 + ε√2k ln 1/δ, we have:

\[ \Pr_{y \sim A(X)} \left[ \ln \frac{Pr[A(X) = y]}{Pr[A(X’) = y]} > \epsilon’ \right] \leq \delta. \]

In particular, for every δ > 0 and every set T ⊆ Y, we have Pr[A(X) ∈ T] ≤ e^{ε’} Pr[A(X’) ∈ T] + δ.

Proof. Without loss of generality, we assume that X, X’ differ in the first k entries. We begin with the privacy loss random variable of A(X), A(X’):

\[ L_{A(X), A(X’)} = \ln \frac{Pr[A(X) = y]}{Pr[A(X’) = y]} = \sum_{i=1}^{k} \ln \frac{Pr[R_i(x_i) = y_i]}{Pr[R_i(x’_i) = y_i]} = \sum_{i=1}^{k} L_{R_i(x_i), R_i(x’_i)} \]
By taking expectation of $L_{R_i(x_i), R_i(x'_i)}$, it becomes KL-divergence. Which now we can apply the Proposition 3.3 of [BS16] (taking $\alpha = 1$) and acquire:

$$\mathbb{E} [L_{A(X), A(X')}] = \sum_{i=1}^{k} \mathbb{E} [L_{R_i(x_i), R_i(x'_i)}] \leq \frac{k \varepsilon^2}{2}.$$  

Since $R_i$ are $\varepsilon$-LDP randomizers, we have $L_{R_i(x_i), R_i(x'_i)} \in [-\varepsilon, \varepsilon]$. Then by Hoeffding’s inequality, for every $t > 0$:

$$\exp \left( \frac{-t^2}{2k\varepsilon^2} \right) \geq \Pr \left[ \sum_{i=1}^{k} L_{R_i(x_i), R_i(x'_i)} > \sum_{i=1}^{k} \mathbb{E} [L_{R_i(x_i), R_i(x'_i)}] + t \right]$$

$$= \Pr \left[ L_{A(X), A(X')} > \sum_{i=1}^{k} \mathbb{E} [L_{R_i(x_i), R_i(x'_i)}] + t \right]$$

$$\geq \Pr \left[ L_{A(X), A(X')} > k\varepsilon^2/2 + t \right] \quad \text{(Since $\mathbb{E} [L_{R_i(x_i), R_i(x'_i)}] \leq \frac{1}{2} \varepsilon^2$)}.$$  

Thus, we can choose $t = \varepsilon \sqrt{2k \ln (1/\delta)}$ such that $\delta = \exp (-t^2/2k\varepsilon^2)$, which completes the proof.

Advanced group position is a useful in the sense that it presents the implication from LDP to DP. We can see that as an analogue of [DMNS06, Advanced Composition Theorem] as they share similar proof techniques.

**Theorem 4.2** (Advanced Group Position for Approximate LDP. [BNS19, Theorem 4.3]). Let $X \in \mathcal{X}^n$, $X' \in \mathcal{X}^n$ differ in at most $k$ entries for some $1 \leq k \leq n$, Let $A = (R_1, \ldots, R_n) : \mathcal{X}^n \rightarrow \mathcal{Y}$ be $(\varepsilon, \delta)$-LDP. Then for every $\delta' > 0$ and $\varepsilon' = k\varepsilon^2/2 + \varepsilon \sqrt{2k \ln 1/\delta'}$, and every set $T \subseteq \mathcal{Y}$, we have: $\Pr[A(X) \in T] \leq e^{\varepsilon'} \Pr[A(X') \in T] + \delta + k\delta'$.

The following is a composition of LDP randomizers.

**Lemma 4.5** (Composition of $(\varepsilon, \delta)$-LDP Randomizers. [EFMRST20, Lemma A.1]). Assume that for every $(\varepsilon_1, \delta_1)$-LDP randomizer $Q_1 : \{0, 1\} \rightarrow \{0, 1\}$ and every $(\varepsilon_2, \delta_2)$-LDP randomizer $Q_2 : \{0, 1\} \rightarrow \{0, 1\}$ we have that $Q_2 \circ Q_1$ is a $(\varepsilon, \delta)$-LDP randomizer. Then for every $(\varepsilon_1, \delta_1)$-LDP randomizer $R_1 : \mathcal{X} \rightarrow \mathcal{Y}$ and $(\varepsilon_2, \delta_2)$-LDP randomizer $R_2 : \mathcal{Y} \rightarrow \mathcal{Z}$, we have that $R_2 \circ R_1$ is an $(\varepsilon, \delta)$-LDP randomizer.

**Proof.** Let $R_1 : \mathcal{X} \rightarrow \mathcal{Y}$ be a $(\varepsilon_1, \delta_1)$-LDP randomizer and $R_2 : \mathcal{Y} \rightarrow \mathcal{Z}$ be a $(\varepsilon_2, \delta_2)$-LDP randomizer.

We will prove this by contradiction. First, assume that for some $(\varepsilon, \delta)$, there exists an event $S \subseteq \mathcal{Z}$ such that for some $x, x'$:

$$\Pr[R_2(R_1(x)) \in S] > e^{\varepsilon} \Pr[R_2(R_1(x')) \in S] + \delta. \quad (1)$$

Then, we will show that there exists an $(\varepsilon_1, \delta_1)$-LDP local randomizer $Q_1 : \{0, 1\} \rightarrow \{0, 1\}$ and $(\varepsilon_2, \delta_2)$-LDP local randomizer $Q_2 : \{0, 1\} \rightarrow \{0, 1\}$ such that:

$$\Pr[Q_2(Q_1(x)) = 1] - e^{\varepsilon} \Pr[Q_2(Q_1(x')) = 1] > \delta,$$

which violates the assumption (that $Q_2 \circ Q_1$ is $(\varepsilon, \delta)$-LDP) of the lemma.
Let:

\[ y_0 := \operatorname{arg\,min}_{y \in \mathcal{Y}} \Pr[R_2(y) \in S], \]

\[ y_1 := \operatorname{arg\,max}_{y \in \mathcal{Y}} \Pr[R_2(y) \in S], \]

\[ P_t := \{ y \in \mathcal{Y} \mid \Pr[R_1(x) = y] - \epsilon^2 \Pr[R_1(x') = y] > 0 \}. \]

By using \( y_0, y_1 \) and our assumption, we can construct the following:

\[ \left( \Pr[R_1(x) \notin P_t] - \epsilon^2 \Pr[R_1(x') \notin P_t] \cdot \Pr[R_2(y_0) \in S] \right) \]

\[ + \left( \Pr[R_1(x) \in P_t] - \epsilon^2 \Pr[R_1(x') \in P_t] \cdot \Pr[R_2(y_1) \in S] \right) \geq \sum_{y \in \mathcal{Y}} \left( \Pr[R_2(y) \in S] \cdot \left( \Pr[R_1(x) = y] - \epsilon^2 \Pr[R_1(x') = y] \right) \right) \]

\[ = \sum_{y \in \mathcal{Y}} \Pr[R_2(y) \in S \land R_1(x) = y] - \epsilon^2 \sum_{y \in \mathcal{Y}} \Pr[R_2(y) \in S \land R_1(x') = y] \]

\[ = \Pr[R_2(R_1(x)) \in S] - \epsilon^2 \Pr[R_2(R_1(x')) \in S] > \delta. \]

Now we can define \( Q_1(0) := 1 \{ R_1(x) \in P_t \} \) and \( Q_1(1) := 1 \{ R_1(x') \in P_t \} \), where \( 1 \{ \cdot \} \) is the indicator function. Since we defined \( Q_1 \) by restraining the set of inputs and post-processing the output, thus \( Q_1 \) is a \((\epsilon_1, \delta_1)\)-LDP randomizer as well.

Next, we can define \( b \in \{0, 1\}, Q_2(b) := 1 \{ R_2(y_b) \in S \} \). It is also easy to see that \( Q_2 \) is a \((\epsilon_2, \delta_2)\)-LDP randomizer. By using the previous result, we have:

\[ \Pr[Q_2(Q_1(x)) = 1] - \epsilon^2 \Pr[Q_2(Q_1(x')) = 1] \]

\[ = \sum_{b \in \{0, 1\}} \left( \Pr[Q_2(b) = 1] \cdot \left( \Pr[Q_1(0) = b] - \epsilon^2 \Pr[Q_1(1) = b] \right) \right) \]

\[ = \left( \Pr[R_1(x) \notin P_t] - \epsilon^2 \Pr[R_1(x') \notin P_t] \cdot \Pr[R_2(y_0) \in S] \right) \]

\[ + \left( \Pr[R_1(x) \in P_t] - \epsilon^2 \Pr[R_1(x') \in P_t] \cdot \Pr[R_2(y_1) \in S] \right) > \delta. \]

which is what we want to show for contradiction. \( \square \)

Following corollary identifies the value of \( \epsilon \) for \( R_2 \circ R_1 \).

**Corollary 4.1** (Followed from Lemma 3.5 Corollary A.2). For every \( \epsilon_2 \)-LDP randomizer \( R_1 : \mathcal{X} \rightarrow \mathcal{Y} \) and every \( \epsilon_2 \)-LDP randomizer \( R_2 : \mathcal{Y} \rightarrow \mathcal{Z} \), we have that \( R_2 \circ R_1 \) is a \( \epsilon \)-LDP randomizer, where \( \epsilon := \ln \left( \frac{e^{\epsilon_2^2+\epsilon_2}+1}{e^{\epsilon_2^2}+1} \right) \). In addition, if \( R_1 \) is removal \( \epsilon_1 \)-LDP, then \( R_2 \circ R_1 \) is a removal \( \epsilon \)-LDP.

**Proof.** We can reduce the problem to \( \mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\} \) using Lemma 3.5. We choose to bound the case of \( R_2 \circ R_1(b) = 1 \) for \( b \in \{0, 1\} \):

\[ \frac{\Pr[R_2 \circ R_1(0) = 1]}{\Pr[R_2 \circ R_1(1) = 1]} = \frac{\Pr[R_1(0) = 0] \cdot \Pr[R_2(0) = 1] + \Pr[R_1(0) = 1] \cdot \Pr[R_2(1) = 1]}{\Pr[R_1(1) = 0] \cdot \Pr[R_2(0) = 1] + \Pr[R_1(1) = 1] \cdot \Pr[R_2(1) = 1]}. \]

Defining \( p_0 := \Pr[R_1(0) = 0], p_1 := \Pr[R_1(1) = 0], \alpha := \Pr[R_2(0) = 1] / \Pr[R_2(1) = 1] \), the above expression simplifies to:

\[ \frac{1 + (\alpha - 1)p_0}{1 + (\alpha - 1)p_1} \]
Since we want to find \( \varepsilon \) such that:
\[
\Pr[R_2 \circ R_1(0) = 1] \leq e^\varepsilon,
\]
we came to solve the following problem:
\[
\max_{p_0, p_1} \frac{1 + (\alpha - 1)p_0}{1 + (\alpha - 1)p_1} \quad \text{s.t.} \quad \alpha = e^{\varepsilon^2},
\]
\[
p_0 > p_1,
\]
\[
e^{-\varepsilon^1} \leq \frac{p_0}{p_1} \leq e^{\varepsilon^1},
\]
\[
e^{-\varepsilon^1} \leq \frac{1 - p_0}{1 - p_1} \leq e^{\varepsilon^2}.
\]
Note that we can assume without loss of generality that \( \alpha \geq 1 \) and thus the expression is maximized while \( \alpha = e^{\varepsilon^2} \) and \( p_0 > p_1 \). Solving this optimization problem yields that \( \varepsilon := \ln \left( \frac{e^{\varepsilon^1} + e^{\varepsilon^2}}{e^{\varepsilon^1} + e^{\varepsilon^2}} \right) \).

The deletion LDP case is achieved if we substitute \( R \in \delta(x') \) for every dataset \( x \). We also give a weaker result of \( \varepsilon \) and show that the bound is tight. They also gives \( \varepsilon := \ln \left( \frac{e^{\varepsilon^1} + e^{\varepsilon^2}}{e^{\varepsilon^1} + e^{\varepsilon^2}} \right) \). See proof of Theorem 4.3 [BNS19, Theorem 6.1]. Let \( \varepsilon \in (0, 1/4] \) and \( \delta \in [0, 1] \), and suppose \( R_1, \ldots, R_n \) are the \((\varepsilon, \delta)\)-LDP randomizers. Then there exists a public-coin protocol with randomizers \( R'_1, \ldots, R'_n \) such that, for every choice of integer
\[
5 \ln \frac{1}{\varepsilon} \leq T \leq \frac{1}{4\delta n e^\varepsilon}
\]
1. \( R'_i \) is \( 10\varepsilon \)-LDP for every \( i \in [n] \)
2. for every dataset \( x = (x_1, \ldots, x_n) \in X^n \), the total variation distance between the distribution \( D_x \) of messages \( (R_1(x_1), \ldots, R_n(x_n)) \) and the distribution \( D'_x \) of \( (R'_1(x_1), \ldots, R'_n(x_n)) \) satisfies
\[
TV(D_x, D'_x) \leq n \left( \left( \frac{1}{2} + \varepsilon \right)^T + 6T \delta \cdot \frac{e^\varepsilon}{1 - e^{-\varepsilon}} \right)
\]
3. Each user sends at most \( \log_2 T \) bits of communication.
4. The protocol uses \( T \sum_{i=1}^n r_i \) bits of public randomness, where \( r_i \) is the number of random bits required by \( R_i \).

5 Relation between Shuffle DP and LDP

**Definition 5.1** (One-Message Shuffle Model). In the one-message shuffled model, each user sends \( m = 1 \) message.

**Theorem 5.1** (From One-Message Shuffle Model to LDP). [CSUZZ19, Theorem 6.2]. If \( P_n = (R, S, A) \) is a one-message shuffled model with \( n \in \mathbb{N} \) users that satisfies \((\varepsilon_S, \delta_S)\)-DP, then the local randomizer \( R \) satisfies \((\varepsilon_L, \delta_L)\)-LDP for \( \varepsilon_L = \varepsilon_S + \ln n \) and \( \delta_L = \delta_S \). Therefore, the symmetric local protocol \( P_L = (R, A \circ S) \) satisfies \((\varepsilon_L, \delta_L)\)-DP.
Proof. Let $E_{R, \delta_S}$ denote the set of parameters $\epsilon' > 0$ for which $R: \mathcal{X} \rightarrow \mathcal{Y}$ is not $(\epsilon', \delta_L)$-LDP, and let $\tilde{\epsilon}$ denote the supremum of $E_{R, \delta_S}$. We can assume without loss of generality that $\epsilon \in E_{R, \delta_S}$ (otherwise, we can use the argument below with $\epsilon' := \epsilon(1 - \alpha)$ for some small $\alpha$ and take the limit as $\alpha \downarrow 0$ in the end).

$$
\Pr[R(x') \in Y] > e^{\epsilon} \Pr[R(x) \in Y] + \delta_L \tag{2}
$$

for some $x, x' \in \mathcal{X}$ and $Y \subseteq \mathcal{Y}$. We define $p' := \Pr[R(x') \in Y]$ and $p := \Pr[R(x) \in Y]$ for brevity, so that the above becomes $p' > e^{\epsilon} p + \delta_S$.

Then we define the set $\mathcal{W} := \{W \in \mathcal{Y}^n \mid \exists i, w_i \in Y\}$, which is a set of output for $P$ where any of its randomizer is not local differentially private. Construct two databases of size $n$: $X := (x, x, \ldots, x)$ and $X' := (x', x, \ldots, x)$.

Since $P_n$ is $(\epsilon, \delta_S)$-differentially private, we can write:

$$
\Pr[P_n(X') \in W] \leq e^{\epsilon} \Pr[P_n(X) \in W] + \delta_S \tag{3}
$$

Now we have:

$$
\Pr[P_n(X) \in W] = \Pr[S(R(x), \ldots, R(x)) \in W]
= \Pr[(R(x), \ldots, R(x)) \in W]
= \Pr[\exists i, R(x) \in Y]
\leq n \Pr[R(x) \in Y] = np, \tag{4}
$$

where the second equality is because $\mathcal{W}$ is closed under permutation, and the inequality results from the union bound.

Similarly, we have:

$$
\Pr[P_n(X') \in W] \geq \Pr[R(x') \in Y] = p'
> e^{\epsilon} \cdot p + \delta_S
$$

where the last inequality comes from (2). Finally we can rewrite (3):

$$
e^{\epsilon} p + \delta_S < \Pr[P_n(X') \in W] \leq e^{\epsilon} \Pr[P_n(X) \in W] + \delta_S
\leq e^{\epsilon} np + \delta_S,
$$

Finally we obtain:

$$
\epsilon < \epsilon + \ln n.
$$

This means that the largest possible $\epsilon$ for which $R$ is not $(\epsilon, \delta_S)$-DP is strictly smaller than $\epsilon_L := \epsilon_S + \ln n$. So $R$ is $(\epsilon_L, \delta_S)$-DP. \hfill \Box

Note that the above proof applies to both public-coin and private-coin protocols, as the only key step where the possibly joint randomization is taken into account is (3). It is not clear whether the analysis can be significantly improved for private-coin protocols, for which $np$ can be replaced by $1 - (1 - p)^n$ in the RHS of (4).

**Theorem 5.2** (From LDP to shuffler. \cite[Theorem 3.2]{FMT22}). Let $R$ be an $(\epsilon_L, \delta)$-differentially private local randomizer. Let $S = (R, \pi)$ be the shuffler that given a dataset, samples a uniform random permutation $\pi$ over $[n]$, then sequentially compute reports $z_i = R^{(i)}(z_{i-1}, x_{\pi(i)})$ for $i \in [n]$ and outputs $z_{1:n}$. Then for any $\delta \in [0, 1]$ such that $\epsilon_L \leq \ln \left( \frac{n}{8 \ln(2/\delta)} - 1 \right)$. $S$ is $(\epsilon, \delta)$-DP, where:

$$
\epsilon \leq \ln \left( 1 + 4(e^{\epsilon_L} - 1) \left( \sqrt{\frac{2 \ln(4/\delta)}{(e^{\epsilon_L} + 1)n}} + \frac{1}{n} \right) \right) \tag{5}
$$

\hfill 10
In particular, for $\varepsilon_L \leq 1$, this gives $\varepsilon = O\left(\varepsilon_L \sqrt{\frac{\log(1/\delta)}{m}}\right)$, and for $\varepsilon_L \geq 1$ we get $\varepsilon = O\left(\varepsilon_L^{1/2} \sqrt{\frac{\log(1/\delta)}{m}}\right)$.

We redirect readers to their work for the proof, and the extension when the local randomizer is itself approximate LDP.

On a side note, above theorem holds true for robust shuffle privacy if we replace $n = \gamma n$ for $\gamma \in (0, 1]$ being the fraction of users that actually follows the protocol (i.e., non-malicious). Indeed, this follows by applying exactly the same argument (amplification by shuffling) as in the proof of the theorem, but only to the $n' := \gamma n$ honest users.

Finally, it is worth noting that the original LDP guarantee does of course still hold after shuffling; and so in particular the output of the shuffler is both $(\varepsilon, \delta)$-DP and $\varepsilon_L$-DP.

6 Privacy Amplification by Subsampling

The chapter "Composition of Differential Privacy & Privacy Amplification by Subsampling" [Ste22 Section 6] written by Thomas Steinke contains more in-depth proofs and claims on this topic. Here, we state a few that is important and provide proofs in alternative methods (which is somewhat simpler and weaker).

**Theorem 6.1** (Privacy Amplification by Subsampling for Approximate DP. [Ste22 Theorem 28]). Let $U \subseteq [n]$ be a random subset and $X$ be a dataset with $n$ records. Define $M: X^n \rightarrow Y$ to satisfy $(\varepsilon, \delta)$-DP. Construct a mechanism $M_U: X^n \rightarrow Y$ where it first samples $|U|$ records from $X$ to form the smaller dataset $X_U$ and then compute $M(X_U)$. We further let $p = \max_{i \in [n]} \Pr[U[i \in U]]$. Then $M_U$ is $(\varepsilon', \delta')$-DP for $\varepsilon' = \ln(1 + p(e^\varepsilon - 1))$ and $\delta' = p\delta$.

A more general version of this theorem is defined in [Ste22 Theorem 28]. The complete proof is available in [Ste22], but here we present a self-contained version for the case of $p = m/n$, using the same notation as lecture notes of Jonathan Ullman.\(^1\)

**Proof.** For datasets $X \sim X'$, let $p = \frac{m}{n} = \Pr[i \in U]$. The goal is to show that $M_U$ is $(\ln(1 + p(e^\varepsilon - 1)), p\delta)$-DP. Fix any measurable $S \subseteq Y$: to show the result, we need to upper bound the following ratio by $1 + p(e^\varepsilon - 1)$:

$$\frac{\Pr[M_U(X) \in S] - \delta p}{\Pr[M_U(X') \in S]} = \frac{\Pr[M(X_U) \in S] - \delta p}{\Pr[M(X'_U) \in S]} = \frac{\Pr[M(X_U) \in S | i \in U]p + \Pr[M(X_U) \in S | i \notin U](1 - p) - \delta p}{\Pr[M(X'_U) \in S | i \in U]p + \Pr[M(X'_U) \in S | i \notin U](1 - p) - \delta p}$$

For convenience, we set:

$$C = \Pr[M(X_U) \in S | i \in U]$$
$$C' = \Pr[M(X'_U) \in S | i \in U]$$
$$D = \Pr[M(X_U) \in S | i \notin U] = \Pr[M(X'_U) \in S | i \notin U]$$

Now we have:

$$\frac{\Pr[M_U(X) \in S] - \delta p}{\Pr[M_U(X') \in S]} = \frac{pC + (1 - p)D - p\delta}{pC' + (1 - p)D}$$

\(^1\)From CS7880 Homework 1 Solution: [http://www.ccs.neu.edu/home/jullman/cs7880s17/HW1sol.pdf](http://www.ccs.neu.edu/home/jullman/cs7880s17/HW1sol.pdf) (Accessed Nov 21, 2022).
\[ \Pr[M_U(X) \in S] - \delta p = pC + (1 - p)D - p\delta \]
\[ \leq p(e^\varepsilon \min(C', D) + \delta) + (1 - p)D - p\delta \]
\[ = p(\min(C', D) + (e^\varepsilon - 1) \min(C', D) + \delta) + (1 - p)D - p\delta \]
\[ \leq p(C' + (e^\varepsilon - 1)(pC' + (1 - p)D) + \delta) + (1 - p)D - p\delta \]
\[ = p(C' + (e^\varepsilon - 1)(pC' + (1 - p)D)) + (1 - p)D \]
\[ = (pC' + (1 - p)D) \times (p(e^\varepsilon - 1) + 1) \]
\[ = \exp(\ln(p(e^\varepsilon - 1) + 1))(pC' + (1 - p)D) \]

The first inequality follows from the definition of \((\varepsilon, \delta)\)-DP. Note that \(C \leq e^\varepsilon D + \delta\) because \(X_U\) with \(x_i\) is still neighboring to subset of \(X\) of size \(U\) without \(x_i\). The second inequality uses the fact that \(\min(x, y) \leq \alpha x + (1 - \alpha)y, \forall \alpha \in [0, 1]\), where we use \(\alpha = p\). This simpler proof gets the same bound as [Ste22] by applying \(e^{\ln}\) on the second-to-last line.

From above, we finally get:
\[ \frac{\Pr[M_U(X) \in S] - \delta p}{\Pr[M_U(X') \in S]} \leq \exp(\ln(p(e^\varepsilon - 1) + 1)) \]

which concludes the proof. \(\square\)

References

[ACFT19] Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi. “Test without Trust: Optimal Locally Private Distribution Testing”. In: The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan. Vol. 89. Proceedings of Machine Learning Research. PMLR, 2019, pp. 2067–2076.

[ASZ18] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. “Differentially Private Testing of Identity and Closeness of Discrete Distributions”. In: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada. 2018, pp. 6879–6891.

[BNS19] Mark Bun, Jelani Nelson, and Uri Stemmer. “Heavy Hitters and the Structure of Local Privacy”. In: ACM Trans. Algorithms 15.4 (2019), 51:1–51:40. doi:10.1145/3344722

[BS16] Mark Bun and Thomas Steinke. “Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds”. In: Lecture Notes in Computer Science 9985 (2016), pp. 635–658. doi:10.1007/978-3-319-53641-4_24

[Che21] Albert Cheu. “Differential Privacy in the Shuffle Model: A Survey of Separations”. In: CoRR abs/2107.11839 (2021).

[CSUZZ19] Albert Cheu, Adam D. Smith, Jonathan R. Ullman, David Zeber, and Maxim Zhilyaev. “Distributed Differential Privacy via Shuffling”. In: Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019. Proceedings, Part I. Vol. 11476. Lecture Notes in Computer Science. Springer, 2019, pp. 375–403. doi:10.1007/978-3-030-17653-2_13

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. “Calibrating noise to sensitivity in private data analysis”. In: Theory of cryptography. Vol. 3876. Lecture Notes in Comput. Sci. Springer, Berlin, 2006, pp. 265–284. doi:10.1007/11681878_14 url:https://doi.org/10.1007/11681878_14
7 Appendix

7.1 Proof of the weaker result in Corollary 4.1

Recall that the goal is to prove:

\[
\frac{1 + e^{x+y}}{e^x + e^y} \leq \exp \left( \frac{xy}{2} \right).
\]

Proof. Suppose w.l.o.g that 0 ≤ x ≤ y. Consider the random variable Z such that:

\[
\Pr[Z = x] = \frac{e^y}{e^x + e^y} \quad \text{and} \quad \Pr[Z = -x] = \frac{e^x}{e^x + e^y}.
\]

Then |Z| ≤ x and E[Z] = x · \frac{e^y - e^{-x}}{e^x + e^y}. We also have, on the one hand, that E[\exp(Z)] = \frac{1 + e^{x+y}}{e^x + e^y}.

This proof is due to Thomas Steinke. See here for the proof in emoji flavor.
On the other hand, by Hoeffding’s Lemma with $\lambda = 1$:

$$
\mathbb{E}[e^Z] \leq \exp \left( \mathbb{E}[Z] + \frac{(2x)^2}{8} \right) = \exp \left( x \cdot \frac{e^y - e^x}{e^y + e^x} + \frac{x^2}{2} \right) = \exp \left( x \cdot \frac{e^{y-x} - 1}{e^{y-x} + 1} + \frac{x^2}{2} \right).$

Then we apply this inequality: for every $t \geq 0$, $\frac{e^t - 1}{e^t + 1} \leq \frac{t}{2}$. With $t := y - x$ and substituting the above result, we get

$$
1 + \frac{e^{x+y}}{e^x + e^y} \leq \exp \left( x \cdot \frac{e^{y-x} - 1}{e^{y-x} + 1} + \frac{x^2}{2} \right) \leq \exp \left( x \cdot \frac{(y-x)}{2} + \frac{x^2}{2} \right) = \exp \left( \frac{xy}{2} \right).
$$

---

3Hoeffding’s Lemma: Let $X$ be any real-valued random variable such that $X \in [a, b]$ (w. prob 1). Then, for all $\lambda \in \mathbb{R}$,

$$
\mathbb{E}[e^{\lambda X}] \leq \exp \left( \lambda \mathbb{E}[X] + \frac{\lambda^2 (b-a)^2}{8} \right)
$$

4This can be proven by the concavity of $f: t \rightarrow \frac{e^t - 1}{e^t + 1}$ on $[0, \infty)$, so $f'(t) \leq f'(0)$, so $f(t) \leq f(0) + f'(0)t = \frac{t}{2}$. 

14