A Novel Approach to Differential Privacy with Alpha Divergence

Yifeng Liu
Department of Electrical and Computer Engineering
The University of British Columbia, Vancouver
Vancouver, Canada
lyf666@student.ubc.ca

Zehua Wang*
Department of Electrical and Computer Engineering
The University of British Columbia, Vancouver
Vancouver, Canada
zwang@ece.ubc.ca

June 23, 2025

Abstract

As data-driven technologies advance swiftly, maintaining strong privacy measures becomes progressively difficult. Conventional (ϵ, δ)-differential privacy, while prevalent, exhibits limited adaptability for many applications. To mitigate these constraints, we present alpha differential privacy (ADP), an innovative privacy framework grounded in alpha divergence, which provides a more flexible assessment of privacy consumption. This study delineates the theoretical underpinnings of ADP and contrasts its performance with competing privacy frameworks across many scenarios. Empirical assessments demonstrate that ADP offers enhanced privacy guarantees in small to moderate iteration contexts, particularly where severe privacy requirements are necessary. The suggested method markedly improves privacy-preserving methods, providing a flexible solution for contemporary data analysis issues in a data-centric environment.

1 Introduction

In the modern data-centric age, protecting individual privacy has emerged as a critical issue for researchers, practitioners, and legislators. Conventional data security techniques frequently fail to maintain an optimal equilibrium between data utility and personal privacy. Differential privacy, initiated by Dwork et al. [1], has gained importance as a standard for privacy-preserving data analysis, supported by strong theoretical guarantees.

The fundamental premise of differential privacy is indistinguishability, which guarantees that the results of studies on datasets differing by one individual are statistically indistinguishable [2]. This essential attribute has enabled the extensive implementation of differential privacy in multiple fields, such as machine learning, statistical analysis, and data mining [3], [4], [5]. Despite various follow-up studies [6], [7] focused on enhancing its efficacy, the traditional (ϵ, δ)-differential privacy framework may still be insufficiently adaptable to meet more nuanced privacy demands or to effectively reconcile the trade-offs between privacy and utility.

Researchers have been actively seeking privacy frameworks that provide improved privacy estimates to overcome these limitations. This paper presents alpha differential privacy (ADP), a privacy architecture based on alpha divergence, a comprehensive set of metrics that assess the dissimilarity across probability distributions [8, 9]. This method provides a more flexible framework that may be customized for various application scenarios and sensitivity levels. ADP enhances the traditional differential privacy framework by leveraging the flexibility of alpha divergence, providing a range of privacy guarantees tailored to particular requirements.

Our research enhances the existing body of knowledge in multiple important aspects: we provide a robust theoretical framework for ADP, clarify its essential characteristics, and demonstrate its robustness and composability, comparable to other privacy frameworks. Empirical evaluations indicate that ADP offers significant advantages over several prominent

privacy frameworks in contexts with a restricted number of iterations or strict privacy requirements, ADP attains a reduced initial privacy consumption with a decent privacy consumption growth in these contexts, underscoring its applicability in privacy-preserving data analysis, while also recognizing its limitations in other contexts.

The organization of the subsequent sections of this paper is as follows: Section II outlines differential privacy and its fundamental principles. Section III examines the relevant literature on differential privacy and divergence metrics. Section IV provides a formal description of alpha differential privacy and examines its principal properties. Section V examines the relationship between alpha differential privacy and approximate differential privacy. Section VI examines diverse mechanisms of alpha differential privacy, including experimental data and analysis that demonstrate its application and versatility. Section VII provides guidance on how to pick an appropriate α for a given task. Section VIII presents the simulation settings, results, and an in-depth discussion. Section IX concludes the paper with a summary of findings and outlines prospective future research directions.

This work aims to advance the theoretical and practical understanding of alpha differential privacy, hence contributing to the development of more resilient and flexible privacy systems. This guarantees that privacy-preserving data analysis continues to be a feasible and powerful pursuit in a progressively data-driven world.

2 Differential Privacy and Divergence

Differential privacy (DP) was initially introduced by Dwork et al. [2] and has since established itself as the fundamental principle of privacy-preserving data analysis. It is a stringent mathematical framework that offers a formal notion of privacy for data analysis methods. The fundamental concept of differential privacy is to guarantee that the addition or removal of an individual from a dataset does not substantially influence the results of an algorithm. This inhibits an assailant from readily deducing any person's information.

2.1 Pure Differential Privacy

Definition 1 (Pure differential privacy [10]). *Let $\mathcal{D}$ denote a set of all possible datasets, and let $D_1, D_2 \in \mathcal{D}$ be two datasets that differ by exactly one element, denoted as $D_1 \sim D_2$. A randomized mechanism $\mathcal{M}$ that maps datasets to outputs in some range $\mathcal{R}$ is said to be pure differential privacy, or ϵ-differential is defined as:*

$$\Pr[\mathcal{M}(D_1) \in \mathcal{R}] \le e^\epsilon \Pr[\mathcal{M}(D_2) \in \mathcal{R}]. \tag{1}$$

The parameter ϵ regulates the degree of privacy. The selection of ϵ entails a compromise between privacy preservation and data utility. Lower values of ϵ enhance privacy guarantees but diminish the usefulness of the mechanism $\mathcal{M}$, while higher values of ϵ augment the utility of the mechanism $\mathcal{M}$ but compromise privacy safeguards. Standard values of ϵ vary from 0.01 to 1, contingent upon the particular privacy stipulations and objectives of the data analysis.

2.2 Approximate Differential Privacy

Definition 2 (Approximate differential privacy [11]). *Let $\mathcal{D}$ denote a set of all possible datasets, and let $D_1, D_2 \in \mathcal{D}$ be two datasets that differ by exactly one element, denoted as $D_1 \sim D_2$. A randomized mechanism $\mathcal{M}$ that maps datasets to outputs in some range $\mathcal{R}$ is said to be approximate differential privacy, or (ϵ, δ)-differential privacy if:*

$$\Pr\left(\Pr[\mathcal{M}(D_1) \in \mathcal{R}] > e^\epsilon \Pr[\mathcal{M}(D_2) \in \mathcal{R}]\right) \le \delta. \tag{2}$$

Approximate differential privacy permits a minor failure possibility δ of breaching the privacy promise. This flexibility enables enhanced utility in data analytic jobs while preserving robust privacy safeguards. Approximate differential privacy is particularly advantageous in situations where the stringent condition of pure differential privacy ($\delta = 0$) is excessively limiting.

The inclusion of the δ parameter recognizes that in real-world applications, attaining complete privacy is frequently unfeasible or impracticable. By allowing a minimal risk of failure, (ϵ, δ)-differential privacy strikes a balance between privacy and usefulness, rendering it a flexible and extensively utilized privacy framework.

2.3 Sensitivity and Noise Addition

A key concept in designing differential private mechanisms is the sensitivity of a function f, defined as the maximum change in f's output when its input dataset changes by one element.

Definition 3 (Sensitivity). *For a function $f : \mathcal{D} \to \mathbb{R}^k$, the sensitivity Δf is defined as:*

$$\Delta f \triangleq \sup_{D_1, D_2 : D_1 \sim D_2} \|f(D_1) - f(D_2)\|_p, \tag{3}$$

where $\|\cdot\|_p$ denotes the ℓ_p norm. The choice of p varies on the mechanism.

To achieve differential privacy, noise proportional to the sensitivity of the function is added to the output. Common mechanisms include the Laplace mechanism and the Gaussian mechanism.

2.3.1   Laplace Mechanism

For a function $f : D \to \mathbb{R}^k$, the Laplace mechanism $\mathcal{M}_L$ ensures ϵ-differential privacy by adding noise drawn from the Laplace distribution to the output of the function [1]. Formally, the mechanism $\mathcal{M}_L$ is defined as:

$$\mathcal{M}_L(D) \triangleq f(D) + \text{Lap}(0, b), \tag{4}$$

where $\text{Lap}(0, b)$ denotes the Laplace distribution with mean 0 and scale parameter b.

The Laplace mechanism ensures $(\epsilon, 0)$-differential privacy when:

$$b = \frac{\Delta f_1}{\epsilon} \tag{5}$$

Here, the sensitivity Δf_1 applies ℓ_1 norm, which is the sum of the absolute differences between the corresponding elements of the vectors, this is also called ℓ_1 sensitivity.

$$\Delta f_1 \triangleq \sup_{D_1, D_2 : D_1 \sim D_2} \|f(D_1) - f(D_2)\|_1, \tag{6}$$

2.3.2   Gaussian Mechanism

For a function $f : D \to \mathcal{R}^k$, the Gaussian mechanism $\mathcal{M}_G$ ensures (ϵ, δ)-differential privacy by adding noise drawn from the Gaussian distribution to the output of the function [11]. Formally, the mechanism $\mathcal{M}_G$ is defined as:

$$\mathcal{M}_G(D) \triangleq f(D) + \mathcal{N}(0, \sigma^2), \tag{7}$$

where $\mathcal{N}(0, \sigma^2)$ denotes the normal distribution with mean 0 and variance σ^2.

The parameter σ is chosen based on ϵ, δ, and Δf to satisfy (ϵ, δ)-differential privacy:

$$\sigma^2 = \frac{2 \log(1.25/\delta) \Delta f_2^2}{\epsilon}. \tag{8}$$

Here, the sensitivity Δf_2 of the function f applies ℓ_2 norm, this is also called ℓ_2 sensitivity.

$$\Delta f_2 \triangleq \sup_{D_1, D_2 : D_1 \sim D_2} \|f(D_1) - f(D_2)\|_2. \tag{9}$$

The Gaussian mechanism is favoured for employing the ℓ_2 norm as its measure of sensitivity. The ℓ_2 norm consolidates the squared differences, which tends to mitigate bigger fluctuations, resulting in a diminished overall sensitivity score. Therefore, the noise introduced by the Gaussian mechanism is generally less intrusive than the noise necessitated by systems reliant on ℓ_1 sensitivity, such as the Laplace mechanism. Furthermore, the Gaussian distribution's bell-shaped curve guarantees that the majority of the added noise is focused around the mean (zero), while the likelihood of extreme noise values decreases swiftly. This characteristic frequently yields outputs that approximate the actual function value more closely, enhancing the usefulness of the disseminated data.

2.4   Divergence Measures

Divergence measures are mathematical instruments employed to assess the disparity between probability distributions. Divergence metrics can be utilized within differential privacy to define and assess the privacy consumption resulting from a randomized method. Frequently utilized divergence metrics encompass Kullback–Leibler divergence, maximum divergence, and Rényi divergence.

Definition 4 (Kullback–Leibler Divergence and Max Divergence). *Let P and Q be two probability measures over a measurable space $(\mathcal{X}, \mathcal{F})$ with property $P \ll Q$. By definition, $P \ll Q$ (absolute continuity of P with respect to Q) means that for any measurable set $A \in \mathcal{F}$, if $Q(A) = 0$, then $P(A) = 0$ as well. Throughout this paper, the notation $\ll$ will consistently refer to absolute continuity. The Kullback–Leibler (KL) divergence from Q to P is defined as:*

$$
\begin{aligned}
D_{KL}(P \parallel Q) &\triangleq \int_{\mathcal{X}} \log\left(\frac{dP}{dQ}\right) dP \\
&= \mathbb{E}_P\left[\log\left(\frac{dP}{dQ}\right)\right].
\end{aligned}
\tag{10}
$$

The max divergence is defined as:

$$
D_{\infty}(P \parallel Q) \triangleq \log(\operatorname*{ess\,sup}_{\mathcal{X}} \frac{dP}{dQ}),
\tag{11}
$$

where $\frac{dP}{dQ}$ is the Radon-Nikodym derivative of P with respect to Q, and $\mathbb{E}_P[\cdot]$ denotes the expectation with respect to the probability measure P.

It is easy to see that the max divergence defined above is the worst-case analog of the KL divergence and it implies that the log-ratio of the probabilities is bounded by ϵ, which directly relates to the max divergence:

$$
D_{\infty}(\mathcal{M}(D) \parallel \mathcal{M}(D')) \leq \epsilon.
\tag{12}
$$

Thus, the max divergence provides a useful and intuitive way to understand the worst-case privacy guarantees offered by a differential privacy mechanism [10].

Definition 5 (Rényi divergence). *Let P and Q be two probability measures over a measurable space $(\mathcal{X}, \mathcal{F})$ with property $P \ll Q$. The Rényi divergence of order α between P and Q is defined as:*

$$
D_{\alpha}(P \parallel Q) = \frac{1}{\alpha - 1} \log\left[\int_{\mathcal{X}} \left(\frac{dP}{dQ}\right)^{\alpha} dQ\right],
\tag{13}
$$

for $\alpha > 1$.

Lemma 1.

$$
\lim_{\alpha \to 1} D_{\alpha}(P \parallel Q) = D_{KL}(P \parallel Q).
\tag{14}
$$

$$
\lim_{\alpha \to \infty} D_{\alpha}(P \parallel Q) = D_{\infty}(P \parallel Q).
\tag{15}
$$

Proof. To prove

$$
\lim_{\alpha \to 1} D_{\alpha}(P \parallel Q) = D_{\mathrm{KL}}(P \parallel Q),
\tag{16}
$$

define

$$
S(\alpha) = \int_{\mathcal{X}} P(x)^{\alpha} Q(x)^{1-\alpha} \, dx,
\tag{17}
$$

and let $f(\alpha) = \log S(\alpha)$. Then,

$$
D_{\alpha}(P \parallel Q) = \frac{f(\alpha)}{\alpha - 1}.
\tag{18}
$$

Using differentiation under the integral sign:

$$
S'(\alpha) = \int_{\mathcal{X}} P(x)^{\alpha} Q(x)^{1-\alpha} (\log P(x) - \log Q(x)) \, dx.
\tag{19}
$$

At $\alpha = 1$, we have:

$$
S(1) = 1, \quad S'(1) = D_{\mathrm{KL}}(P \parallel Q).
\tag{20}
$$

Thus,

$$
f'(1) = \frac{S'(1)}{S(1)} = D_{\mathrm{KL}}(P \parallel Q).
\tag{21}
$$

Applying L'Hôpital's rule:

$$\lim_{\alpha \to 1} D_\alpha(P \parallel Q) = \lim_{\alpha \to 1} \frac{f(\alpha)}{\alpha - 1} = f'(1) = D_{\mathrm{KL}}(P \parallel Q). \tag{22}$$

Now, to prove

$$\lim_{\alpha \to \infty} D_\alpha(P \parallel Q) = D_\infty(P \parallel Q), \tag{23}$$

define

$$I(\alpha) = \int_{\mathcal{X}} \left(\frac{P(x)}{Q(x)} \right)^\alpha Q(x) \, dx, \tag{24}$$

such that

$$D_\alpha(P \parallel Q) = \frac{1}{\alpha - 1} \log I(\alpha). \tag{25}$$

Let $M = \operatorname{ess\,sup}_{x \in \mathcal{X}} \frac{P(x)}{Q(x)}$. For any $\epsilon > 0$, define

$$E_\epsilon = \left\{ x \in \mathcal{X} \;\middle|\; \frac{P(x)}{Q(x)} > M - \epsilon \right\}, \tag{26}$$

where $Q(E_\epsilon) > 0$ since M is the essential supremum. We obtain the following bounds for $I(\alpha)$:
Lower bound:

$$I(\alpha) \geq (M - \epsilon)^\alpha Q(E_\epsilon), \tag{27}$$

Upper bound:

$$I(\alpha) \leq M^\alpha. \tag{28}$$

Taking logarithms and dividing by α:

$$\log(M - \epsilon) + \frac{1}{\alpha} \log Q(E_\epsilon) \leq \frac{\log I(\alpha)}{\alpha} \leq \log M. \tag{29}$$

As $\alpha \to \infty$, $\frac{1}{\alpha} \log Q(E_\epsilon) \to 0$, so:

$$\lim_{\alpha \to \infty} \frac{\log I(\alpha)}{\alpha} = \log M. \tag{30}$$

Note that

$$\lim_{\alpha \to \infty} \frac{\alpha}{\alpha - 1} = 1. \tag{31}$$

Therefore,

$$\begin{aligned} \lim_{\alpha \to \infty} D_\alpha(P \parallel Q) &= \lim_{\alpha \to \infty} \frac{1}{\alpha - 1} \log I(\alpha) \\ &= \log M \\ &= D_\infty(P \parallel Q). \end{aligned} \tag{32}$$

$\square$

3   Related Work

Given the properties outlined in Lemma 1, Rényi divergence is a natural choice for analyzing and developing new frameworks for differential privacy.

3.1   Rényi differential privacy

A direct application of Rényi divergence is Rényi Differential Privacy (RDP) [12]. RDP provides a more flexible and fine-grained privacy analysis compared to traditional differential privacy. A randomized mechanism $\mathcal{M} : D \to \mathbb{R}$ satisfies $(\alpha, \bar{\epsilon})$-Rényi differential privacy if, for all adjacent datasets D and D',

$$D_\alpha(\mathcal{M}(D) \parallel \mathcal{M}(D')) \leq \bar{\epsilon}. \tag{33}$$

By adjusting α, RDP allows precise control over the privacy-utility trade-off. Different α values provide varying sensitivity to outliers, enabling tailored privacy guarantees. RDP's strong composability properties simplify the analysis of cumulative privacy consumption.

A common instantiation of the Rényi mechanism involves adding Gaussian noise. The parameter σ_G^2 is chosen to satisfy (α, ϵ)-RDP. Specifically, σ_G^2 is calibrated as:

$$\sigma_G^2 = \frac{\alpha \Delta f_2^2}{2\bar{\epsilon}}. \tag{34}$$

Additionally, RDP can be converted to (ϵ, δ)-differential privacy, allowing for flexible privacy budget management. Specifically, given a mechanism that satisfies $(\alpha, \bar{\epsilon})$-RDP. The parameter ϵ of (ϵ, δ)-differential privacy can be chosen as:

$$\epsilon = \bar{\epsilon} + \frac{\log(1/\delta)}{\alpha - 1}. \tag{35}$$

3.2   Zero-Concentrated Differential Privacy

Zero-Concentrated Differential Privacy (zCDP) is another refinement leveraging Rényi divergence [13]. A randomized mechanism $\mathcal{M} : D \to \mathbb{R}$ satisfies ρ-zCDP if for all adjacent datasets D and D' and for all $\alpha \in (1, \infty)$:

$$D_\alpha(\mathcal{M}(D) \parallel \mathcal{M}(D')) \le \rho\alpha. \tag{36}$$

zCDP simplifies privacy analysis compared to $(\bar{\epsilon}, \delta)$-differential privacy. The Gaussian mechanism is a natural fit for zCDP, where for a function f with ℓ_2 sensitivity Δf_2, adding Gaussian noise with variance σ_G^2 satisfies ρ-zCDP with:

$$\sigma_G^2 = \frac{\Delta f_2^2}{2\rho}. \tag{37}$$

zCDP also can be converted to (ϵ, δ)-differential privacy. Given a mechanism that satisfies $(\alpha, \bar{\epsilon})$-zCDP. The parameter ϵ (ϵ, δ)-differential privacy can be chosen as:

$$\epsilon = \rho + 2\sqrt{\rho \log(1/\delta)}. \tag{38}$$

4   Alpha differential Privacy

Rényi divergence and Zero-Concentrated Differential Privacy offer robust and flexible frameworks for analyzing privacy consumption in differential privacy mechanisms. Their ability to balance privacy and utility, combined with strong composability properties, makes them essential tools in the design of privacy-preserving data analysis algorithms.

Inspired by Rényi divergence, this paper aims to explore the connection and practical significance of alpha divergence, a notable subset of f-divergence closely related to Rényi divergence, within the context of differential privacy [15].

Definition 6 (f-divergence [16]). *Given a convex function $f : (0, \infty) \to \mathbb{R}$ with $f(1) = 0$, if $P \ll Q$, the f-divergence between two probability measures P and Q over a measure space $(\mathcal{X}, \mathcal{F})$ is defined as:*

$$D_f(P \parallel Q) \triangleq \int_{\mathcal{X}} f\left(\frac{dP}{dQ}\right) dQ. \tag{39}$$

In the context of probability density functions, let λ be the Lebesgue measure, if $Q \ll \lambda$, the f-divergence $D_f(P \parallel Q)$ is just defined as:

$$D_f(P \parallel Q) \triangleq \int_{\mathcal{X}} f\left[\frac{\left(\frac{dP}{d\lambda}\right)}{\left(\frac{dQ}{d\lambda}\right)}\right] \frac{dQ}{d\lambda} d\lambda = \int_{\mathcal{X}} f\left(\frac{p}{q}\right) q \, d\lambda. \tag{40}$$

Where p and q are corresponding density functions of P and Q with respect to the Lebesgue measure (A more general version can be found in [15]).

Definition 7 (Alpha divergence). *Let P and Q be two probability measures over a measure space $(\mathcal{X}, \mathcal{F})$ and λ be the Lebesgue measure with the property of $P \ll Q \ll \lambda$. Let p and q be the density functions of P and Q with respect to the Lebesgue measure. The alpha divergence is a special case of f-divergence, generated by the f-function defined on $\mathbb{R} \setminus \{0, 1\}$ [14]:*

$$f(u) = \frac{u^\alpha - \alpha u - (1 - \alpha)}{\alpha(\alpha - 1)}, \tag{41}$$

where $u = \frac{dP}{dQ} = \frac{p}{q}$, reader can easily check the convexity of $f(u)$.
The alpha divergence can be expressed as:

$$
\begin{aligned}
\widetilde{D}_\alpha(P \parallel Q) &\triangleq \frac{1}{\alpha(\alpha - 1)} \left[\int_{\mathcal{X}} p^\alpha q^{1-\alpha} - \alpha p - (1 - \alpha) q \, d\lambda \right] \\
&= \frac{1}{\alpha(\alpha - 1)} \left[\int_{\mathcal{X}} p^\alpha q^{1-\alpha} \, d\lambda - (1 - \alpha) - \alpha \right] \\
&= \frac{1}{\alpha(\alpha - 1)} \left[\int_{\mathcal{X}} p^\alpha q^{1-\alpha} \, d\lambda - 1 \right].
\end{aligned}
\tag{42}
$$

Lemma 2. *An f-divergence is always non-negative.*

Proof. Let P and Q be two probability measures over a measurable space $(\mathcal{X}, \mathcal{F})$. The f-divergence between P and Q is given by:

$$
\begin{aligned}
D_f(P \parallel Q) &= \int_{\mathcal{X}} f\left(\frac{dP}{dQ}\right) dQ \\
&= \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right) \right].
\end{aligned}
\tag{43}
$$

Since f is a convex function, by Jensen's inequality, for any random variable X,

$$f(\mathbb{E}[X]) \le \mathbb{E}[f(X)]. \tag{44}$$

Applying this to $X = \frac{dP}{dQ}$, we get:

$$f\left(\mathbb{E}_Q\left[\frac{dP}{dQ} \right] \right) \le \mathbb{E}_Q\left[f\left(\frac{dP}{dQ} \right) \right]. \tag{45}$$

Given that $\mathbb{E}_Q\left[\frac{dP}{dQ} \right] = 1$, it follows:

$$f(1) = 0 \le \mathbb{E}_Q\left[f\left(\frac{dP}{dQ} \right) \right]. \tag{46}$$

$\square$

Based on the non-negativity property of f-divergence (Lemma 2), we can define alpha differential privacy in a well-defined manner.

Definition 8 ((α, ϵ)-ADP). *A randomized mechanism $\mathcal{M} : D \to \mathbb{R}$ satisfies (α, ϵ)-alpha differential privacy with $\alpha > 1$ if, for all adjacent datasets D and D',*

$$\widetilde{D}_\alpha(\mathcal{M}(D) \parallel \mathcal{M}(D')) \le \epsilon. \tag{47}$$

Remark 1. *Although α in alpha divergence can take any value except 0 and 1, ADP typically considers the case where $\alpha > 1$. This restriction is adopted because, in this range, ADP exhibits properties that are particularly advantageous for practical applications. These beneficial properties will be shown in detail below.*

4.1   Preservation of alpha differential privacy under Post-Processing

Differential privacy methods possess a crucial characteristic which is their capacity to tolerate post-processing, as stated in the works of [2]. This indicates that if a mechanism $\mathcal{M}$ satisfies the criteria for α-differential privacy, then any additional operations performed on the output of $\mathcal{M}$ will not compromise the privacy safeguards that it provides. Due to the fact that it ensures that any further analysis or modification of data that has been anonymized by a differential privacy mechanism will not jeopardize the data's privacy, this trait is particularly significant in applications that are used in the real world.

Differential privacy is strong and highly relevant in a wide variety of data analytic workflows as a result of this resilience. One example of this would be in the field of machine learning, where a model could be trained on differential privacy data in order to prevent it from memorizing sensitive information. Following the completion of the training process, the model may undergo a number of evaluations and transformations, including parameter tuning and model compression, among others. The post-processing property ensures that the privacy protection that was initially provided to the training data will not be diminished as a result of these actions.

In the following, we present a substantial verification of this property, which demonstrates that ADP maintains its guarantees even after post-processing has been performed.

Proposition 1 (Data Processing Inequality). *ADP is preserved under post-processing.*

Proof. The proof follows the approach outlined in Erven's work [17]. Let P and Q be two probability measures over a measurable space $(\mathcal{X}, \mathcal{F})$, with $P \ll Q$. Let $\mathcal{G}$ be the sub-σ-algebra of $\mathcal{F}$ generated by a measurable map f. We need to show that $P_\mathcal{G}$ and $Q_\mathcal{G}$, the restrictions of P and Q to $\mathcal{G}$, satisfy

$$\widetilde{D}_\alpha(P_\mathcal{G} \parallel Q_\mathcal{G}) \leq \widetilde{D}_\alpha(P \parallel Q). \tag{48}$$

To prove this, we need to show that:

$$\int_\mathcal{X} \left(\frac{dP_\mathcal{G}}{dQ_\mathcal{G}} \right)^\alpha dQ_\mathcal{G} \leq \int_\mathcal{X} \left(\frac{dP}{dQ} \right)^\alpha dQ. \tag{49}$$

Recall that the conditional expectation $\mathbb{E}_Q \left[\frac{dP}{dQ} \middle| \mathcal{G} \right]$ is the Radon-Nikodym derivative $\frac{dP_\mathcal{G}}{dQ_\mathcal{G}}$.

$$\frac{dP_\mathcal{G}}{dQ_\mathcal{G}} = \mathbb{E}_Q \left[\frac{dP}{dQ} \middle| \mathcal{G} \right]. \tag{50}$$

By Jensen's inequality for the convex function $x \mapsto x^\alpha$ (since by the definition of ADP, $\alpha > 1$), we have:

$$\begin{aligned}
\int_\mathcal{X} \left(\frac{dP_\mathcal{G}}{dQ_\mathcal{G}} \right)^\alpha dQ_\mathcal{G} &= \int_\mathcal{X} \left(\frac{dP_\mathcal{G}}{dQ_\mathcal{G}} \right)^\alpha dQ \\
&= \int_\mathcal{X} \left(\mathbb{E} \left[\frac{dP}{dQ} \middle| \mathcal{G} \right] \right)^\alpha dQ \\
&\leq \int_\mathcal{X} \mathbb{E} \left[\left(\frac{dP}{dQ} \right)^\alpha \middle| \mathcal{G} \right] dQ \\
&= \int_\mathcal{X} \left(\frac{dP}{dQ} \right)^\alpha dQ.
\end{aligned} \tag{51}$$

The first line holds since $dQ_\mathcal{G}$ is the restriction of dQ to the sub-σ-algebra $\mathcal{G}$. Hence, we have shown that:

$$\widetilde{D}_\alpha(P_\mathcal{G} \parallel Q_\mathcal{G}) \leq \widetilde{D}_\alpha(P \parallel Q). \tag{52}$$

This proves that ADP is preserved under post-processing. $\square$

4.2   Adaptive composability of alpha differential privacy

The ability of differential privacy methods to be composed is yet another important characteristic of these techniques. This composability guarantees that the cumulative privacy consumption that arises from numerous applications of differential privacy techniques may be controlled and regulated in a systematic manner, as stated in the works of [1, 11]. In order to demonstrate the efficacy and adaptability of ADP in comparison to general differential privacy methods, we will demonstrate that ADP possesses composability qualities that are comparable to those with general mechanisms.

Furthermore, this promise extends to situations in which the next mechanism is picked in an adaptive manner based on the output of the mechanism that came before it.

In machine learning pipelines, where models are frequently trained in an iterative manner, composability and adaptability are especially useful. It is possible that differential privacy safeguards will be implemented throughout each iteration. This will ensure that the model does not overfit the training data and so accidentally disclose sensitive information. ADP ensures that privacy guarantees are valid during the training process by retaining composability. This provides strong protection against data leakage and ensures that the training session is successful.

Proposition 2 (Adaptive Sequential Composition). *Let mechanisms $\mathcal{M}_1 : D \to \mathcal{A}$ and $\mathcal{M}_2 : \mathcal{A} \times D \to \mathcal{B}$ be (α, ϵ_1)-ADP and (α, ϵ_2)-ADP mechanisms, respectively. Then the mechanism defined as (X, Y), where $X \sim \mathcal{M}_1(D)$ and $Y \sim \mathcal{M}_2(X, D)$, satisfies $(\alpha, \epsilon_1 + \epsilon_2 + \alpha(\alpha - 1)\epsilon_1\epsilon_2)$-ADP.*

Proof. Let $\mathcal{M}_3 : D \to \mathcal{A} \times \mathcal{B}$ be the mechanism obtained by sequentially applying $\mathcal{M}_1$ and $\mathcal{M}_2$. Denote the probability measures induced by $\mathcal{M}_1$ on $\mathcal{A}$, $\mathcal{M}_2$ given X on $\mathcal{B}$, and the joint probability measure on $\mathcal{A} \times \mathcal{B}$ as P_X, $P_{Y|X}$, and $P_{X,Y}$, respectively. Similarly, let $P_{X'}$, $P_{Y'|X'}$, and $P_{X',Y'}$ represent the corresponding probability measures when the input dataset is D'.

Before proceeding with the calculations, note that $P_{X,Y} \ll P_{X',Y'}$ holds because $P_X \ll P_{X'}$ holds (from the ADP property of $\mathcal{M}_1$) and $P_{Y|X} \ll P_{Y'|X'}$ holds (from the ADP property of $\mathcal{M}_2$). The product measure $P_{X,Y} = P_X \times P_{Y|X}$ is therefore absolutely continuous with respect to $P_{X',Y'} = P_{X'} \times P_{Y'|X'}$ given that all of the measures here is σ-finite. Then:

$$
\begin{aligned}
&\alpha(\alpha - 1)\widetilde{D}_\alpha(\mathcal{M}_3(D) \| \mathcal{M}_3(D')) + 1 \\
&= \int_{\mathcal{A} \times \mathcal{B}} \left(\frac{dP_{X,Y}}{dP_{X',Y'}} \right)^\alpha dP_{X',Y'} \\
&= \int_{\mathcal{A} \times \mathcal{B}} \left(\frac{d(P_X \times P_{Y|X})}{d(P_{X'} \times P_{Y'|X'})} \right)^\alpha d(P_{X'} \times P_{Y'|X'}) \\
&= \int_{\mathcal{A}} \left(\frac{dP_X}{dP_{X'}} \right)^\alpha dP_{X'} \int_{\mathcal{B}} \left(\frac{dP_{Y|X}}{dP_{Y'|X'}} \right)^\alpha dP_{Y'|X'} \\
&\leq (\alpha(\alpha - 1)\epsilon_1 + 1)(\alpha(\alpha - 1)\epsilon_2 + 1) \\
&= \alpha(\alpha - 1)(\epsilon_1 + \epsilon_2 + \alpha(\alpha - 1)\epsilon_1\epsilon_2) + 1,
\end{aligned}
\tag{53}
$$

which proves the claim. $\square$

It should be noted that ϵ in alpha-differential privacy is not entirely consistent with the intuitive understanding of the privacy parameter. Higher values do not necessarily correspond to weaker privacy guarantees. Since the mapping of parameters in alpha-differential privacy to those in traditional privacy frameworks is not linearly positively correlated. We will show this in section V.

4.3   Group privacy of alpha differential privacy

An extension of the individual privacy guarantees that are included in differential privacy frameworks is the concept of group privacy. Group privacy ensures that the privacy of any group of persons is also kept, in contrast to the standard differential privacy approach, which focuses on safeguarding the privacy of individual entries within a dataset [18]. In the context of ADP, the concept of group privacy addresses situations in which the adversary may possess auxiliary information about many persons contained within the dataset. This is of utmost significance in applications that deal with sensitive data, since it is necessary to safeguard the privacy of subgroups within the data against the possibility of inference attacks.

Suppose that there is a database in the healthcare industry that contains confidential patient information. In the case of traditional differential privacy, it is possible to guarantee that the inclusion or exclusion of the data of a single patient does not significantly impact the outcomes of an analysis. On the other hand, if an adversary is aware that a group of patients are members of the same family or community, then they could be able to piece together sensitive information about the group by accessing data that is relevant to that group. Group privacy helps limit this danger by extending privacy guarantees to groups. This ensures that even if someone has access to additional information, they are unable to readily jeopardize the privacy of the individuals who are a part of the group.

The ADP that we have described can be extended to safeguard the privacy of groups of varying sizes. This will ensure that the appropriate level of privacy is provided for a variety of practical contexts, including healthcare, finance, social sciences, and other areas.

Lemma 3 (Triangle inequality of alpha divergence). *Let P, Q, and R be three probability measures defined in a measurable space $(\mathcal{X}, \mathcal{F})$, where λ is the Lebesgue measure such that $\lambda \ll P \ll Q \ll R \ll \lambda$. Define $I_\alpha(A \parallel B) = \alpha(\alpha - 1)\widetilde{D}_\alpha(A \parallel B) + 1$, where $\widetilde{D}_\alpha(A \parallel B)$ denotes the alpha divergence between the probability measures A and B for $\alpha > 1$. Then, the following inequality holds:*

$$I_\alpha(P \parallel Q) \leq (I_{2\alpha}(P \parallel R))^{\frac{1}{2}} (I_{2\alpha-1}(R \parallel Q))^{\frac{1}{2}}.$$

Proof. The proof follows directly from an application of Hölder's inequality. We begin by expressing $I_\alpha(P \parallel Q)$ in terms of the Radon-Nikodym derivatives:

$$
\begin{aligned}
&I_\alpha(P \parallel Q) \\
&= \int_{\mathcal{X}} \left(\frac{dP}{dQ}\right)^\alpha dQ \\
&= \int_{\mathcal{X}} \left(\frac{dP}{d\lambda}\right)^\alpha \left(\frac{dR}{d\lambda}\right)^{\frac{1}{2}-\alpha} \left(\frac{dR}{d\lambda}\right)^{\alpha-\frac{1}{2}} \left(\frac{dQ}{d\lambda}\right)^{1-\alpha} d\lambda.
\end{aligned}
\tag{54}
$$

Applying Hölder's inequality with exponents $p = 2$ and $q = 2$, we obtain:

$$
\begin{aligned}
I_\alpha(P \parallel Q) &\leq \left(\int_{\mathcal{X}} \left(\frac{dP}{d\lambda}\right)^{2\alpha} \left(\frac{dR}{d\lambda}\right)^{1-2\alpha} d\lambda\right)^{\frac{1}{2}} \\
&\quad \times \left(\int_{\mathcal{X}} \left(\frac{dR}{d\lambda}\right)^{2\alpha-1} \left(\frac{dQ}{d\lambda}\right)^{2-2\alpha} d\lambda\right)^{\frac{1}{2}} \\
&= \left(\int_{\mathcal{X}} \left(\frac{dP}{dR}\right)^{2\alpha} dR\right)^{\frac{1}{2}} \left(\int_{\mathcal{X}} \left(\frac{dR}{dQ}\right)^{2\alpha-1} dQ\right)^{\frac{1}{2}} \\
&= (I_{2\alpha}(P \parallel R))^{\frac{1}{2}} (I_{2\alpha-1}(R \parallel Q))^{\frac{1}{2}}.
\end{aligned}
\tag{55}
$$

This completes the proof. $\qquad\square$

Proposition 3 (Group privacy in ADP). *Let $\mathcal{M}$ be a mechanism that satisfies (α, ϵ)-ADP with $\alpha > 2^k$. For any group of $2^k + 1$ sizes, let D and D' be two datasets that differ in at most 2^k entries. The mechanism $\mathcal{M}$ provides $\left(\frac{\alpha}{2^k}, \frac{\alpha(\alpha-1)}{\frac{\alpha}{2^k}\left(\frac{\alpha}{2^k}-1\right)}\epsilon\right)$-group privacy for any such pair of datasets D and D'.*

Proof. Let D_1, D_2, and D_3 be three datasets such that D_1 is adjacent to D_2 and D_2 is adjacent to D_3. Let P, R, and Q be the probability measures induced by $\mathcal{M}(D_1)$, $\mathcal{M}(D_2)$, and $\mathcal{M}(D_3)$ over the measurable space $(\mathcal{X}, \mathcal{F})$, respectively. Assume λ is the Lebesgue measure, and suppose the condition $\lambda \ll P \ll Q \ll R \ll \lambda$ is satisfied, as required by Lemma 3. Let $I_\alpha(A \parallel B) = \alpha(\alpha - 1)\widetilde{D}_\alpha(A \parallel B) + 1$, as defined in Lemma 3. Under these conditions, we can apply Lemma 3 to obtain:

$$I_\alpha(P \parallel Q) \leq (I_{2\alpha}(P \parallel R))^{\frac{1}{2}} (I_{2\alpha-1}(R \parallel Q))^{\frac{1}{2}}.
\tag{56}$$

Let's consider the relationship between $I_{2\alpha-1}(R \parallel Q)$ and $I_{2\alpha}(R \parallel Q)$. We start with the following expression:

$$
\begin{aligned}
I_{2\alpha-1}(R \parallel Q) &= \mathbb{E}_Q\left[\left(\frac{dR}{dQ}\right)^{2\alpha-1}\right] \\
&= \mathbb{E}_Q\left[\left(\frac{dR}{dQ}\right)^{2\alpha \cdot \frac{2\alpha-1}{2\alpha}}\right].
\end{aligned}
\tag{57}
$$

By Jensen's inequality for the concave function $f(x) = x^{\frac{2\alpha-1}{2\alpha}}$, we have:

$$\mathbb{E}_Q\left[\left(\frac{dR}{dQ}\right)^{2\alpha \cdot \frac{2\alpha-1}{2\alpha}}\right] \leq I_{2\alpha}(R \parallel Q)^{\frac{2\alpha-1}{2\alpha}}.
\tag{58}$$

Then,

$$
\begin{aligned}
I_\alpha(P \parallel Q) &\leq I_{2\alpha}(P \parallel R)^{\frac{1}{2}} I_{2\alpha}(R \parallel Q)^{\frac{2\alpha-1}{4\alpha}} \\
&\leq I_{2\alpha}(P \parallel R),
\end{aligned}
\tag{59}
$$

which implies:

$$
\begin{aligned}
\widetilde{D}_{\frac{\alpha}{2}}(P \parallel Q) &\leq \frac{\alpha(\alpha-1)}{\frac{\alpha}{2}\left(\frac{\alpha}{2}-1\right)} \widetilde{D}_\alpha(P \parallel R) \\
&= \frac{\alpha(\alpha-1)}{\frac{\alpha}{2}\left(\frac{\alpha}{2}-1\right)} \epsilon.
\end{aligned}
\tag{60}
$$

Now, let D_1 and D_3 be two datasets differing in at most 2^k entries, where P and Q are the probability measures induced by $\mathcal{M}(D_1)$ and $\mathcal{M}(D_3)$, respectively. We maintain similar settings as before. By induction, we have:

$$
\begin{aligned}
&\widetilde{D}_{\frac{\alpha}{2^k}}(P \parallel Q) \\
&\leq \frac{\frac{\alpha}{2^{k-1}}\left(\frac{\alpha}{2^{k-1}}-1\right)}{\frac{\alpha}{2^k}\left(\frac{\alpha}{2^k}-1\right)} \cdot \frac{\frac{\alpha}{2^{k-2}}\left(\frac{\alpha}{2^{k-2}}-1\right)}{\frac{\alpha}{2^{k-1}}\left(\frac{\alpha}{2^{k-1}}-1\right)} \cdots \frac{\alpha(\alpha-1)}{\frac{\alpha}{2}\left(\frac{\alpha}{2}-1\right)} \epsilon \\
&= \prod_{i=0}^{k-1} \frac{\frac{\alpha}{2^i}\left(\frac{\alpha}{2^i}-1\right)}{\frac{\alpha}{2^{i+1}}\left(\frac{\alpha}{2^{i+1}}-1\right)} \epsilon \\
&= \frac{\alpha(\alpha-1)}{\frac{\alpha}{2^k}\left(\frac{\alpha}{2^k}-1\right)} \epsilon,
\end{aligned}
\tag{61}
$$

which proves the claim. $\qquad\square$

Remark 2. *It is worth noting that maintaining a fixed α while extending ADP to group privacy appears to be infeasible under the constraints of not assuming any specific distribution. Therefore, we must adjust α by dividing it by 2^k to get the bound of group privacy.*

Remark 3. *Regarding the absolute continuity chain $\lambda \ll P \ll Q \ll R \ll \lambda$, it is good to know that many commonly used differential privacy mechanisms, such as the Laplace and Gaussian mechanisms, satisfy this requirement. These mechanisms induce probability measures with well-defined Radon-Nikodym derivatives with respect to the Lebesgue measure λ (e.g., the Laplace and Gaussian densities), thereby ensuring absolute continuity. Additionally, these measures are mutually absolutely continuous because the mechanisms generate overlapping supports and assign nonzero probability density to the same regions of the space.*

5   ADP and (ϵ, δ)-DP

The relationship between alpha differential privacy (ADP) and approximate differential privacy ((ϵ, δ)-DP) is an important aspect of privacy analysis. By adjusting the parameters α and ϵ, ADP offers a flexible approach to privacy guarantees. This flexibility allows us to map the guarantees of ADP to the (ϵ, δ) framework, thereby connecting these two important privacy models.

Proposition 4 (Relationship between ADP and (ϵ, δ)-DP). *If $\mathcal{M}$ is an (α, ϵ)-ADP mechanism, then it also satisfies $(\bar{\epsilon}, \delta)$-DP with $\delta \in (0, 1)$, where $\bar{\epsilon} \geq \frac{\log\left(\frac{e^\epsilon \alpha(\alpha-1)+1}{\delta}\right)}{\alpha-1}$.*

Proof. Let P and Q be the probability measures induced by $\mathcal{M}(D)$ and $\mathcal{M}(D')$, respectively, over a measurable space $(\mathcal{X}, \mathcal{F})$. We have:

$$
\begin{aligned}
\int_{\mathcal{X}} \left(\frac{dP}{dQ}\right)^\alpha dQ &= \mathbb{E}_Q\left[\left(\frac{dP}{dQ}\right)^\alpha\right] \\
&= \mathbb{E}_P\left[\left(\frac{dP}{dQ}\right)^{\alpha-1}\right] \\
&\leq \alpha(\alpha-1)\epsilon + 1.
\end{aligned}
\tag{62}
$$

By Markov's inequality, we restrain:

$$
\begin{aligned}
\Pr\left[\frac{dP}{dQ} > e^{\bar{\epsilon}}\right] &= \Pr\left[\left(\frac{dP}{dQ}\right)^{\alpha-1} \geq e^{\bar{\epsilon}(\alpha-1)}\right] \\
&\leq \frac{\mathbb{E}\left[\left(\frac{dP}{dQ}\right)^{\alpha-1}\right]}{e^{\bar{\epsilon}(\alpha-1)}} \\
&\leq \frac{\alpha(\alpha-1)\epsilon + 1}{e^{\bar{\epsilon}(\alpha-1)}} \\
&\leq \delta,
\end{aligned}
\tag{63}
$$

which implies

$$
\bar{\epsilon} \geq \frac{\log\left(\frac{e^{\epsilon}\alpha(\alpha-1)+1}{\delta}\right)}{\alpha - 1}.
\tag{64}
$$

This completes the proof. □

From the equation above, it is evident that in comparison to the parameter ϵ, the approximate privacy guarantee corresponding to ADP is more constrained by the value of α. Specifically, a larger α value tends to result in a more stringent approximate differential privacy guarantee.

In Section VII, we will conduct a detailed comparison of the privacy accumulation for various differential privacy.

6   Various Mechanisms of ADP

In this section, we explore three widely used mechanisms—Randomized response, Laplace, and Gaussian—and demonstrate how they can be adapted to the framework of alpha differential privacy (ADP). Each mechanism offers unique advantages and can be leveraged effectively depending on the specific requirements of a given privacy-preserving application.

6.1   Randomized response mechanism

The randomized response mechanism is frequently employed in privacy-preserving surveys and questionnaires. It incorporates randomization into responses to guarantee plausible deniability, complicating the identification of an individual's authentic response.

The mechanism $\mathcal{M}_R(f)$ for a predicate $f : \mathcal{D} \to \{0, 1\}$ is defined as follows:

$$
\mathcal{M}_R(f(D)) = \begin{cases} f(D) & \text{with probability } p \\ 1 - f(D) & \text{with probability } 1 - p \end{cases}
\tag{65}
$$

Here, probability p controls the amount of noise introduced into the mechanism.

Proposition 5 (Randomized response mechanism and ADP). *If $\mathcal{M}_R$ is a randomized response mechanism, it satisfies* $\left(\alpha, \frac{1}{\alpha(\alpha-1)}\left(p^{\alpha}(1-p)^{1-\alpha} + (1-p)^{\alpha}p^{1-\alpha} - 1\right)\right)$-*ADP.*

Proof. Without loss of generality, we assume that $f(D) = 1$ and the worst-case response generated by D' is $f(D') = 0$. Using the definition of the randomized response mechanism, the probability distributions for D and D' are:

$$
\begin{aligned}
Pr(\mathcal{M}_R(f(D)) = 1) &= p, \ Pr(\mathcal{M}_R(f(D)) = 0) = 1 - p, \\
Pr(\mathcal{M}_R(f(D')) = 1) &= 1 - p, \ Pr(\mathcal{M}_R(f(D')) = 0) = p.
\end{aligned}
\tag{66}
$$

Therefore,

$$
\begin{aligned}
&\widetilde{D}_{\alpha}(\mathcal{M}_R(f(D))\|\mathcal{M}_R(f(D'))) \\
&= \frac{1}{\alpha(\alpha-1)}\left(\sum_{\{0,1\}} p^{\alpha}q^{1-\alpha} - 1\right) \\
&= \frac{1}{\alpha(\alpha-1)}\left(p^{\alpha}(1-p)^{1-\alpha} + (1-p)^{\alpha}p^{1-\alpha} - 1\right).
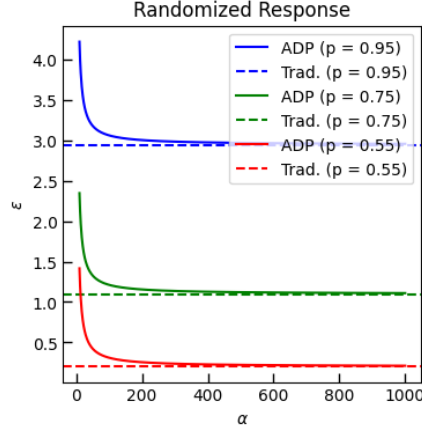\end{aligned}
\tag{67}
$$

□

Figure 1: privacy consumption of the randomized response mechanism under different probabilities p, with the horizontal axis representing the value of α and the vertical axis indicating the privacy consumption (ϵ). The solid line represents the privacy consumption evaluated by ADP for a failure probability of $\delta = 1e\text{-}5$ across varying values of α, while the dashed line shows the privacy consumption as assessed by the traditional ϵ-differential privacy framework.

The randomized response mechanism stands out for its simplicity and applicability to categorical data, particularly in scenarios where the data consists of binary or discrete attributes. By flipping the result of a predicate with a certain probability, Randomized Response offers a straightforward yet effective way to ensure privacy while maintaining the utility of individual query results. As illustrated in Figure 1, when ADP is used to evaluate privacy consumption of the randomized response mechanism, increasing the value of α causes the evaluation results to converge with those of the traditional privacy framework, which means this mechanism can be tightly integrated into the ADP framework, achieving a privacy-utility trade-off that through the choice of α.

6.2   Laplace mechanism

Recall that the definition of Laplace mechanism defined in II is:

$$\mathcal{M}_L(D) \triangleq f(D) + \text{Lap}(0, b), \tag{68}$$

with ℓ_1 sensitivity.

Proposition 6 (Laplace mechanism and ADP). *If $\mathcal{M}_L$ is a Laplace mechanism, with sensitivity Δf_1 and scale b, it satisfies $\left(\alpha, \frac{\exp\left(\frac{(\alpha-1)\mu}{b}\right)}{(\alpha-1)(2\alpha-1)} + \frac{\exp\left(-\frac{\alpha\mu}{b}\right)}{\alpha(2\alpha-1)} - \frac{1}{\alpha(\alpha-1)} \right)$-ADP.*

Proof. Without loss of generality, assume the distribution of $\mathcal{M}_L(D)$ is $\text{Lap}(0, b)$, the distribution $\mathcal{M}_L(D')$ generated by D's adjacent dataset D' is $\text{Lap}(\mu, b)$. Notice that the Laplace distribution is symmetrical. Thus, we can assume

$\mu > 0$, we have:

$$\widetilde{D}_\alpha(\mathcal{M}_L(D)\|\mathcal{M}_L(D'))$$

$$=\frac{1}{\alpha(\alpha-1)}\frac{1}{2b}\left(\int_{-\infty}^{0}\exp\left(\frac{x-\mu}{b}\right)dx\right.$$

$$+\int_{0}^{\mu}\exp\left(\frac{\alpha x-\mu}{b}\right)dx$$

$$\left.+\int_{\mu}^{+\infty}\exp\left(-\frac{\alpha x+\mu}{b}\right)dx\right)-1$$

$$=\frac{1}{\alpha(\alpha-1)}\frac{1}{2b}\left(b\exp\left(\frac{(\alpha-1)\mu}{b}\right)\right.$$

$$+\frac{b}{2\alpha-1}\left(\exp\left(\frac{(\alpha-1)\mu}{b}\right)-\exp\left(-\frac{\alpha\mu}{b}\right)\right)$$

$$\left.+b\exp\left(-\frac{\alpha\mu}{b}\right)\right)-1$$

$$=\frac{\exp\left(\frac{(\alpha-1)\mu}{b}\right)}{(\alpha-1)(2\alpha-1)}+\frac{\exp\left(-\frac{\alpha\mu}{b}\right)}{\alpha(2\alpha-1)}-\frac{1}{\alpha(\alpha-1)}. \tag{69}$$

For the multivariate Laplace mechanism, assume $\mu \in \mathbb{R}^d$, it is immediate that:

$$\widetilde{D}_\alpha(\mathcal{M}_L(D)\|\mathcal{M}_L(D'))$$

$$=\frac{\exp\left(\frac{(\alpha-1)\|\mu\|_1}{b}\right)}{(\alpha-1)(2\alpha-1)}+\frac{\exp\left(-\frac{\alpha\|\mu\|_1}{b}\right)}{\alpha(2\alpha-1)}-\frac{1}{\alpha(\alpha-1)}. \tag{70}$$

We know that Δf_1 is an ℓ_1 sensitivity. Therefore, we have:

$$\Delta f_1 = \|\mu - 0\|_1 = \|\mu\|_1. \tag{71}$$
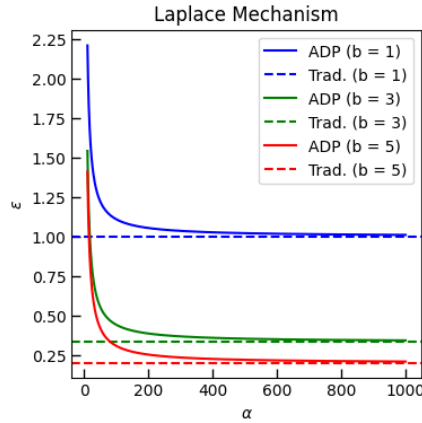
This proves the claim. □

Figure 2: privacy consumption of the Laplace mechanism under different scales b and a fixed sensitivity $\Delta f_1 = 1$, with the horizontal axis representing the value of α and the vertical axis indicating the privacy consumption (ϵ). The solid line represents the privacy consumption evaluated by ADP for a failure probability of $\delta = 1e\text{-}5$ across varying values of α, while the dashed line shows the privacy consumption as assessed by the traditional ϵ-differential privacy framework.

For the Laplace mechanism, the trend in privacy consumption (Figure 2) closely mirrors that of the randomized response mechanism. As the value of alpha increases, the privacy consumption gradually converges to the results predicted by the traditional differential privacy framework.

6.3   Gaussian Mechanism

Recall that the definition of Gaussian mechanism defined in II is:

$$\mathcal{M}_G(D) \triangleq f(D) + \mathcal{N}(0, \sigma_G^2), \tag{72}$$

with ℓ_2 sensitivity.

Proposition 7 (Gaussian mechanism and ADP). *If $\mathcal{M}_G$ is a Gaussian mechanism, with sensitivity Δf_2 and variance σ_G^2, it satisfies $\left(\alpha, \frac{1}{\alpha(\alpha-1)} \left(\exp\left(\frac{(\alpha^2-\alpha)\Delta f_2^2}{2\sigma_G^2}\right) - 1\right)\right)$-ADP.*

Proof. Similar to the proof of the Laplace mechanism. Without loss of generality, assume the distribution of $\mathcal{M}_G(D)$ is $\mathcal{N}(0, \sigma_G^2)$, the distribution of $\mathcal{M}_G(D')$ generated by D's adjacent dataset D' is $\mathcal{N}(\mu, \sigma_G^2)$. Hence, we have:

$$\widetilde{D}_\alpha(\mathcal{M}_G(D)\|\mathcal{M}_G(D'))$$
$$= \frac{1}{\alpha(\alpha-1)} \int_{-\infty}^{+\infty} \frac{\exp\left(\frac{-\alpha x^2 - (1-\alpha)(x-\mu)^2}{2\sigma_G^2}\right)}{\sigma_G\sqrt{2\pi}} - 1$$
$$= \frac{1}{\alpha(\alpha-1)} \left(\exp\left(\frac{(\alpha^2-\alpha)\mu^2}{2\sigma_G^2}\right) - 1\right). \tag{73}$$

For the multivariate Gaussian mechanism, assume $\mu \in \mathbb{R}^d$, then, the distribution of $\mathcal{M}_G(D)$ is $\mathcal{N}(0, \sigma_G^2 I_d)$, the distribution of $\mathcal{M}_G(D')$ is $\mathcal{N}(\mu, \sigma_G^2 I_d)$, it is immediate that:

$$\widetilde{D}_\alpha(\mathcal{M}_G(D)\|\mathcal{M}_G(D'))$$
$$= \frac{1}{\alpha(\alpha-1)} \left(\exp\left(\frac{(\alpha^2-\alpha)\|\mu\|_2^2}{2\sigma_G^2}\right) - 1\right). \tag{74}$$

We know that Δf_2 is an ℓ_2 sensitivity. Therefore, we have:

$$\Delta f_2^2 = \|\mu - 0\|_2^2 = \|\mu\|_2^2. \tag{75}$$

This proves the claim. □

Corollary 1. *A Gaussian mechanism with variance $\frac{\alpha(\alpha-1)\Delta f_2^2}{2\log(\alpha(\alpha-1)\epsilon+1)}$ satisfies (α, ϵ)-ADP*

Proof. The proof is immediate from Proposition 7. □

Under the ADP framework, the privacy consumption trend of the Gaussian mechanism displays a unique pattern in contrast to the preceding two mechanisms. Figure 3 demonstrates that the privacy consumption initially exhibits a convergence pattern as the amount of α increases, closely aligning with the privacy bounds anticipated by conventional differential privacy. This convergence is transient; once reaching a specific threshold, the privacy consumption diverges from the standard of traditional differential privacy. This trend is especially pronounced when the variance parameter σ_G is minimal (e.g., $\sigma_G = 1$), as the privacy consumption under ADP markedly surpasses that of the conventional framework with increasing α. This non-monotonic behaviour underscores the intricate relationship between α and privacy assurances in ADP, indicating that the selection of an optimal α necessitates meticulous evaluation, as an inappropriate choice may lead to greater cumulative privacy consumption than conventional privacy methods.

The strength of the ADP framework does not lie in evaluating privacy consumption for a single query. Instead, its real advantage lies in providing an effective upper bound estimation for the cumulative privacy consumption across multiple iterations, which will be analyzed in detail in Section VIII.

7   Guidance on Choosing α

In actual applications that require multiple iterations. A carefully chosen α ensures the privacy budget is utilized efficiently, minimizing the cumulative privacy loss. This section takes the Gaussian mechanism as an example to show how to select the optimal α.

Figure 4 illustrates the relationship between the cumulative privacy consumption of a Gaussian mechanism over 1000 iterations and α. The privacy consumption function exhibits a convex-like behaviour: for small values of α, privacy
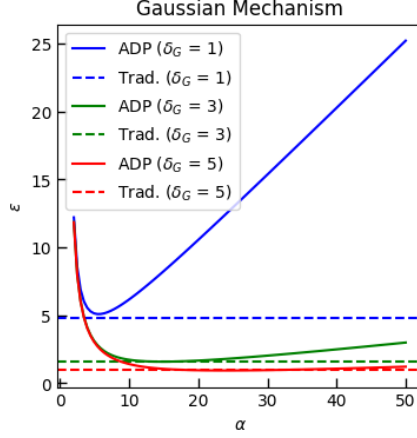
Figure 3: privacy consumption of the Gaussian mechanism under different standard deviation σ_G and a fixed sensitivity $\Delta f_2 = 1$, with the horizontal axis representing the value of α and the vertical axis indicating the privacy consumption (ϵ). The solid line represents the privacy consumption evaluated by ADP for a failure probability of $\delta = 1e\text{-}5$ across varying values of α, while the dashed line shows the privacy consumption as assessed by the traditional $(\epsilon, \delta)-$differential privacy framework.
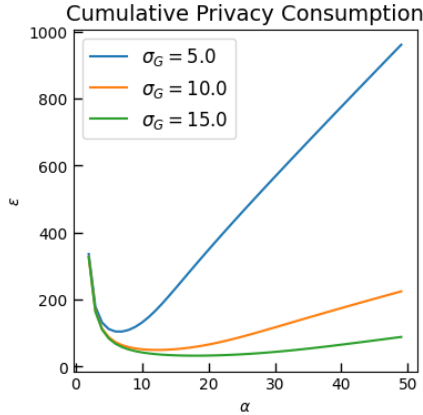

Figure 4: Cumulative privacy consumption (ϵ) of the Gaussian mechanism over 1000 iterations under varying required standard deviations, with a required failure rate of $\delta = 1e\text{-}5$ and a required sensitivity of $\Delta f_2 = 1$. The horizontal axis represents the value of α, while the vertical axis indicates the cumulative privacy consumption (ϵ).

consumption decreases rapidly as α increases, reaching a minimum within an optimal range. Beyond this range, further increases in α result in a gradual rise in privacy consumption. Notably, this trend closely resembles that of a single query using the Gaussian mechanism, as depicted in Figure 3.

Similar to α in Rényi Differential Privacy, α in ADP is dynamically determined based on user-defined constraints. Algorithm 1 demonstrates a scenario on how to determine the optimal α to minimize cumulative privacy consumption based on user-defined constraints, including the failure probability, total number of iterations, standard deviation, and sensitivity. In this algorithm, α is dynamically selected to minimize cumulative privacy consumption under specific requirements.

For scenarios where users have different requirements—such as minimizing the standard deviation of the Gaussian mechanism to enhance data utility—a similar algorithm can be employed (Algorithm 2). By leveraging the specified failure probability, total iterations, sensitivity, and an upper bound on the overall privacy consumption, the optimal α and its corresponding standard deviation can be effectively determined. As shown in Figure 5.

In conclusion, our empirical results show that, over multiple iterations, the relationship between α and other privacy parameters is non-monotonic, instead exhibiting a behaviour that resembles convexity. This means that excessively

---

Algorithm 1 Find Alpha that Minimizes Privacy Consumption for Gaussian Mechanism

---

Input: Number of iterations ℓ, Standard deviation σ_G, Failure probability δ, Sensitivity Δf_2
Output: Minimum privacy consumption ϵ_{min} and optimal alpha α^*
1: Initialize $\epsilon_{min} \leftarrow \infty$, $\alpha^* \leftarrow 2$
2: // E.g. we can take α from 2 to 100
3: **for** α in a suitable range **do**
4:    Compute the ADP privacy consumption for a single query.
5:    $\epsilon \leftarrow \frac{1}{\alpha(\alpha-1)} \left(\exp\left(\frac{(\alpha^2-\alpha)\Delta f_2^2}{2\sigma_G^2} \right) - 1 \right)$
6:    Compute the cumulative ADP privacy consumption for ℓ iterations.
7:    $\epsilon_{new} \leftarrow 0$
8:    **for** $i = 1$ to ℓ **do**
9:      Update $\epsilon_{new} \leftarrow \epsilon_{new} + \epsilon + \alpha(\alpha - 1) \cdot \epsilon \cdot \epsilon_{new}$
10:    **end for**
11:    Convert ADP privacy consumption to traditional privacy consumption.
12:    $\epsilon_{temp} \leftarrow \frac{\log(\epsilon_{new} \cdot \alpha(\alpha-1)+1)}{\delta(\alpha-1)}$
13:    **if** $\epsilon_{temp} < \epsilon_{min}$ **then**
14:      Update $\epsilon_{min} \leftarrow \epsilon_{temp}$, $\alpha^* \leftarrow \alpha$
15:    **end if**
16: **end for**
17: **Return** ϵ_{min}, α^*

Algorithm 2 Find Alpha that Minimizes Standard Deviation for Gaussian Mechanism

---

Input: Number of iterations ℓ, Privacy consumption bound ϵ_{bound}, Failure probability δ, Sensitivity Δf_2
Output: Minimum standard deviation σ_{min} and optimal alpha α^*
1: Initialize $\alpha^* \leftarrow 2$, $\sigma_{min} \leftarrow \infty$
2: // E.g. we can take α from 2 to 100
3: **for** α in a suitable range **do**
4:    // E.g. we can take σ_G from 1 to 500
5:    **for** σ_G in a suitable range **do**
6:      **if** $\sigma_G \geq \sigma_{min}$ **then**
7:         **break**
8:      **end if**
9:    Compute the ADP privacy consumption for a single query.
10:    $\epsilon \leftarrow \frac{1}{\alpha(\alpha-1)} \left(\exp\left(\frac{(\alpha^2-\alpha)\Delta f_2^2}{2\sigma_G^2} \right) - 1 \right)$
11:    Compute the cumulative ADP privacy consumption for ℓ iterations.
12:    $\epsilon_{new} \leftarrow 0$
13:    **for** $i = 1$ to ℓ **do**
14:      Update $\epsilon_{new} \leftarrow \epsilon_{new} + \epsilon + \alpha(\alpha - 1) \cdot \epsilon \cdot \epsilon_{new}$
15:    **end for**
16:    Convert ADP privacy consumption to traditional privacy consumption.
17:    $\epsilon_{temp} \leftarrow \frac{\log(\epsilon_{new} \cdot \alpha(\alpha-1)+1)}{\delta(\alpha-1)}$
18:    **if** $\epsilon_{temp} < \epsilon_{bound}$ **then**
19:      Update $\sigma_{min} \leftarrow \sigma_G$, $\alpha^* \leftarrow \alpha$
20:      **break**
21:    **end if**
22:    **end for**
23: **end for**
24: **Return** σ_{min}, α^*

---

high or low values of α can adversely impact the dependent privacy parameters, thereby affecting overall performance. This highlights the necessity of dynamically selecting α based on specific constraints to optimize privacy requirements or data utility. Our simulations indicate that evaluating a small range of α values—typically between 2 and 300—is generally sufficient to identify the optimal choice.
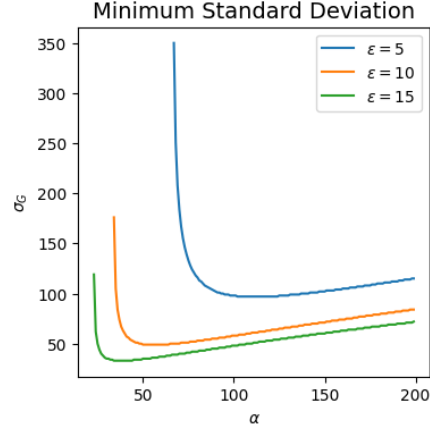

Figure 5: Optimal standard deviation (σ_G) of the Gaussian mechanism over 1000 iterations under varying required upper bounds of cumulative privacy consumption, with a required failure rate of $\delta = 1e\text{-}5$ and a required sensitivity of $\Delta f_2 = 1$. The horizontal axis represents the value of α, while the vertical axis indicates the standard deviation (σ_G).

8   Simulation and Discussion

In this section, we conduct simulations to explore the privacy consumption of different differential privacy frameworks, including alpha differential privacy (ADP), Rényi Differential Privacy (RDP), Zero-Concentrated Differential Privacy (zCDP), and the Advanced Composition (Adv.) theorem [19] under different iteration scenarios. This section is further divided into three parts: the simulation settings, the obtained results, and a detailed discussion of the observed trends.

8.1   Simulation Settings

Our simulations concentrate on the privacy consumption of the Gaussian mechanism, which is extensively employed in numerous differential privacy applications owing to its advantageous characteristics, especially in maintaining the utility of processed data. We choose the results generated by Gaussian as the basis for comparison to ensure that the results reflect real-world settings where different privacy approaches are often adopted. The simulations adjust the number of iterations and the failure probability δ, to compare the performance of each mechanism under different conditions, ranging from a minimum to a wide range of iterations and varying δ values. We also establish the variance parameter σ_G at various levels ($\sigma_G = 10, 50, 100$) to assess the sensitivity of each privacy framework to this variable. For alpha differential privacy (ADP) and Rényi differential privacy (RDP), we demonstrate the results under their optimal parameter choice for α, providing an evaluation of their performance when optimally configured. It should be noted that for the curves of Adv., δ represents the overall δ after applying the advanced composition theorem.

For Figure 6, the ϵ values for a single query under the ADP framework are $5.00e\text{-}5$, $5.05e\text{-}5$, and $5.24e\text{-}5$, respectively. Similarly, for Figure 7, the ϵ values for a single query under ADP are $5.29e\text{-}5$, $5.19e\text{-}5$, and $5.12e\text{-}5$, respectively. For Figure 8, the ϵ values for a single query under ADP are $5.02e\text{-}3$, $2.01e\text{-}4$, and $5.04e\text{-}5$, respectively. It is important to emphasize that these ϵ values represent the privacy parameter ϵ defined within the ADP frameworks, rather than the privacy consumption in the traditional (ϵ, δ)-differential privacy framework mentioned earlier. These data are provided here for the readers' reference.

8.2   Simulation Results

Our simulation results are shown in Figures 6, 7, and 8, which compare the privacy consumption of different differential privacy mechanisms, including ADP, RDP, zCDP, and Advanced Composition, in detail. These figures aim to illustrate the effectiveness of each privacy framework under varying conditions, such as different numbers of iterations and different values of the failure probability δ. In these figures, the horizontal axis represents the number of iterations, while

the vertical axis shows the corresponding privacy consumption. This visualization allows for a detailed comparison of how each mechanism performs in terms of cumulative privacy consumption over multiple iterations.

Figure 6 shows the privacy consumption trends of different mechanisms under three different δ values (1e-5, 1e-10, and 1e-15) with a small number of iterations. The main observation is that although the privacy consumption of both ADP and RDP estimates shows a linear growth, the lower intercept of ADP shows that it provides a stronger initial privacy estimate.

Figure 7 shows the privacy consumption when the number of iterations is relatively large, where δ is set to $1e$-5. Here, we observe the performance difference between ADP and other frameworks as the number of iterations increases. While ADP starts with a relatively lower privacy consumption, it has a steeper slope compared to RDP, increasing the cumulative privacy consumption as the number of iterations increases.
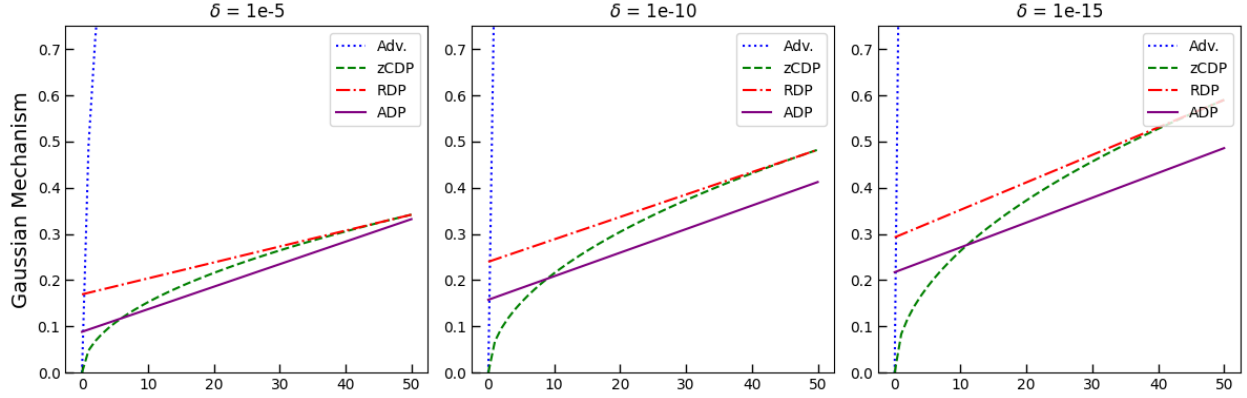

Figure 6: Comparison of cumulative privacy consumption between alpha differential privacy and other mainstream differential privacy frameworks for the Gaussian mechanism under small iterations, with a fixed standard deviation $\sigma_G = 100$ and a fixed sensitivity $\Delta f_2 = 1$. The horizontal axis represents the number of iterations, and the vertical axis represents the corresponding privacy consumption (ϵ). The blue line represents the advanced composition of differential privacy, the red line represents Rényi differential privacy, the green line represents zero-concentrated differential privacy, and the purple line represents alpha differential privacy. The results are shown for three different values of the failure probability: $\delta =1e$-5, $\delta =1e$-10, and $\delta =1e$-15. For ADP, the selected α values to minimize cumulative privacy consumption are 136, 152, and 164, respectively, while for RDP, the corresponding α values are 69, 97, and 119.

Figure 8 provides the privacy consumption under strict failure probability requirements ($\delta =1e$-25). In this case, ADP maintains a clear advantage over RDP regarding the increasing slope of the privacy consumption curve, indicating that its cumulative privacy consumption grows slower than RDP as δ becomes smaller.

Figures 6, 7, and 8 illustrate that although zCDP initially exhibits a higher growth rate of privacy consumption, its logarithmic growth becomes favourable over a wide range of iterative scenarios.

Each figure also highlights the consistent behaviour of the advanced combination mechanism, which shows higher privacy consumption in all scenarios compared to ADP, RDP, and zCDP.

8.3 Discussion

One of the key observations from the results is the behaviour of ADP versus RDP under small iterations. Figure 6 illustrates that both ADP and RDP demonstrate linear growth in privacy consumption. Nonetheless, ADP has a continually lower intercept, signifying a diminished initial privacy consumption. This attribute indicates that ADP is very efficient in situations necessitating a limited number of repeats, hence providing enhanced initial privacy assurances relative to RDP. However, this initial benefit gradually diminishes as the number of iterations increases, with ADP having a higher growth rate of privacy consumption than RDP, ultimately leading to a larger cumulative privacy consumption over a large number of iterations. This trade-off must be meticulously evaluated when choosing a differential privacy method, especially for applications that entail repetitive queries.

Another important finding in the results is related to zCDP. As the number of iterations increases, its logarithmic growth rate under the combination becomes increasingly favourable. The continued decline in the growth rate of zCDP with increasing iterations allows it to remain stable across various privacy settings. Despite the high initial growth rate of
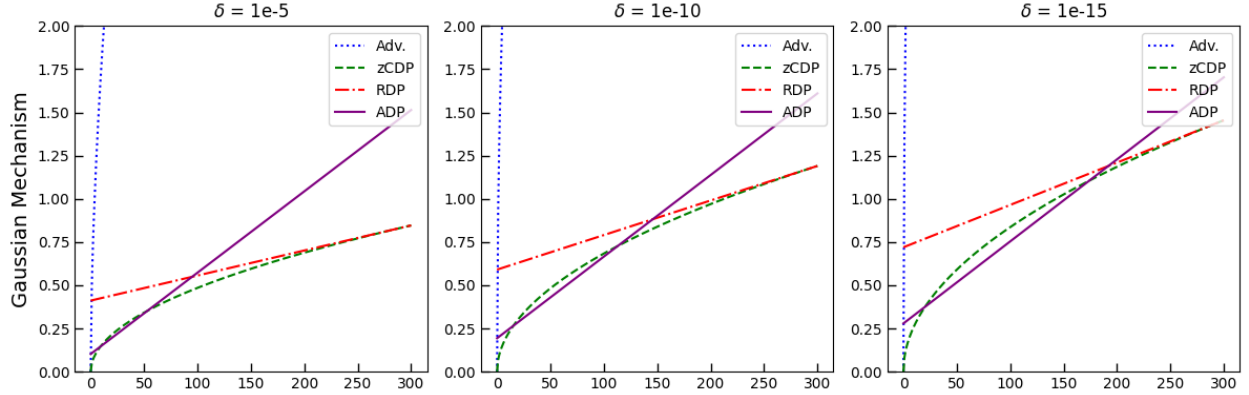
Figure 7: Comparison of cumulative privacy consumption between alpha differential privacy and other mainstream differential privacy frameworks for the Gaussian mechanism under relatively large iterations, with a fixed standard deviation $\sigma_G = 100$ and a fixed sensitivity $\Delta f_2 = 1$. The horizontal axis represents the number of iterations, and the vertical axis represents the corresponding privacy consumption (ϵ). The blue line represents the advanced composition of differential privacy, the red line represents Rényi differential privacy, the green line represents zero-concentrated differential privacy, and the purple line represents alpha differential privacy. The results are shown for three different values of the failure probability $\delta = 1e$-5, $\delta = 1e$-10, and $\delta = 1e$-15. For ADP, the selected α values to minimize cumulative privacy consumption are 13, 64, and 127, respectively, while for RDP, the corresponding α values are 6, 25, and 49.
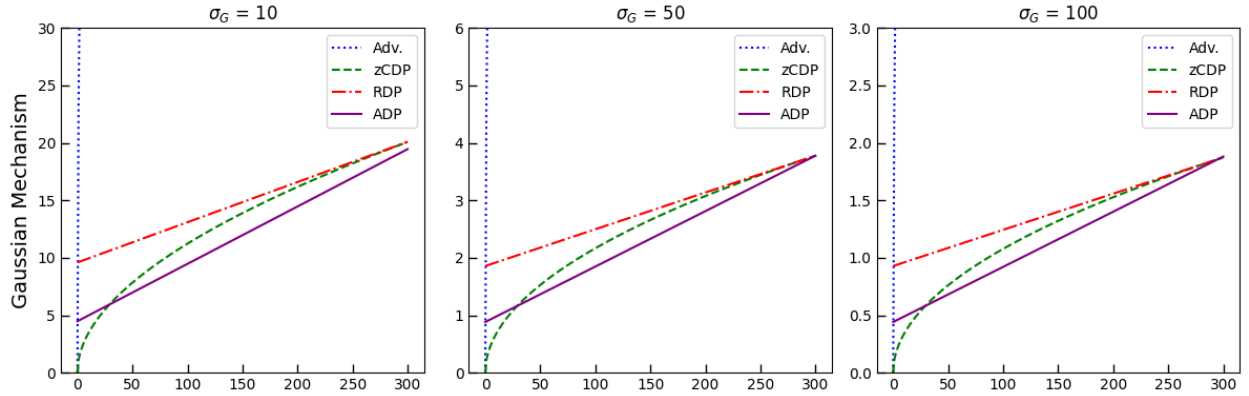

Figure 8: Comparison of cumulative privacy consumption between alpha differential privacy and other mainstream differential privacy frameworks for the Gaussian mechanism under relatively large iterations, with a fixed failure probability $\delta = 1e$-25 and a fixed sensitivity $\Delta f_2 = 1$. The horizontal axis represents the number of iterations, and the vertical axis represents the corresponding privacy consumption (ϵ). The blue line represents the advanced composition of differential privacy, the red line represents Rényi differential privacy, the green line represents zero-concentrated differential privacy, and the purple line represents alpha differential privacy. The results are shown for three different values of the standard deviation $\sigma_G = 10$, $\sigma_G = 50$, and $\sigma_G = 100$. For ADP, the selected α values to minimize cumulative privacy consumption are 14, 67, and 133, respectively, while for RDP, the corresponding α values are 7, 32, and 63.

privacy consumption, zCDP is well suited for situations where data needs to be accessed frequently or for long-term continuous analysis, where managing the cumulative privacy consumption is essential.

The findings also underscore a notable aspect of ADP under rigorous δ criteria, as illustrated in Figure 8. With a tight failure probability promise ($\delta = 1e$-25), the slope of the privacy consumption curve for ADP increases at a slower rate than that of RDP, indicating that ADP is especially appropriate for situations necessitating exceptionally rigorous privacy assurances. This attribute renders ADP beneficial for applications dealing with extremely sensitive data, where minimizing privacy consumption during repeated accesses is essential.

In contrast, the advanced composition framework, indicated by the blue line in all figures, consistently demonstrates the largest privacy consumption in every scenario. This persistently elevated expense constrains its applicability in contexts where reducing privacy consumption is a primary goal. The advanced composition approach may remain relevant in situations when simpler privacy accounting is favoured and computing speed is emphasized over the reduction of cumulative privacy consumption.

Observations above demonstrate that ADP exhibits compelling advantages in practical scenarios characterized by small to moderate iterations and stringent failure probability requirements (δ). These conditions are particularly prevalent in highly sensitive domains such as healthcare and finance, where robust privacy guarantees are imperative. In healthcare applications, particularly electronic health record (EHR) analysis, privacy regulations such as Health Insurance Portability and Accountability Act (HIPAA) mandate extraordinarily stringent privacy safeguards. Typical scenarios involve constrained query patterns (approximately less than 100 iterations) with extremely low failure probabilities (e.g., $\delta = 1e$-15 or smaller) to protect sensitive patient information. In such cases, ADP enables dynamic selection of α to minimize cumulative privacy consumption while meeting the given constraints. Our empirical analysis, as illustrated in Figure 6, demonstrates that with $\delta = 1e$-15 and 50 iterations, ADP achieves approximately 20% reduction in cumulative privacy consumption compared to existing frameworks like RDP and zCDP by determining an optimal α.

The advantages of ADP extend similarly to financial applications, where protecting sensitive financial data (e.g., account transactions, credit histories, investment portfolios) is crucial. In scenarios such as credit risk assessment and fraud detection systems, which typically require 50 to 200 iterations, ADP's adaptive framework demonstrates superior performance. As evidenced in Figure 8, under the extreme constraint of failure probability ($\delta = 1e$-25), ADP outperforms other privacy frameworks in minimizing cumulative privacy consumption. These scenarios highlight the practical significance of ADP in real-world applications.

9   Conclusion and Future Work

This section concludes the findings of our research and outlines potential directions for future work.

9.1   Conclusion

The results of this study demonstrate that alpha differential privacy (ADP) is particularly appropriate for applications with small to moderate iterations, especially in settings where the failure probability needs to be strictly limited. Alpha divergence provides ADP with the necessary flexibility to fine-tune privacy consumption while achieving a customized balance between privacy and utility. In the small iteration setting, ADP has a unique advantage in that it can evaluate the initial privacy consumption more strictly than other privacy frameworks. This feature is particularly advantageous in privacy-sensitive applications where low privacy consumption in small iterations and failure probability are essential, such as in healthcare or financial analytics.

In instances with high iteration counts, the performance of ADP requires careful evaluation due to the relatively large growth rate of privacy consumption. Simulation results show that the total privacy consumption under ADP can become significant as the number of iterations increases, especially when the failure probability δ is less restricted. Therefore, although ADP offers specific advantages in the initial stage, its overall privacy cost may exceed that of other differential privacy frameworks such as Rényi Differential Privacy (RDP) or zero-concentrated differential privacy (zCDP) during long-term iterations. Practitioners must carefully evaluate the iteration requirements and privacy constraints of their specific applications before choosing ADP as a privacy framework.

ADP offers a promising enhancement to conventional differential privacy models, providing refined privacy assurances that can be adjusted to satisfy particular needs. Nonetheless, its constraints in extensive iteration scenarios underscore the necessity of evaluating context-specific criteria while selecting among various privacy frameworks. Evaluating ADP's early advantages alongside its possible disadvantages over extended durations is a crucial factor in its effective application.

9.2 Future Work

Future research could focus on advancing the practical applications of ADP to enhance its robustness and adaptability in diverse privacy-preserving contexts. Expanding ADP beyond the Gaussian mechanism to include other mechanisms like the Laplace and Exponential mechanisms may provide insights into its flexibility across different data distributions and queries, reinforcing its role as a versatile privacy framework.

Evaluating ADP in practical settings such as healthcare, and finance will be crucial to determining its real-world utility and assessing how its theoretical benefits translate into practice. Understanding its performance amidst data heterogeneity, dynamic updates, and varying privacy requirements will be key to optimizing its deployment.

Moreover, integrating ADP into machine learning and deep learning systems could open up new possibilities for privacy-preserving models. This research could explore how ADP can be effectively incorporated into federated learning or privacy-preserving optimization while ensuring model accuracy and managing privacy consumption over multiple training iterations.

References

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference*, 2006, pp. 265-284.

[2] C. Dwork, "Differential privacy," in *International Colloquium on Automata, Languages, and Programming*, 2006, pp. 1-12.

[3] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308-318.

[4] N. Wang, X. Xiao, Y. Yang, J. Zhao, S. C. Hui, H. Shin, J. Shin, and G. Yu, "Collecting and analyzing multidimensional data with local differential privacy," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 2019, pp. 638-649.

[5] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 493-502.

[6] B. Balle and Y.-X. Wang, "Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising," in *International Conference on Machine Learning*, 2018, pp. 403-412.

[7] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," *IEEE Transactions on Information Theory*, vol. 61, no. 6, pp. 3391-3406, 2015.

[8] S. Amari, *Information Geometry and Its Applications*. Springer, 2016.

[9] I. Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 4, pp. 417-528, 2004.

[10] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211-407, 2014.

[11] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology-EUROCRYPT 2006*, St. Petersburg, Russia, 2006, pp. 486-503, Springer.

[12] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 263-275.

[13] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *Theory of Cryptography Conference*, 2016, pp. 635-658, Springer.

[14] A. Cichocki, H. Lee, Y.-D. Kim, and S. Choi, "Non-negative matrix factorization with α-divergence," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1433–1440, 2008.

[15] T. Villmann and S. Haase, "Mathematical aspects of divergence based vector quantization using Fréchet-derivatives," *University of Applied Sciences Mittweida*, 2010.

[16] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299-318, 1967.

[17] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797-3820, Jul. 2014. [Online]. Available: arxiv.org/abs/1206.2459.

[18] F. D. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," in *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, C. Binnig and B. Dageville, Eds., 2009, pp. 19-30.

[19] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS)*, 2010, pp. 51–60. doi: 10.1109/FOCS.2010.12.