A unifying framework for differentially private quantum algorithms

Armando Angrisani, Mina Doosti, Elham Kashefi

A unifying framework for differentially private quantum algorithms

Armando Angrisani *,[1], Mina Doosti[2], and Elham Kashefi[1,2]

[1]LIP6, CNRS, Sorbonne Université, 75005 Paris, France
[2]School of Informatics, University of Edinburgh, EH8 9AB Edinburgh, United Kingdom

July 11, 2023

Abstract

Differential privacy is a widely used notion of security that enables the processing of sensitive information. In short, differentially private algorithms map "neighbouring" inputs to close output distributions. Prior work proposed several quantum extensions of differential privacy, each of them built on substantially different notions of neighbouring quantum states. In this paper, we propose a novel and general definition of neighbouring quantum states. We demonstrate that this definition captures the underlying structure of quantum encodings and can be used to provide exponentially tighter privacy guarantees for quantum measurements. Our approach combines the addition of classical and quantum noise and is motivated by the noisy nature of near-term quantum devices. Moreover, we also investigate an alternative setting where we are provided with multiple copies of the input state. In this case, differential privacy can be ensured with little loss in accuracy combining concentration of measure and noise-adding mechanisms. *En route*, we prove the advanced joint convexity of the quantum hockey-stick divergence and we demonstrate how this result can be applied to quantum differential privacy. Finally, we complement our theoretical findings with an empirical estimation of the certified adversarial robustness ensured by differentially private measurements.

*corresponding author: armando.angrisani@lip6.fr

1 Introduction

In recent years, the availability of large datasets and advanced computational tools has sparked progress across various fields, including natural sciences, medicine, finance, and social sciences. This advance came also with privacy concerns since even the release of aggregated data can compromise the sensitive information contained in the original dataset. This poses a significant challenge for the researcher, who must adopt privacy-preserving techniques to avoid the exposure of private data. Privacy-preserving data processing is a non-trivial task, and ill-defined notions of privacy led to impressive privacy breaches [1]. This motivated the quest for a robust framework to assess privacy.

Over the last decade, differential privacy (DP) has become the de facto standard for ensuring privacy both in statistical data analysis and machine learning applications [2, 3, 4, 5, 6, 7, 8]. Intuitively, a differentially private algorithm $\mathcal{A}(\cdot)$ can learn a statistical property of a dataset consisting of n elements, yet it leaks *almost* nothing about each individual element. In other words, given two inputs x and x' which are very close according to some chosen metric, the output distributions $\mathcal{A}(x)$ and $\mathcal{A}(x')$ should be almost indistinguishable. We call x and x' neighbouring inputs. If x and x' represent datatsets about n individuals, then it's customary to consider x and x' neighbouring if one of such individuals is present in x and absent in x'. Then, if $\mathcal{A}(\cdot)$ is differentially private, the output alone doesn't allow for inferring whether the input contained a given individual. This goal is pursued by combining various techniques, that usually involve randomising the input or perturbing the output by adding noise. The challenge is then to achieve the desired level of privacy by adding less noise as possible, hence preserving accuracy.

Apart from privacy-preserving data analysis and machine learning, differential privacy has also found several applications in other fields of computer science such as statistical learning theory [9, 10, 11, 12], adaptive data analysis [13, 14, 15] and mechanism design [16].

More recently, the major influence of quantum computing and quantum information has led to the exploration of differentially private quantum algorithms. Since many near-term quantum algorithms involve a classical optimiser as a subroutine, one possible approach consists in privatising such optimiser and leaving the rest of the algorithm unchanged. This strategy is adopted in [17, 18, 19, 20].

Alternatively, we can rely on several notions of *quantum* differential privacy. Quantum differential privacy allows the design of private measurements and channels combining classical and quantum noise. This is extremely relevant with the emergence of Noisy Intermediate Scale Quantum devices (NISQ) today [21]. The noisy nature of these devices on the one hand, and the potential capabilities of quantum algorithms, on the other hand, make such quantum or hybrid quantum-classical mechanisms, an interesting subject of study from the point of view of privacy. Several efforts has been made in this area of research, including [22, 23, 24, 25, 26]. Furthermore, the connection between machine learning and differential privacy [15, 27] suggests that exploring quantum differential privacy can lead to intriguing insights into the capabilities of quantum machine learning.

One of the main challenges in translating the definition of DP in the quantum setting is to characterise the notion of neighbouring quantum states, i.e. choose the right metric to measure the similarity between the input states. The first notion of quantum differential privacy was proposed in [22] and it's based on bounded trace distance, whereas the definition introduced in [23] is based on reachability by a single-qudit operation. Another possible definition is based on the quantum Wasserstein distance of order 1. This metric was introduced in [28] and

the authors mention quantum differential privacy as one potential application of their work. Furthermore, quantum private PAC learning has been defined in [29] and a quantum analogue of the equivalence between private classification and online prediction has been shown in [12]. Moreover, an equivalence between learning with quantum local differential privacy and quantum statistical query (QSQ) learning was provided in [30]. Other authors compared classical and quantum mechanisms in the context of local differential privacy [31, 32]. Building upon these prior contributions, the present paper aims at establishing a general framework for differentially private quantum algorithms, providing a more general definition of neighbouring quantum states and attaining better privacy guarantees combining classical and quantum noisy channels.

1.1 Motivation: connecting neighbouring relationships with quantum encodings

Our work is motivated by a practical goal: we want to design quantum algorithms that satisfy differential privacy with respect to a classical input x. We assume that this input belongs to a set equipped with a neighbouring relationship. Moreover, we consider quantum algorithms that include a quantum encoding as a subroutine, where x is mapped to a quantum state $\rho(x)$. Thus, we want to define a quantum neighbouring relationship that mimics the underlying classical neighbouring relationship. In particular, we require the following property:

$$x \text{ and } x' \text{ are neighbouring} \implies \rho(x) \text{ and } \rho(x') \text{ are neighbouring.}$$

It's easy to see why the above property is extremely useful. If an algorithm $\mathcal{A}$ is (ε, δ)-differentially private with respect to $\rho(x)$, then $\mathcal{A} \circ \rho$ is (ε, δ)-differentially private with respect to x (this is stated more formally in Proposition 3.2). In the meantime, we want to avoid neighbouring relationships that are excessively loose, as this would make the output almost independent of the input. A paradigmatic example of a "pathological" relationship is the one based on *constant* trace distance:

$$\rho \text{ and } \rho' \text{ are neighbouring} \iff \frac{1}{2}\|\rho - \rho'\|_1 \leqslant \tau = \Theta(1).$$

To fix the ideas, let $\tau = 0.1$. It's easy to see that for any pair of states ρ, σ we can build a sequence $\rho_0, \rho_2, \ldots, \rho_{10}$, such that

$$\begin{cases} \rho_0 = \rho \\ \rho_{10} = \sigma \end{cases} \qquad \text{and for all } i,\ \rho_i \sim \rho_{i+1}.$$

By triangle inequality, the outputs of ρ and σ will be (10ε)-close. This, significantly limits the capabilities of private algorithms, since, independently of the input states, the output distribution would be highly concentrated around the same value. We discuss this more formally in Section 7.

Surprisingly, the current quantum neighbouring relationships fulfil these two natural requirements only for a limited number of specific quantum encodings. Given that this property is crucial for the framework of differential privacy and the need to handle various types of quantum and classical data in the quantum setting, it becomes necessary to develop an approach that can account for different encodings. Therefore, we present a generalised neighbouring relationship that enables the handling of a wide range of near-term and long-term algorithms.

1.2 Overview of main results

In this work, we tackle several technical problems arising in the field of quantum differential privacy and we try to address them within a broader framework using different tools and techniques from quantum information. The following is a summary of our main contributions.

- **Improved privacy bounds for noisy channels.** Our first contribution consists of tighter privacy guarantees for a general family of noisy channels, which includes local Pauli noise and particularly as a special case, the depolarising channel. To this end, we prove the advanced joint convexity of the quantum hockey-stick divergence. Moreover, we provide a tighter analysis of the privacy of quantum measurements post-processed with classical stochastic channels, such as the Laplace or Gaussian noise. This approach allows us to be able to study both classical and quantum noisy mechanisms for differential privacy, within a unified framework.

- **Generalised neighbouring relationship.** Our second contribution is a generalised neighbouring relationship, that allows us to recover the previous definition as special cases. We demonstrate how to design differentially private measurements according to this definition by introducing both classical and quantum noise into the computation. Notably, we show that local measurements can be made differentially private by adding a modest amount of noise. Our work is the first to incorporate the locality in the analysis of quantum differential privacy.

- **Privacy-utility tradeoff for quantum differential privacy.** There exists an unavoidable tradeoff between the desired level of privacy and the resulting loss in accuracy. Here, we make a crucial observation: different neighbouring relationships have different tradeoffs. In particular, this limits the applicability of neighbouring relationships based solely on the bounded trace distance. We also show no-go results for pure quantum differential privacy under the Wasserstein distance of order 1.

- **Private estimation with multiple copies.** We provide differentially private mechanisms for estimating the expected values of observables given m copies of a quantum state. These mechanisms can find applications in privatising the results of experiments on physical devices where estimating the expectation value is the main figure of merit.

- **Applications.** Our results can be applied to variational quantum algorithms and other quantum machine learning models to enhance or certify privacy. We specifically focus on certified adversarial robustness through differential privacy and we perform numerical simulations to assess the robustness to adversarial attacks of private quantum classifiers.

1.3 Organisation of the paper

The paper is organised as follows. First, in Section 2 we give a brief background on the notion and definition of differential privacy in the classical world, as well as introduce the notations we are using throughout the paper. In Section 3, we discuss quantum differential privacy and the differences among different approaches to defining differential privacy in the quantum world, specifically the neighbouring relationship between quantum states, and we prove that classical differential privacy can be ensured through quantum differential privacy. In Section 4

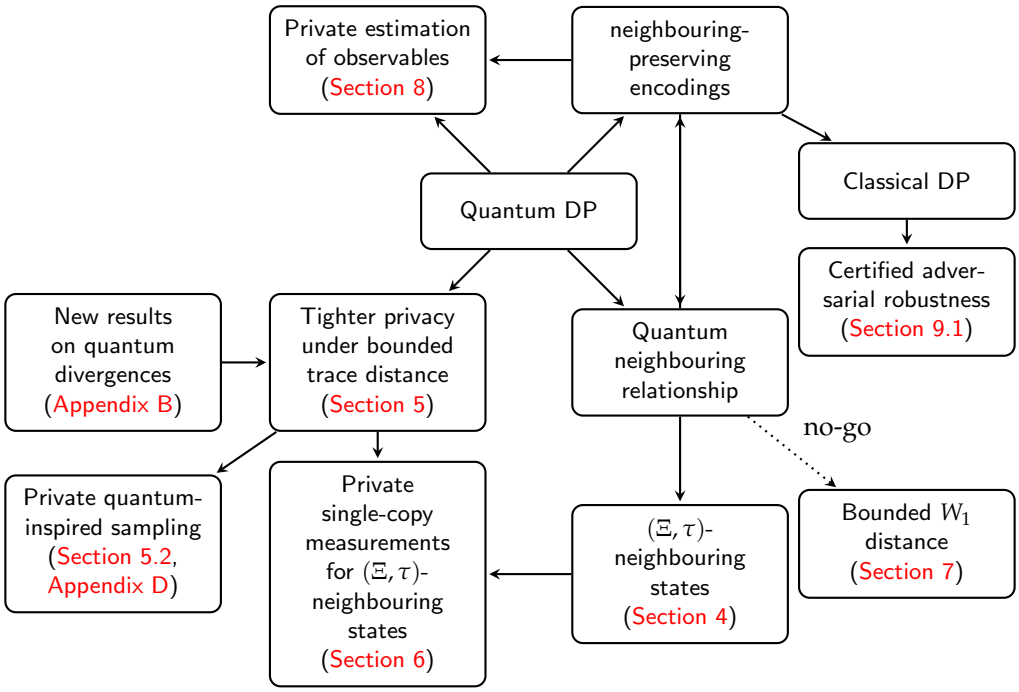

Figure 1: High-level summary of the main concepts covered in the paper and their mutual dependencies. Our contributions spans includes tighter information-theoretic bounds, novel approaches to quantum differential privacy and applications to private quantum machine learning. We remark that the no-go results for W_1 distance holds for single-copy measurements under pure differential privacy with respect to mixed states with bounded W_1 distance. On the other hand, the W_1 distance can be conveniently used for the private estimation of expected values of observables with multiple copies.

we introduce our generalised framework for quantum differential privacy and we discuss its properties. Within this formal framework, we prove several results. Starting with Section 5, we provide several improved privacy bounds for the case where the neighbouring relationship is specified with a bounded trace distance between quantum states. Our results include classical post-processing and quantum-inspired sampling mechanisms. Our techniques hinge on a novel result information-theoretic result, namely the advanced joint-convexity of the quantum hockey-stick divergences, discussed in detail in Appendix B, along with a proof of the quantum Bretagnolle-Huber inequality. We then turn to the unique properties of our framework in Section 6 which allows us to study local measurements as quantum differentially private mechanisms, as well as addressing the question of how quantum and classical noise can be studied together in the context of differential privacy. In Section 7 we define the cost of differential privacy and benchmark different approaches and notions of neighbouring, under this lens, providing negative and positive results which clarify and justify the applicability of our framework. In Section 8 we introduce mechanisms for privately estimating expectation values. Finally, in Section 9, we discuss applications of some of our results in quantum machine learning, particularly for certified adversarial robustness, and we support our theoretical findings with numerical simulations.

2 Background

We start by introducing the notations we use in the paper as well as essential definitions of classical differential privacy. Other technical tools such as different norms and divergences are introduced in Appendix A.

2.1 Notation

We denote by $\log(\cdot)$ the natural logarithm. We denote by $P(X)$ the power set of a set X, i.e. the set of all subsets of X. For a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, we denote as $\|\boldsymbol{x}\|_p$ its p-norm, where $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $1 \leqslant p < \infty$ and $\|\boldsymbol{x}\|_\infty = \max_i \|x_i\|$. It's convenient to introduce also the 0-norm (which is technically not a norm): $\|\boldsymbol{x}\|_0 = |\{i : x_i \neq 0\}|$, which is the number of the non-zero entries of $\boldsymbol{x}$.

We consider a set V corresponding to a system of $|V| = n$ qudits, and denote by $\mathcal{H}_n = \bigotimes_{v \in V} \mathbb{C}^d$ the Hilbert space of n qudits. We denote by $\mathcal{L}(\mathcal{H}_n)$ the set of linear operators on $\mathcal{H}_n$. We denote by $\mathcal{O}_n$ the set of self-adjoint linear operators on $\mathcal{H}_n$. By $\mathcal{O}_n^+$ we denote the subset of positive semidefinite linear operators on $\mathcal{H}_n$, and $\mathcal{S}_n \subset \mathcal{O}_n^+$ denotes the set of quantum states. Similarly, we denote by $\mathcal{P}_n$ the set of probability measures on $[d]^V$. For any subset $A \subseteq V$, we use the standard notation $\mathcal{O}_A, \mathcal{S}_A, \ldots$ for the corresponding objects defined on subsystem A. Given a state $\rho \in \mathcal{S}_n$, we denote by ρ_A its marginal on subsystem A. For any $X \in \mathcal{O}_n$, we denote by $\|X\|_p$ its Schatten p norm. For any subset $A \subseteq V$, the identity on $\mathcal{O}_A$ is denoted by $\mathbb{1}_A$, or more simply $\mathbb{1}$. Given an observable O, we define $\langle O \rangle_\sigma = \mathrm{Tr}[\sigma O]$. Moreover, given a number $a \in \mathbb{R}$, we define $\{O \geqslant a\}$ to be the projector onto the subspace spanned by the eigenvectors of O corresponding to eigenvalues greater than or equal to a. We denote the probability of measuring an eigenvalue of O greater than $a \in \mathbb{R}$ in state σ as $\Pr_\sigma(O \geqslant a) := \mathrm{Tr}[\sigma\{O \geqslant a\}]$. For a subset F of the spectrum of O, we denote the probability of that the measurement outcome lies in F as $\Pr_\sigma[O \in F]$. For an observable O, we write its Pauli expansion as $O = \sum_{P \in \{X,Y,Z,\mathbb{1}\}^n} c_P P$. We say that a Pauli string $P = P_1 P_2 \ldots P_n \in \{X, Y, Z, \mathbb{1}\}^n$ acts non trivially on $\mathcal{I} \subseteq [n]$ if $c_P \neq 0$ and $\exists i \in \mathcal{I} : P_i \neq \mathbb{1}$. A quantum channel $\mathcal{N} : V \to W$ is a linear completely positive and trace-preserving map from the operators on $\mathcal{H}_{|V|}$ to the operators on $\mathcal{H}_{|W|}$. Similarly, a classical channel can be defined as a randomised mapping from V to W. We'll refer to a channel as an algorithm when we want to emphasise the input-output relationship. For a channel Φ, either classical or quantum, we denote as $\mathsf{range}(\Phi)$ the set of all the possible outputs of Φ. Given a quantum channel Φ acting on n qubits, we define its light-cone as follows: first, for any qubit i, we denote by $\mathcal{I}_i$ the minimal subset of qubits such that $\mathrm{Tr}_{\mathcal{I}_i} \Phi(\rho) = \mathrm{Tr}_{\mathcal{I}_i} \Phi(\sigma)$ for any two n-qubit states ρ and σ such that $\mathrm{Tr}_i(\rho) = \mathrm{Tr}_i(\sigma)$. Then, the light-cone of Φ is defined as $|\mathcal{I}| := \max_{i \in [k]} |\mathcal{I}_i|$.

2.2 Classical differential privacy

We concisely introduce the definition of differential privacy. For a comprehensive introduction to the topic, we refer to [3], [33] and [4]. Throughout this paper, we'll denote by $\sim$ the neighbouring condition, i.e. a relationship between two inputs, consisting of either classical vectors or quantum states. We'll write $\overset{Q}{\sim}$ when we want to emphasise that the neighbouring relationship refers to quantum states. The choice of the relationship is problem-dependent. In many practical cases, it's convenient to say that two binary vectors $x, x' \in \{0,1\}^n$ are neighbouring if

their Hamming distance is at most one, i.e.

$$x \sim x' \iff d_H(x, x') \leqslant 1.$$

In alternative, we can select a p-norm and a threshold $\gamma \geqslant 0$ and opt for the following neighbouring relationship:

$$x \sim x' \iff \|x - x'\|_p \leqslant \gamma.$$

We say that a randomised algorithm $\mathcal{A}(\cdot)$ is (ε, δ)-differentially private (DP) if for all $x \sim x'$ and for all $S \subseteq \mathsf{range}(\mathcal{A})$, it satisfies

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

We say that $\mathcal{A}(\cdot)$ is ε-DP when it is $(\varepsilon, 0)$-DP. Equivalently, differential privacy can be defined in terms of hockey-stick divergence E_γ and the smooth max-relative entropy (or smooth max-divergence) D_∞^δ:

$$\mathcal{A} \text{ is } (\varepsilon, \delta)\text{-DP} \iff \forall x \sim x' : E_{e^\varepsilon}(\mathcal{A}(x) \| \mathcal{A}(x')) \leqslant \delta \iff \forall x \sim x' : D_\infty^\delta(\mathcal{A}(x) \| \mathcal{A}(x')) \leqslant \varepsilon,$$

where the (classical) hockey-stick divergence E_γ between two distributions P and Q is defined as follows [34]:

$$E_\gamma(P \| Q) := \frac{1}{2} \int |\mathrm{d}P - \gamma \mathrm{d}Q| - \frac{1}{2}(\gamma - 1),$$

for $\gamma \geqslant 1$. These information-theoretic divergences can be thought of as a measure of closeness between distributions, thus these reformulations are consistent with the intuition that private algorithms map neighbouring inputs to "close" output distributions. Differential privacy with $\delta = 0$ is also referred to as *pure* differential privacy, whereas the case with $\delta \neq 0$ is referred to as *approximate* differential privacy. Roughly speaking, an (ε, δ)-DP algorithm can be thought of as an algorithm that is ε-DP with probability $1 - \delta$. We remark that this intuition is slightly imprecise, and thus we refer to the following references for a more detailed explanation [35, 36, 33].

It's also worth noticing that the max-divergence corresponds to the Rényi divergence of order ∞. Thus, it's possible to relax pure differential privacy by replacing the max-divergence with the Rényi divergence of order α, for $\alpha \geqslant 1$ [37]. We say that $\mathcal{A}$ is (α, ε)-RDP (Rényi differentially private) if for all $x \sim x'$,

$$D_\alpha(\mathcal{A}(x) \| \mathcal{A}(x')) \leqslant \varepsilon.$$

As a consequence, for all $S \subseteq \mathsf{range}(\mathcal{A})$, we have

$$\Pr[\mathcal{A}(x) \in S] \leqslant e^{\varepsilon} \Pr[\mathcal{A}(x') \in S]^{(\alpha-1)/\alpha}.$$

If $\mathcal{A}$ is (α, ε)-RDP then it is also $(\varepsilon + \frac{\log(1/\delta)}{\alpha - 1}, \delta)$-DP for any $0 < \delta < 1$. Similarly, if $\mathcal{A}$ is $(\varepsilon, 0)$-DP then it is also $(\alpha, 2\alpha\varepsilon^2)$-RDP for any $\alpha \geqslant 1$.

Privacy via classical noisy channels

Now we present two widely used mechanisms that ensure differential privacy by injecting noise into the output. To this end, we introduce two classical channels $\Lambda_{\mathcal{L},b} : \mathbb{R} \to \mathbb{R}$ and $\Lambda_{\mathcal{G},\sigma} : \mathbb{R} \to \mathbb{R}$, that corresponds to an additive noise coming from either the Laplace distribution

of scale b or the Gaussian distribution of variance σ^2, both centred in zero. The channels are defined as follows:

$$\Lambda_{\mathcal{L},b}(x) = x + \eta \quad \text{and} \quad \Lambda_{\mathcal{G},\sigma}(x) = x + \zeta,$$

where $\eta \sim \frac{1}{2b} \exp\left(-\frac{|\eta|}{b}\right)$ and $\zeta \sim \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\zeta^2}{2\sigma^2}\right)$.

Let $f : \mathcal{X} \to \mathbb{R}$ be a scalar function. We define the *sensitivity* of f as

$$\Delta_f := \max_{\substack{x,x' \in \mathcal{X} \\ x \sim x'}} |f(x) - f(x')|. \tag{1}$$

Then $\Lambda_{\mathcal{L},b}(f(\cdot))$ is ε-DP if $b \geqslant \Delta/\varepsilon$. Similarly, $\Lambda_{\mathcal{G},\sigma}(f(\cdot))$ is (ε, δ)-DP if $\sigma^2 \geqslant 2\ln(1.25/\delta)\Delta^2/\varepsilon^2$. The addition of Laplace noise is referred to as *Laplace mechanism* [2], whereas the addition of Gaussian noise is referred to as *Gaussian mechanism* [3]. Both mechanisms can also be analysed within the relaxed framework of Rényi differential privacy [37].

3 Quantum differential privacy

Let ρ, σ two neighbouring quantum states, i.e. $\rho \overset{Q}{\sim} \sigma$. We'll discuss appropriate neighbouring conditions for quantum states in the next sections and for the moment we use the letter Q as a placeholder. We also say that ρ and σ are Q-neighbouring in order to emphasise that we selected a suitable relationship Q over quantum states. Following [22, 24], we say that a quantum channel $\mathcal{C}(\cdot)$ is (ε, δ)-DP if for all $\rho \overset{Q}{\sim} \sigma$, for all POVM $M = \{M_m\}$ and for all m, we have that

$$\mathrm{Tr}[M_m \mathcal{C}(\rho)] \leqslant e^{\varepsilon} \mathrm{Tr}[M_m \mathcal{C}(\sigma)] + \delta.$$

As in the classical case, this can be equivalently expressed in terms of the quantum hockey-stick divergence or the quantum smooth max-relative entropy:

$$\begin{aligned} \mathcal{C} \text{ is } (\varepsilon, \delta)\text{-DP} &\iff \forall \rho, \sigma : \rho \overset{Q}{\sim} \sigma : E_{e^\varepsilon}(\mathcal{C}(\rho) \| \mathcal{C}(\sigma)) \leqslant \delta \\ &\iff \forall \rho, \sigma : \rho \overset{Q}{\sim} \sigma : D_\infty^\delta(\mathcal{C}(\rho) \| \mathcal{C}(\sigma)) \leqslant \varepsilon, \end{aligned}$$

where the quantum hockey-stick divergence E_γ is defined as follows:

$$E_\gamma(\rho \| \sigma) := \mathrm{Tr}(\rho - \gamma\sigma)^+,$$

for $\gamma \geqslant 1$. Here X^+ denotes the positive part of the eigendecomposition of a Hermitian matrix $X = X^+ - X^-$. We refer to Lemma III.2 in [24] for more details. A special case of particular interest is one of quantum-to-classical channels (i.e. POVM measurements), mapping states to probability distributions. For a measurement $\mathcal{M}$, denote as $\mathcal{M}(\rho)$ the probability distribution induced by measuring $\mathcal{M}$ on input ρ. Quantum differential privacy shares many useful properties with classical differential privacy. Notably, it is robust to parallel composition and post-processing (also referred to as sequential composition).

Proposition 3.1 (Adapted from Corollary III.3, [24])**.** *The following properties hold.*

- *(Post-processing) Let $\mathcal{A}$ be (ε, δ)-differentially private and $\mathcal{N}$ be an arbitrary quantum channel, then $\mathcal{N} \circ \mathcal{A}$ is also (ε, δ)-differentially private.*

- *(Parallel composition) Let $\mathcal{A}_1$ be $(\varepsilon_1, \delta_1)$-differentially private and $\mathcal{A}_2$ be (ε_2, δ)-differentially private. Define that $\rho_1 \otimes \rho_2 \overset{Q}{\sim} \sigma_1 \otimes \sigma_2$ if $\rho_1 \overset{Q}{\sim} \sigma_1$ and $\rho_2 \overset{Q}{\sim} \sigma_2$. Then $\mathcal{A}_1 \otimes \mathcal{A}_2$ is $(\varepsilon_1 + \varepsilon_2, \bar{\delta})$-differentially private on such product states, with $\bar{\delta} = \min\{\delta_1 + e^{\varepsilon_1}\delta_1, e^{\varepsilon_2}\delta_1 + \delta_2\}$.*

  Moreover, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are quantum-classical channels (measurements), we have that $\mathcal{A}_1 \otimes \mathcal{A}_2$ is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-differentially private.

Proof. The proposition coincides with Corollary III.3 in [24], except for the final statement about the parallel composition of differentially private measurements. Since the output of a measurement is a classical distribution, the proof of this part is identical to the one of Theorem 3.16 in [3]. □

In short, the composition theorem ensures that performing k times an ε-DP algorithm is (εk)-differentially private, and then the privacy budget scales as the number of repetitions k. However, under mild assumptions, this scaling can be improved to $O(\sqrt{k})$. This result is called *advanced composition* (we refer to Theorem 3.20 in [3] for the classical case). Moreover, advanced composition holds also for quantum measurements under suitable assumptions (Theorem 6, [22]).

Rényi quantum differential privacy has also been defined in [24]. Due to the non-commutative nature of quantum mechanics, the quantum generalisation of the Rényi divergence is not unique. However, we don't need to fix a particular definition of the quantum Rényi divergence, since we can define Rényi quantum differential privacy in terms of an arbitrary family of Rényi divergences $\mathbb{D}_\alpha$, as defined in [38]. Thus, a quantum channel $\mathcal{C}$ is (ε, α)-Rényi differentially private if

$$\sup_{\rho \sim \sigma} \mathbb{D}_\alpha(\mathcal{C}(\rho) \| \mathcal{C}(\sigma)) \leqslant \varepsilon.$$

3.1 From quantum to classical differential privacy

Now we show how quantum differential privacy can be used as a proxy to ensure the privacy of a classical input encoded in a quantum state. First, we introduce a preliminary definition.

Definition 3.1 (Privacy-preserving quantum encodings)**.** *Let $\mathcal{X}$ a set equipped with a neighbouring relationship $\sim$. A quantum encoding $\rho(\cdot)$ is Q-neighbouring-preserving if*

$$x \sim x' \implies \rho(x) \overset{Q}{\sim} \rho(x').$$

The following proposition formalizes the intuitive fact that Q-neighbouring-preserving encodings can be used to transfer privacy guarantees and ensure the privacy of the underlying classical input.

Proposition 3.2 (Transferring privacy guarantees)**.** *Let $\rho(\cdot)$ a quantum encoding, i.e. a function mapping a classical vector $x \in \mathcal{X}$ to a quantum state $\rho(x)$. Assume $\mathcal{X}$ is equipped with a neighbouring relationship $\sim$ and $\mathcal{S}_n$ is equipped with a neighbouring relationship $\overset{Q}{\sim}$. Assume that $\rho(\cdot)$ is Q-neighbouring-preserving. Let $\mathcal{M}$ be a measurement. We have,*

$$\mathcal{M} \text{ is } (\varepsilon, \delta)\text{-DP with respect to } \overset{Q}{\sim} \implies \mathcal{M}(\rho(\cdot)) \text{ is } (\varepsilon, \delta)\text{-DP with respect to } \sim.$$

Proof. The proposition follows from the definition of differential privacy. Assuming $\mathcal{M}(\cdot)$ is (ε, δ)-DP, we have

$$\forall \sigma, \sigma' : \sigma \overset{Q}{\sim} \sigma',\ \forall S \subseteq \mathsf{range}(\mathcal{M}) : \Pr[\mathcal{M}(\sigma) \in S] \leqslant e^{\varepsilon} \Pr[\mathcal{M}(\sigma') \in S] + \delta.$$

Since $\rho(\cdot)$ is Q-neighbouring-preserving, the above inequality still holds if we set $\sigma := \rho(x)$ and $\sigma' := \rho(x')$ for $x \sim x'$. Moreover, we replace $\mathsf{range}(\mathcal{M})$ with $\mathsf{range}(\mathcal{M} \circ \rho(\cdot))$ (we can do it since $\mathsf{range}(\mathcal{M} \circ \rho(\cdot))$ is a subset of $\mathsf{range}(\mathcal{M})$). The result readily follows.

$$\forall x, x' : x \sim x',\ \forall S \subseteq \mathsf{range}(\mathcal{M} \circ \rho(\cdot)) : \Pr[\mathcal{M}(\rho(x)) \in S] \leqslant e^{\varepsilon} \Pr[\mathcal{M}(\rho(x')) \in S] + \delta.$$

□

4 Generalised neighbouring relationship

In this section, we present the cornerstone of our work, which is a general definition of neighbouring quantum states.

Definition 4.1. *Let $\rho, \sigma \in \mathcal{S}_n$ and let $\Xi \subset P([n])$, i.e. let Ξ be a collection of subsets of $[n]$. Let $\tau > 0$ be a parameter. We say that ρ and σ are (Ξ, τ)-neighbouring and we write $\rho \overset{(\Xi,\tau)}{\sim} \sigma$ if*

$$\exists \mathcal{I} \in \Xi : \mathrm{Tr}_{\mathcal{I}} \rho = \mathrm{Tr}_{\mathcal{I}} \sigma \ \wedge\ \frac{1}{2} \|\rho - \sigma\|_1 \leqslant \tau.$$

If $\Xi = \{\mathcal{I} : \mathcal{I} = \{i, i+1, \ldots, i+\ell\}$ for some $i\}$, i.e. each subset $\mathcal{I}$ is a collection of ℓ consecutive integers (modulo n), we say that ρ and σ are (ℓ, τ)-neighbouring and we write $\rho \overset{(\ell,\tau)}{\sim} \sigma$. When $\Xi = \{[n]\}$, we simply write $\rho \overset{\tau}{\sim} \sigma$ and we say that ρ and σ are τ-neighbouring.

This definition extends one of the neighbouring states used in previous works. In [22, 39, 24], two states are neighbouring if they have bounded trace distance τ, i.e. if they are τ-neighbouring. Moreover, setting $\ell = 1$ and $\tau = 1$ we recover the definition of quantum differential privacy based on convertibility by local measurements, used in [23].

This notion is particularly suitable to handle local measurements, i.e. measurements expressible as sums of local terms, as we show in Section 6. We remark that local measurements are of particular interest since they can be considered practically feasible measurements for extracting classical information from quantum data (or quantum systems). They also play a major role in variational learning algorithms as they are provably resilient to barren plateaus [40].

On the other hand, several encodings widely used in quantum machine learning are (Ξ, τ)-neighbouring-preserving, for appropriate choices of (Ξ, τ). We include upper bounds for $\max_{\mathcal{I} \in \Xi} |\mathcal{I}|$ and τ in Table (1). We delay to Appendix C the definition of the various encodings and the proof of upper bounds.

We also show that the notion of (Ξ, τ)-neighbouring states degrades gently under quantum postprocessing, assuming that the post-processing channel has a bounded light-cone.

Proposition 4.1 (Robustness to quantum post-processing)**.** *Let ρ and σ be two (Ξ, τ)-neighbouring states and consider a channel Φ with light-cone bounded by K. Then $\Phi(\rho)$ and $\Phi(\sigma)$ are (Ξ', τ)-neighbouring, where*

$$\max_{\mathcal{I} \in \Xi'} |\mathcal{I}| \leqslant K \max_{\mathcal{I} \in \Xi} |\mathcal{I}|.$$

Table 1: As we discuss in details in Section C, the encodings above are (Ξ, τ)-neighbouring-preserving for appropriate Ξ and τ depending on the encodings. We assume that the initial vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ are neighbouring if $\|\boldsymbol{x} - \boldsymbol{x}'\|_0 \leqslant \gamma_0$, $\|\boldsymbol{x} - \boldsymbol{x}'\|_1 \leqslant \gamma_1$ and $\|\boldsymbol{x} - \boldsymbol{x}'\|_2 \leqslant \gamma_2$. We also assumed that the Hamiltonian encoding is implemented by a $1D$ circuit of depth at most L. We refer to Section C for more details on the noise models.

ENCODING $\rho(\cdot)$	$\max_{\mathcal{I}\in\Xi} \lvert\mathcal{I}\rvert$	τ
AMPLITUDE ENCODING	n	γ_2
ROTATION ENCODING	γ_0	1
COHERENT STATE ENCODING	γ_0	$\sqrt{1 - e^{-\gamma_2^2}}$
1D-HAMILTONIAN ENCODING	$2L\gamma_0$	$O(1)\gamma_1$
1D-HAMILTONIAN ENCODING (LOW NOISE)	$2L\gamma_0$	$O(1)\sqrt{n}\exp(-L)$
1D-HAMILTONIAN ENCODING (HIGH NOISE)	$2L\gamma_0$	$O(1)\exp(-L)\gamma_1$

Proof. The proposition follows from the fact that the trace distance is non-increasing and from the definition of light-cone provided in Section 2. We have

$$\frac{1}{2}\|\Phi(\rho) - \Phi(\sigma)\|_1 \leqslant \frac{1}{2}\|\rho - \sigma\|_1 \leqslant \tau$$

Moreover,

$$\mathrm{Tr}_{\mathcal{J}}\rho = \mathrm{Tr}_{\mathcal{J}}\sigma$$

for $\mathcal{J} \in \Xi$. Since the channel Φ has bounded light-cone K, there exists $\mathcal{J}' \subseteq [n]$

$$\mathrm{Tr}_{\mathcal{J}'}\rho = \mathrm{Tr}_{\mathcal{J}'}\sigma$$

where $|\mathcal{J}'| \leqslant K|\mathcal{J}|$. This implies the desired result. □

We conclude this section by observing that our definition can be easily related to the quantum Wasserstein distance of order 1. Combining Lemma A.2 and Eq. (24), we obtain

$$\rho \overset{(\Xi,\tau)}{\sim} \sigma \implies W_1(\rho, \sigma) \leqslant \min\left\{\max_{\mathcal{I}\in\Xi} |\mathcal{I}|\frac{3}{2}\tau, n\tau\right\}. \tag{2}$$

It's natural to ask whether it would be convenient to define neighbouring quantum states in terms of the W_1 distance. The answer to this question is twofold. On the one hand, when we dispose of a single copy of the input state, the W_1 distance leads to a suboptimal tradeoff between privacy and accuracy, as we show in Theorem 7.2. On the other hand, when we dispose of multiple copies of the input state, neighbouring quantum states can be suitably defined in terms of the W_1 distance. We will discuss this alternative setting in Section 8.

5 Improved privacy for states with bounded trace distance

Before dealing with the general case of (Ξ, τ)-neighbouring states, we provide several new results for τ-neighbouring states, i.e. states with trace distance bounded by τ. This corresponds

to the definition previously explored in [22, 24]. In particular, we provide tighter guarantees for two private mechanisms, namely a generalised noisy channel and the addition of classical noise on the output of a quantum measurement. Following the convention used in [24], we state the results of this section using the quantum hockey-stick divergence.

Let $\mathcal{M}(\cdot)$ an arbitrary channel and let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}(\cdot)$. We briefly discuss how several noisy channels can be recovered as special cases of $\mathcal{N}_p(\cdot)$. For $\mathcal{M}(\cdot)$ equal to the identity channel $\mathrm{Id}(\cdot)$, $\mathcal{N}_p$ is the depolarising channel. For $n = 1$ we can also recover the single qubit Pauli channel $\mathcal{P}$ as a special case. Following [41], the action of $\mathcal{P}$ on a local Pauli operator $\sigma \in \{X, Y, Z\}$ can be expressed as

$$\mathcal{P}(\sigma) = q_\sigma \sigma,$$

where $-1 < q_X, q_Y, q_Z < 1$. It's customary to characterize the noise strength with a single parameter $q = \sqrt{\max\{|q_X|, |q_Y|, |q_Z|\}}$. Then for a single qubit state $\rho = \frac{1}{2}(\mathbb{1} + r_X X + r_Y Y + R_Z Z)$ we have

$$\begin{aligned}\mathcal{P}(\rho) &= \frac{1}{2}(\mathbb{1} + q_X r_X X + q_Y r_Y Y + q_Z r_Z Z) \\ &= (1-q^2)\frac{\mathbb{1}}{2} + q^2 \times \frac{1}{2}\left(\mathbb{1} + \frac{q_X}{q^2} r_X X + \frac{q_Y}{q^2} r_Y Y + \frac{q_Z}{q^2} r_Z Z\right) \\ &:= (1-q^2)\frac{\mathbb{1}}{2} + q^2 \mathcal{M}'(\rho),\end{aligned}$$

where we defined $\mathcal{M}'(\rho) := \frac{1}{2}\left(\mathbb{1} + \frac{q_X}{q^2} r_X X + \frac{q_Y}{q^2} r_Y Y + \frac{q_Z}{q^2} r_Z Z\right)$. We proceed by analysing the privacy guarantees of the channel $\mathcal{N}_p$.

Lemma 5.1. *Let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}(\cdot)$ a channel. For $0 \leqslant p \leqslant 1$ and $\gamma \geqslant 1$ we have*

$$E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) \leqslant (1-p)(1-\beta)E_\gamma(\rho\|\mathbb{1}/2^n) + (1-p)\beta E_\gamma(\rho\|\sigma),$$

where $\gamma' = 1 + (1-p)(\gamma - 1)$ and $\beta = \gamma'/\gamma$.

Proof. The result follows from Lemma B.2 by plugging $\rho_0 = \mathbb{1}/2^n$, $\rho_1 = \rho$ and $\rho_2 = \sigma$. □

Recall that from Lemma IV.1 in [24] we have that for the depolarising noise (hence for $\mathcal{M} = \mathrm{Id}$) and for any $\gamma \geqslant 1$,

$$E_\gamma(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) \leqslant \max\left\{0, (1-\gamma)\frac{p}{2^n} + (1-p)E_\gamma(\rho\|\sigma)\right\}.$$

In the following theorem, we extend this previous bound to an arbitrary channel $\mathcal{M}$ and we combine it with *Lemma* 5.1.

Theorem 5.1. *Let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}(\cdot)$ a channel. For $0 \leqslant p \leqslant 1$ and $\gamma' \geqslant 1$ we have*

$$\begin{aligned}&E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) \leqslant \\ &\min\left\{(1-p)(1-\beta)E_\gamma(\rho\|\mathbb{1}/2^n) + (1-p)\beta E_\gamma(\rho\|\sigma), \max\left\{0, (1-\gamma')\frac{p}{2^n} + (1-p)E_{\gamma'}(\rho\|\sigma)\right\}\right\}.\end{aligned}$$

where $\gamma = 1 + (\gamma' - 1)/(1-p)$ and $\beta = \gamma'/\gamma$.

Proof. Lemma 5.1 implies that

$$E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) \leqslant (1-p)(1-\beta)E_\gamma(\rho\|\mathbb{1}/2^n) + (1-p)\beta E_\gamma(\rho\|\sigma),$$

Then it remains to show that

$$E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) \leqslant \max\left\{0, (1-\gamma')\frac{p}{2^n} + (1-p)E_{\gamma'}(\rho\|\sigma)\right\}.$$

The proof closely follows the one of Lemma IV.1 and Lemma IV.4 in [24]. We have

$$\begin{aligned}
&E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) \\
&= \mathrm{Tr}((1-\gamma')p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}((\rho-\gamma'\sigma)))^+ \\
&= \mathrm{Tr}P^+((1-\gamma')p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}((\rho-\gamma'\sigma))),
\end{aligned}$$

where P^+ is the projector onto the positive subspace of $((1-\gamma')p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}((\rho-\gamma'\sigma)))$. Observe that

$$E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) > 0 \quad \Rightarrow \quad \mathrm{Tr}P^+ \geqslant 1.$$

Considering this case we get

$$\begin{aligned}
&E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) \\
&= (1-\gamma')\frac{p}{2^n}\mathrm{Tr}P^+ + (1-p)(\mathrm{Tr}P^+(\mathcal{M}(\rho-\gamma'\sigma))) \\
&\leqslant (1-\gamma')\frac{p}{2^n} + (1-p)E_{\gamma'}(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \\
&\leqslant (1-\gamma')\frac{p}{2^n} + (1-p)E_{\gamma'}(\rho\|\sigma) \\
&\leqslant (1-\gamma')\frac{p}{2^n} + (1-p).
\end{aligned}$$

Note that for sufficiently large γ' the upper bound could become negative, but one can easily check that in this case $E_{\gamma'}(\mathcal{N}_p(\rho)\|\mathcal{N}_p(\sigma)) = 0$ implying that we are in the other case.

□

For single-qubit product channels, we give the following bound:

Theorem 5.2. *Let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2} + (1-p)\mathcal{M}(\cdot)$ a single-qubit channel. For $0 \leqslant p \leqslant 1$ and $\gamma' \geqslant 1$ we have*

$$E_{\gamma'}(\mathcal{N}_p^{\otimes k}(\rho)\|\mathcal{N}_p^{\otimes k}(\sigma)) \leqslant$$
$$\min\left\{(1-p^k)(1-\beta)E_\gamma(\rho\|\mathbb{1}/2^k) + (1-p^k)\beta E_\gamma(\rho\|\sigma), \max\left\{0, (1-\gamma')\frac{p^k}{2^k} + (1-p^k)E_{\gamma'}(\rho\|\sigma)\right\}\right\}.$$

where $\gamma = 1 + (\gamma'-1)/(1-p)$ and $\beta = \gamma'/\gamma$.

Proof. It suffices to note that $\mathcal{N}_p^{\otimes k}$ can be rearranged as:

$$\mathcal{N}_p^{\otimes k}(\cdot) = p^k\frac{\mathbb{1}}{2^k} + (1-p^k)\mathcal{M}'(\cdot),$$

where $\mathcal{M}'$ is a quantum channel. Then the result follows from Theorem 5.1. □

These first two technical results show that several quantum noisy channels contract the quantum hockey-stick divergence. This can be used to prove that those channels ensure quantum differential privacy for τ-neighbouring states. In particular, we derive the following corollaries, that improve Lemma IV.2 and Lemma IV.5 in [24].

Corollary 5.1. *Let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}(\cdot)$ a channel. $\mathcal{N}_p$ is (ε, δ)-DP with respect to τ-neighbouring states with*

$$\delta \leqslant \max\left\{0, (1-e^{\varepsilon})\frac{p}{2^n} + (1-p)\tau\right\}. \tag{3}$$

Let $\gamma = 1 + (e^{\varepsilon}-1)/(1-p)$ and $\beta = e^{\varepsilon}/\gamma$. Under the additional assumption that the input state ρ satisfies $E_\gamma(\rho\|\frac{\rho}{\mathbb{1}/2^n}) \leqslant \eta$, we also have

$$\delta \leqslant (1-p)(1-\beta)\eta + (1-p)\beta\tau. \tag{4}$$

It's not straightforward whether Eq. (4) provides any advantage over Eq. (3). Thus, in Fig. 2 we plot both bounds of δ as a function of ε, for a specific set of parameters, and we observe that no bound is always tighter, and thus the choice of the bound will depend on the value of ε. An upper bound of δ as a function of ε is also referred to as *privacy profile*, a concept introduced in [42].

Corollary 5.2. *Let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2} + (1-p)\mathcal{M}(\cdot)$ single-qubit a channel. $\mathcal{N}_p^{\otimes k}$ is (ε, δ)-DP with respect to τ-neighbouring states with*

$$\delta \leqslant \max\left\{0, (1-e^{\varepsilon})\frac{p^k}{2^k} + (1-p^k)\tau\right\}.$$

Let $\gamma = 1 + (e^{\varepsilon}-1)/(1-p^k)$ and $\beta = e^{\varepsilon}/\gamma$. Under the additional assumption that the input state ρ satisfies $E_\gamma(\rho\|\frac{\rho}{\mathbb{1}/2^k}) \leqslant \eta$, we also have

$$\delta \leqslant (1-p^k)(1-\beta)\eta + (1-p^k)\beta\tau.$$

Bounding privacy with the purity

Our results improve the prior bounds under the additional assumption that the divergence $E_\gamma(\rho\|\mathbb{1}/2^n)$ is relatively small. The value of $E_\gamma(\rho\|\mathbb{1}/2^n)$ can be thought as a "distance" between the state ρ and the maximally mixed state, thus small values of $E_\gamma(\rho\|\mathbb{1}/2^n)$ are associated to high levels of noise. Hence, we can connect it to the purity $\mathrm{Tr}[\rho^2]$ of the state ρ, or the related D_2 divergence. By definition, we have

$$\mathrm{Tr}[\rho^2] = 2^{-n+D_2(\rho\|\mathbb{1}/2^n)}.$$

The hockey stick divergence and the Rényi divergence satisfy the following relationship ([38], Proposition 6.22)

$$E_{e^{\varepsilon}}(\rho\|\mathbb{1}/2^n) \leqslant \delta, \tag{5}$$

where $\varepsilon = D_2(\rho\|\mathbb{1}/2^n) - \log(1-\sqrt{1-\delta^2}) \leqslant D_2(\rho\|\mathbb{1}/2^n) + \log(2/\delta^2)$. We also note that two states with low purity are also close in hockey-stick divergence:

$$E_\gamma(\rho\|\sigma) \leqslant E_1(\rho\|\sigma) \leqslant E_1(\rho\|\mathbb{1}/2^n) + E_1(\sigma\|\mathbb{1}/2^n).$$

And then $E_1(\rho\|\mathbb{1}/2^n) = \frac{1}{2}\|\rho - \mathbb{1}/2^n\|_1$ can be bounded either with the quantum Bretagnolle Huber inequality (Lemma B.1) or the Pinsker's inequality. Now, we show how Corollary 5.1 and Corollary 5.2 can be rephrased in terms of the purity of the input state.

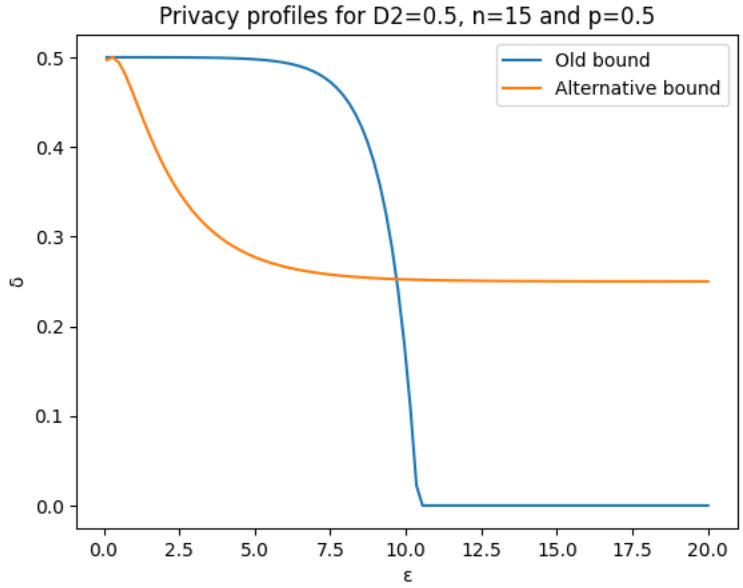

Figure 2: In this figure we compare the former upper bound from [24] (Eq. (3)) with the novel upper bound provided in this section (Eq. (4)). We emphasise that each bound outperforms the other for some values of ε. We assumed that the input state satisfy $D_2(\rho\|\mathbb{1}/2^n) \leqslant 0.5, n = 15$ and $p = 0.5$. The upper bound on τ is derived from $\|\rho - \sigma\|_1 \leqslant \|\rho - \mathbb{1}/2^n\|_1 + \|\rho - \mathbb{1}/2^n\|_1 \leqslant 2\sqrt{2D_2(\rho\|\mathbb{1}/2^n)}$, i.e. combining the triangle inequality and the Pinsker's inequality.

Corollary 5.3. *Let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}(\cdot)$ a channel that acts on state ρ with bounded purity $\mathrm{Tr}[\rho^2] \leqslant \zeta < 1$. Let $\gamma = 1 + (e^\varepsilon - 1)/(1-p)$, $\beta = e^\varepsilon/\gamma$ and $\eta = \sqrt{2n\zeta^{\frac{1}{\log 2}}\gamma^{-1}}$. Then $\mathcal{N}_p$ is (ε, δ)-DP with respect to τ-neighbouring states with*

$$\delta \leqslant (1-p)(1-\beta)\eta + (1-p)\beta\tau.$$

Proof. The proof follows by plugging the relation between purity and hockey-stick divergence into Corollary 5.1. We have

$$D_2(\rho\|\mathbb{1}/2^n) \leqslant \log_2(\zeta) + n,$$

and hence, by Eq. (5),

$$E_\gamma(\rho\|\mathbb{1}/2^n) \leqslant \sqrt{2n\zeta^{\frac{1}{\log 2}}\gamma^{-1}} := \eta,$$

which satisfies the hypothesis of Corollary 5.1. □

Proceeding in a similar way can also prove a purity-based bound for local channels.

Corollary 5.4. *Let $\mathcal{N}_p(\cdot) = p\frac{\mathbb{1}}{2^n} + (1-p)\mathcal{M}(\cdot)$ a single-qubit channel and assume that $\mathcal{N}_p^{\otimes k}$ acts on state ρ with bounded purity $\mathrm{Tr}[\rho^2] \leqslant \zeta < 1$. Let $\gamma = 1 + (e^\varepsilon - 1)/(1-p^k)$, $\beta = e^\varepsilon/\gamma$ and $\eta = \sqrt{2n\zeta^{\frac{1}{\log 2}}\gamma^{-1}}$. Then $\mathcal{N}_p$ is (ε, δ)-DP with respect to τ-neighbouring states with*

$$\delta \leqslant (1-p^k)(1-\beta)\eta + (1-p^k)\beta\tau.$$

5.1 Privacy via classical post-processing

Now, we show that the output of a quantum measurement can be privatised by adding classical noise. This result is particularly interesting since firstly, it provides a practical approach for using the existing tools and techniques from classical differential privacy to privatize the outputs of our quantum systems and algorithms, and secondly allows us to be able to combine classical noise with the output distributions resulting from a quantum measurement. In particular, we can account for quantum and classical noise in the analysis by noting that quantum noisy channels contract the trace distance between any two quantum states, and, moreover, the privacy guarantees obtained by adding classical noise are inversely proportional to the trace distance between two neighbouring states.

Lemma 5.2. *Let ρ, σ such that $\frac{1}{2}\|\rho - \sigma\|_1 \leqslant \tau$. Let M be a POVM measurement and Λ a classical channel such that $\forall x, x' \in \mathsf{range}(M) : E_{e^\varepsilon}(\Lambda(x)\|\Lambda(x')) \leqslant \delta$. Then we have that*

$$E_{e^{\varepsilon'}}(\Lambda(M(\rho))\|\Lambda(M(\sigma))) \leqslant \tau\delta,$$

where $\varepsilon' = \log(1 + \tau(e^\varepsilon - 1))$, which for small ε gives $\varepsilon' \simeq \tau\varepsilon$.

Proof. Let $\nu := M(\rho)$ and $\nu' := M(\sigma)$. We have that

$$d_{TV}(\nu, \nu') := \eta \leqslant \tau,$$

which follows from the data processing inequality. Moreover, there always exists some distributions ν_0, ν_1, ν_1' such that

$$\nu = (1-\eta)\nu_0 + \eta\nu_1, \qquad \nu' = (1-\eta)\nu_0 + \eta\nu_1'.$$

The above identities are discussed in detail in ([42], Section 3). We also have,

$$\max\{E_{e^\varepsilon}(\Lambda(\nu_1)\|\Lambda(\nu_0)), E_{e^\varepsilon}(\Lambda(\nu_1)\|\Lambda(\nu_1'))\} \leqslant \delta$$

This follows by noting that ν_0, ν_1, ν_1' are supported in $\mathsf{range}(M)$ and applying the (standard) joint-convexity of the hockey-stick divergence. By advanced joint convexity (Lemma B.2), we have that for all states ρ_0, ρ_1, ρ_2 and $\gamma' = 1 + (1-p)(\gamma - 1)$,

$$E_{\gamma'}(p\rho_0 + (1-p)\rho_1\|p\rho_0 + (1-p)\rho_2) \leqslant (1-p)(1-\beta)E_\gamma(\rho_1\|\rho_0) + (1-p)\beta E_\gamma(\rho_1\|\rho_2),$$

Then,

$$E_{e^{\varepsilon'}}(\Lambda(M(\rho))\|\Lambda(M(\sigma))) \leqslant \tau\delta.$$

□

Lemma 5.2 is stated in terms of a general classical noisy channel. In the following theorem we consider the special cases of the Laplace and Gaussian mechanisms, two noisy channels widely used in many differentially private classical algorithms and defined in Section 2.2.

Theorem 5.3. *Let M a measurement with range $[a, a + \Delta]$ for $a \in \mathbb{R}$.*

- *(Laplace mechanism) Let $\Lambda_{\mathcal{L},b}$ the Laplace noise of scale b. Then $\Lambda_{\mathcal{L},b}(M(\cdot))$ is ε'-DP with respect to τ-neighbouring states, where*

$$\varepsilon' = \log(1 + \tau(e^{\Delta/b} - 1)).$$

- *(Gaussian mechanism) Let $\Lambda_{\mathcal{G},\sigma}$ the Gaussian noise of variance $\sigma^2 \geqslant 2\ln(1.25/\delta)\Delta^2/\varepsilon^2$. Then $\Lambda_{\mathcal{G},\sigma}(M(\cdot))$ is (ε', δ')-DP with respect to τ-neighbouring states, where*

$$\varepsilon' = \log(1 + \tau(e^{\varepsilon} - 1)) \quad \textit{and} \quad \delta' = \tau\delta.$$

Proof. The theorem follows by replacing the channel Λ in Lemma 5.2 with the Laplace and Gaussian noise, respectively. □

5.2 Implications for quantum-inspired sampling

As the trace distance generalizes the total variation distance, the range of applicability of Theorem 5.3 includes also classical algorithms. In particular, we show here an application for private quantum-inspired sampling. In quantum-inspired algorithms [43, 44, 45, 46, 47], a classical vector $u \in \mathbb{C}^N$ is accessed through quantum-inspired sampling: i.e. an entry u_i is sampled with probability proportional to $|u_i|^2$. This is equivalent to encoding u into the state

$$|u\rangle = \frac{1}{\|u\|_2} \sum_{i=1}^{N} u_i |i\rangle,$$

and performing a computational-basis measurement. Let p_u be the distribution induced by such measurements. Say that $u \sim u'$ if u and u' differ in only one entry. In particular, let $u_i = u'_i$ for all $i \neq j$.

$$\begin{aligned} |\|u\|_2^2 - \|u'\|_2^2| &= \left| \sum_i |u_i|^2 - \sum_i |u'_i|^2 \right| \\ \leqslant \left| \sum_{i\neq j} |u_i|^2 - \sum_{i\neq j} |u'_i|^2 + |u_j|^2 - |u'_j|^2 \right| &\leqslant \max\{|u_j|^2, |u'_j|^2\} \end{aligned}$$

It's easy to see that p_u and $p_{u'}$ are close in total variation distance.

$$\begin{aligned} |p_u - p_{u'}|_{\mathrm{tv}} &= \frac{1}{2} \sum_i \left| \frac{|u_i|^2}{\|u\|_2^2} - \frac{|u'_i|^2}{\|u'\|_2^2} \right| \\ &\leqslant \frac{1}{2} \left(\sum_{i\neq j} |u_i|^2 \left| \frac{1}{\|u\|_2^2} - \frac{1}{\|u'\|_2^2} \right| + \left| \frac{|u_j|^2}{\|u\|_2^2} - \frac{|u'_j|^2}{\|u'\|_2^2} \right| \right) \\ &\leqslant \frac{1}{2} \left(\min\{\|u\|_2^2, \|u'\|_2^2\} \frac{\max\{|u_j|^2, |u'_j|^2\}}{\|u\|_2^2 \|u'\|_2^2} + \frac{|u_j|^2}{\|u\|_2^2} + \frac{|u'_j|^2}{\|u'\|_2^2} \right) \\ &\leqslant \frac{3}{2} \max\left\{ \frac{|u_j|^2}{\|u\|_2^2}, \frac{|u'_j|^2}{\|u'\|_2^2} \right\} := \alpha. \end{aligned}$$

Then, by subadditivity of the total variation distance,

$$|p_u^{\otimes m} - p_{u'}^{\otimes m}|_{\mathrm{tv}} \leqslant m\alpha.$$

We will show the intuitive fact that quantum-inspired subsampling amplifies DP. First, we can consider the encoding $u \mapsto p_u^{\otimes m}$ and derive the following special case of Theorem 5.3.

Corollary 5.5. *Let u, u' be neighbouring if they differ in at most one entry. Consider the oracle O_u that returns a u_i with probability $\frac{|u_i|^2}{\|u\|_2^2}$. For $a \in \mathbb{R}$ and $\Delta \geqslant 0$, let $\mathcal{S}$ a randomised algorithm with range $[a, a+\Delta]$ that makes m queries to O_u and assume that $\frac{3}{2}\frac{|u_j|^2}{\|u\|_2^2} \leqslant \alpha$.*

- *(Laplace mechanism) Let $\Lambda_{\mathcal{L},b}$ the Laplace noise of scale b. Then $\Lambda_{\mathcal{L},b}(\mathcal{S}(\cdot))$ is ε'-DP, where*

$$\varepsilon' = \log(1 + \alpha m(e^{\Delta/b} - 1)).$$

- *(Gaussian mechanism) Let $\Lambda_{\mathcal{G},\sigma}$ the Gaussian noise of variance $\sigma^2 \geqslant 2\ln(1.25/\delta)\Delta^2/\varepsilon^2$. Then $\Lambda_{\mathcal{G},\sigma}(\mathcal{S}(\cdot))$ is (ε', δ')-DP, where*

$$\varepsilon' = \log(1 + \alpha m(e^{\varepsilon} - 1)) \quad \text{and} \quad \delta' = \alpha m \delta.$$

The approach described above is tailored to noise-adding mechanisms. In Appendix D we provide a more general result that applies to any private mechanism and it builds upon prior work on privacy amplification by subsampling [42, 48].

6 Differential privacy for (Ξ, τ)-neighbouring states

While in Section 5 we provided tighter bounds for quantum differential privacy with respect to states with bounded trace distance, here we add two additional ingredients: the locality of the measurements and the generalised neighbouring relationship defined in Section 4. Under these stronger assumptions, we can improve the privacy guarantees of local noisy channels and classical post-processing. First, we need to introduce the following quantity.

Definition 6.1 (Worst-case quantum sensitivity)**.** *Let O be an observable expressed as a weighted sum of Pauli operators, $O = \sum_{P \in \{X,Y,Z,\mathbb{1}\}^n} c_P P$. Let $\mathcal{I} \subseteq [n]$ and consider the subset $\mathcal{S}_{\mathcal{I}}$ of all the Pauli strings that act non trivially on $\mathcal{I}$. The worst-case quantum sensitivity of O with respect to $\mathcal{I}$ is defined as*

$$\Delta(O; \mathcal{I}) := 2 \sum_{P \in \mathcal{S}_{\mathcal{I}}} |c_P|.$$

Let $\Xi \subseteq P([n])$, i.e. Ξ is a collection of subsets of $[n]$. The worst-case quantum sensitivity of O with respect to Ξ is defined as

$$\Delta_{\Xi}(O) := \max_{\mathcal{I} \in \Xi} \Delta(O; \mathcal{I}).$$

We will omit the index Ξ and simply write $\Delta(O)$ when there is no ambiguity.

So, if $O = \sum_{i=1}^n Z_i$ and $\Xi = \{\{1\}, \{2\}, \ldots, \{n\}\}$, the worst-case quantum sensitivity equals $\Delta(O) = 2$. This is consistent with the fact that, if ρ and σ satisfy $\mathrm{Tr}_j \rho = \mathrm{Tr}_j \sigma$, then all the terms but Z_j induce the same distributions when measured on either ρ or σ. Moreover, the outcome of term Z_j will be either 1 or -1, then it belongs to an interval of length 2. We can also consider the more general case where $O_\ell = \sum_{i=1}^n \bigotimes_{j=i}^{i+\ell-1} Z_j$ and $\Xi = \{\{i, i+1, \ldots, i+k\} | \text{ for } i = 1, 2, \ldots, n-k\}$. It's easy to see that $\Delta(O_\ell) = 2k + 4\ell - 4$.

We can now state the first result of this section, concerning a class of local noisy channels, which includes the local Pauli noise.

Theorem 6.1 (Generalised private measurement via local noisy channels). *Let $O = \sum_P c_P P$ be an observable consisting of a weighted sum of commuting Pauli operators. Let $\mathcal{M}$ an arbitrary single qubit channel and let $\mathcal{N}(\cdot) = p\mathbb{1}/2 + (1-p)\mathcal{M}(\cdot)$. Let $k = \max_{\mathcal{I}\in\Xi} |\mathcal{I}|$. Then $\mathcal{O}\circ\mathcal{N}^{\otimes n}$ satisfies (ε, δ_k)-DP with respect to (Ξ, τ)-neighbouring states, where*

$$\delta_k \leqslant \max\left\{0, (1-e^{\varepsilon})\frac{p^k}{2^k} + (1-p^k)\tau\right\}.$$

Let $\gamma = 1 + (e^{\varepsilon}-1)/(1-p)$ and $\beta = e^{\varepsilon}/\gamma$. Under the additional assumption the the input state ρ satisfies $E_\gamma(\rho\|\mathbb{1}/2^n) \leqslant \eta$, the following inequality also holds

$$\delta_k \leqslant (1-p^k)(1-\beta)\eta + (1-p^k)\beta\tau.$$

Proof. Since $\rho \overset{(\Xi,\tau)}{\sim} \sigma$, there exists $\mathcal{I} \in \Xi$ such that

$$\mathrm{Tr}_{\mathcal{I}}\rho = \mathrm{Tr}_{\mathcal{I}}\sigma \quad \text{and} \quad |\mathcal{I}| \leqslant k. \tag{6}$$

We also have

$$\mathcal{N}^{\otimes n}(\rho) = p^{|\mathcal{I}|}\left(\mathrm{Tr}_{\mathcal{I}}\mathcal{M}(\rho)\otimes\frac{\mathbb{1}}{2^{|\mathcal{I}|}}\right) + (1-p^{|\mathcal{I}|})\mathcal{M}(\rho).$$

The measurement O can be implemented by measuring each qubit in a different Pauli basis and then performing classical postprocessing. As quantum differential privacy is robust to postprocessing, we only need to prove that Pauli measurements preserve (ε, δ_k)-DP. We can assume without loss of generality that the qubits in the subsystem $\mathcal{I}^c$ are measured first, since we assumed that O is a weighted sum of commuting Pauli operators, and hence the measurement order doesn't alter the overall statistics. Assume that measuring the subsystem $\mathcal{I}^c$ produces the outcome $\boldsymbol{y} \in \{\pm 1\}^{n-|\mathcal{I}|}$. Eq. (6) implies that

$$p(\boldsymbol{y}) := \Pr[\boldsymbol{y} \text{ is obtained on input } \rho] = \Pr[\boldsymbol{y} \text{ is obtained on input } \sigma].$$

Denote by $\rho_{\boldsymbol{y}}$ the post-measurement state produced by measuring the system $\mathcal{I}^c$ and obtaining outcome $\boldsymbol{y}$. Let $\mathcal{T}_{\boldsymbol{y}}$ be the quantum channel mapping ρ to $\rho_{\boldsymbol{y}}$.

$$\begin{aligned}
&\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{N}^{\otimes n}(\rho)) \\
&= p^{|\mathcal{I}|}\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}\left(\mathrm{Tr}_{\mathcal{I}}\mathcal{M}(\rho)\otimes\frac{\mathbb{1}}{2^{|\mathcal{I}|}}\right) + (1-p^{|\mathcal{I}|})\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{M}(\rho)) \\
&= p^{|\mathcal{I}|}\frac{\mathbb{1}}{2^{|\mathcal{I}|}} + (1-p^{|\mathcal{I}|})\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{M}(\rho)) \\
&:= p^{|\mathcal{I}|}\frac{\mathbb{1}}{2^{|\mathcal{I}|}} + (1-p^{|\mathcal{I}|})\mathcal{M}'(\rho).
\end{aligned}$$

where the second equality follows from $\mathcal{T}_{\boldsymbol{y}}\left(\mathrm{Tr}_{\mathcal{I}}\mathcal{M}(\rho)\otimes\frac{\mathbb{1}}{2^{|\mathcal{I}|}}\right) = \mathrm{Tr}_{\mathcal{I}}\mathcal{T}(\mathcal{M}(\rho))\otimes\frac{\mathbb{1}}{2^{|\mathcal{I}|}}$ and we defined $\mathcal{M}' := \mathrm{Tr}_{\mathcal{I}^c}\circ\mathcal{T}_{\boldsymbol{y}}\circ\mathcal{M}$. By Corollary 5.2,

$$E_{e^{\varepsilon}}(\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{N}^{\otimes n}(\rho))\|\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{N}^{\otimes n}(\sigma))) \leqslant \delta_k. \tag{7}$$

In order to prove that measuring O on $\mathcal{N}^{\otimes n}(\rho)$ preserves (ε, δ_k)-DP, it's sufficient consider the outcome $\boldsymbol{y}$ and the partial post-measurement state $\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{N}^{\otimes n}(\rho))$. Thus we need to ensure that

$$E_{e^{\varepsilon}}\left(\sum_{\boldsymbol{y}} p_{\boldsymbol{y}}(\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{N}^{\otimes n}(\rho))\otimes|\boldsymbol{y}\rangle\langle\boldsymbol{y}|)\middle\|\sum_{\boldsymbol{y}} p_{\boldsymbol{y}}(\mathrm{Tr}_{\mathcal{I}^c}\mathcal{T}_{\boldsymbol{y}}(\mathcal{N}^{\otimes n}(\sigma))\otimes|\boldsymbol{y}\rangle\langle\boldsymbol{y}|)\right) \leqslant \delta(\varepsilon, k)$$

We also have, for all $\gamma \geqslant 1$,

$$\begin{aligned} E_\gamma \left(\sum_y p_y (\mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\rho)) \otimes |y\rangle\langle y|) \middle\| \sum_y p_y (\mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\sigma)) \otimes |y\rangle\langle y|) \right) \\ \leqslant \sum_y p_y E_\gamma \left(\mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\rho)) \otimes |y\rangle\langle y| \, \| \mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\sigma)) \otimes |y\rangle\langle y| \right) \\ \leqslant \sum_y p_y E_\gamma \left(\mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\rho)) \| \mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\sigma)) \right) \\ \leqslant \max_y E_\gamma \left(\mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\rho)) \| \mathrm{Tr}_{\mathcal{I}^c} \mathcal{T}_y(\mathcal{N}^{\otimes n}(\sigma)) \right), \end{aligned} \tag{8}$$

where the second line follows from the convexity of the hockey-stick divergence Eq. (22) and the third line follows from the stability of the hockey-stick divergence Eq. (23). Combining Eq. (7) with Eq. (8) gives the desired result:

$$E_{e^\varepsilon}(\mathcal{O}(\mathcal{N}^{\otimes n}(\rho)) \| \mathcal{O}(\mathcal{N}^{\otimes n}(\sigma))) \leqslant \delta_k.$$

□

We emphasise that the number of qubits n appearing in the guarantees of Corollary 5.2 is now replaced by $k = \max_{\mathcal{I} \in \Xi} |\mathcal{I}|$. Thus if $k = \mathrm{polylog}(n)$, this new bound is exponentially tighter than the previous one. In a similar fashion, we can adapt Theorem 5.3 to the generalised neighbouring relationship.

Theorem 6.2 (Generalised private measurement via classical post-processing)**.** *Let ρ and σ two (Ξ, τ)-neighbouring quantum states, i.e. $\rho \overset{(\Xi,\tau)}{\sim} \sigma$. Let O be an observable, and denote $\mathcal{O}$ as a quantum-to-classical channel implementing a measurement of O.*

- *(Laplace mechanism) Let $\Lambda_{\mathcal{L},b}$ the Laplace noise of scale b. Then $\Lambda_{\mathcal{L},b}(\mathcal{O}(\cdot))$ is ε'-DP with respect to (Ξ, τ)-neighbouring states, where*

$$\varepsilon' = \log(1 + \tau(e^{\Delta(O)/b} - 1)).$$

- *(Gaussian mechanism) Let $\Lambda_{\mathcal{G},\sigma}$ the Gaussian noise of variance $\sigma^2 \geqslant 2\log(1.25/\delta)\Delta(O)^2/\varepsilon^2$. Then $\Lambda_{\mathcal{G},\sigma}(\mathcal{O}(\cdot))$ is (ε', δ')-DP with respect to (Ξ, τ)-neighbouring states, where*

$$\varepsilon' = \log(1 + \tau(e^\varepsilon - 1)) \quad \text{and} \quad \delta' = \tau\delta.$$

Proof. Proceeding as in the proof of Theorem 6.1, consider $\mathcal{I} \in \Xi$ such that $\mathrm{Tr}_{\mathcal{I}}\rho = \mathrm{Tr}_{\mathcal{I}}\sigma$ and let $\mathcal{S}_{\mathcal{I}}$ be the subset of all the Pauli strings that act non trivially on $\mathcal{I}$. Thus, we can decompose O as $O = O_1 + O_2$, where $O_1 = \sum_{P \notin \mathcal{S}_{\mathcal{I}}} c_P P$ and $O_2 = O - O_1 = \sum_{P \in \mathcal{S}_{\mathcal{I}}} c_P P$. Assume without loss of generality that O_1 is measured first. Since $\mathrm{Tr}_{\mathcal{I}}\rho = \mathrm{Tr}_{\mathcal{I}}\sigma$ and O_1 acts non trivially only on $\mathcal{I}^c = [n] \backslash \mathcal{I}$, then this measurement produces no loss of privacy, i.e.

$$\forall y : p(y) := \Pr_\rho[O_1 = y] = \Pr_\sigma[O_1 = y].$$

Observe that O_2 is a measurement whose output is comprised into $[-\Delta(O)/2, \Delta(O)/2]$. Moreover, let ρ_y be the post-measurement state obtained when O_1 returns outcome y. As the trace distance is non-increasing, we have,

$$\frac{1}{2}\|\rho_y - \sigma_y\| \leqslant \frac{1}{2}\|\rho - \sigma\| \leqslant \tau,$$

Conditioning on input y, the output of $O = O_1 + O_2$ lies in $[y - \Delta/2, y + \Delta/2]$. Then Theorem 5.3 yields

$$E_{e^{\varepsilon'}}\left(\sum_y p(y)\Lambda_{\mathcal{L},b}(\mathcal{O}(\rho_y))\middle\|\sum_y p(y)\Lambda_{\mathcal{L},b}(\mathcal{O}(\rho_y))\right) \leqslant \max_y E_{e^{\varepsilon'}}\left(\Lambda_{\mathcal{L},b}(\mathcal{O}(\rho_y))\|\Lambda_{\mathcal{L},b}(\mathcal{O}(\rho_y))\right) \leqslant 0.$$

for $\varepsilon' = \log(1 + \tau(e^{\Delta(O)/b} - 1))$. Similarly, replacing the Laplace noise with the Gaussian noise and applying again Theorem 5.3,

$$E_{e^{\varepsilon}}\left(\sum_y p(y)\Lambda_{\mathcal{G},\sigma}(\mathcal{O}(\rho_y))\middle\|\sum_y p(y)\Lambda_{\mathcal{G},\sigma}(\mathcal{O}(\rho_y))\right) \leqslant \max_y E_{e^{\varepsilon}}\left(\Lambda_{\mathcal{G},\sigma}(\mathcal{O}(\rho_y))\|\Lambda_{\mathcal{G},\sigma}(\mathcal{O}(\rho_y))\right) \leqslant \delta',$$

where $\sigma^2 \geqslant 2\log(1.25/\delta)\Delta(O)^2/\varepsilon^2$, $\varepsilon' = \log(1 + \tau(e^{\varepsilon} - 1))$ and $\delta' = \tau\delta$. □

We observe that similar results can be derived for multiple sources of noise, beyond the Laplace or the Gaussian channels, along the lines of Lemma 5.2. We leave it to the reader to extend Theorem 6.2 to alternative stochastic channels.

7 The cost of quantum differential privacy

Differential privacy, both in the classical and in the quantum setting, can be achieved by introducing noise into the computation, thus reducing the final accuracy. Intuitively, large values of ε can be attained with little loss in accuracy, while for $\varepsilon = 0$ the output is totally independent of the input. In particular, if an algorithm is ε-DP with respect to Hamming distance, we have that

$$\forall x, x' : D_\infty(\mathcal{A}(x)\|\mathcal{A}(x')) \leqslant \varepsilon n, \tag{9}$$

thus if $\varepsilon = O(1/n)$, any pair of inputs (not necessarily neighbouring) are mapped to outputs $O(1)$-close in max-divergence. This result follows from the fact that the max-relative entropy satisfies the triangle inequality (both in the classical and in the quantum cases), i.e. $\forall \rho_1, \rho_2, \sigma$: $D_\infty(\rho_1\|\rho_2) \leqslant D_\infty(\rho_1\|\sigma) + D_\infty(\sigma\|\rho_2)$. We can pick a sequence of $n+1$ inputs $x_0, x_1, \ldots, x_n$ such that $x = x_0$, $x' = x_n$ and $x_i \sim x_{i+1}$. Then iterating the triangle inequality yields Eq. (9). However, for most applications ε can be chosen as a constant independent of n, avoiding this undesired concentration of the output around a unique value.

A vast portion of the literature about differential privacy is devoted to optimising the tradeoff between the value of ε and the loss in utility. In this section we make a crucial observation: the privacy-utility tradeoff doesn't depend solely on the value of ε, but also on the notion of neighbouring inputs. Thus, the privacy-utility tradeoff is an important figure of merit for the comparison of different approaches to quantum differential privacy.

In particular, we argue that some prior definitions of neighbouring quantum states suffer from a poor tradeoff between privacy and accuracy, leading to a suboptimal scaling with respect to the number of qubits n. This is the case, for instance, if we require two neighbouring states to have bounded trace distance $\tau = \Theta(1)$. We also provide a similar result for the Wasserstein distance of order 1.

7.1 Concentration inequalities for private measurements

It's well known that noisy quantum algorithms suffer from severe limitations, that often hinder quantum advantage. Prior works [49, 50] showed that, if the noise exceeds a given threshold, the output of noisy devices is concentrated around the maximally mixed state, and then it can be efficiently approximated with a classical computer. Since quantum differential privacy involves the injection of noise, it's not surprising that similar concentration inequalities hold for quantum private algorithms. In the remainder of this section, we will show how this concentration affects the accuracy of private measurements. For the sake of simplicity, we will state our results in terms of simple, local observables such as $O = \sum_{i=1}^{n} Z_i$. Similar results can be obtained for any observable with bounded Lipschitz constant, as also discussed in [50], but our choice is sufficient to display the shortcomings of a poor choice of the neighbouring relationship. If we measure O on the maximally mixed state $\mathbb{1}/2^n$, the outcome satisfies a Gaussian concentration inequality [50]:

$$\Pr_{\mathbb{1}/2^n}(|O| \geqslant an) \leqslant Ke^{-a^2 n}, \tag{10}$$

for $K = 1$. So, if a state ρ satisfies $D_\infty(\rho\|\mathbb{1}/2^n) \leqslant \varepsilon$, the definition of the quantum max-relative entropy yields,

$$\Pr_{\rho}(|O| \geqslant an) \leqslant e^{\varepsilon} \Pr_{\mathbb{1}/2^n}(|O| \geqslant an) \leqslant K'e^{-a^2 n}, \tag{11}$$

where $K' = e^{\varepsilon}$. For the sake of simplicity, throughout this section, we consider the special case of *pure* differential privacy, i.e. $(\varepsilon, 0)$-DP, but our results can be suitably extended to the more general *approximate* differential privacy, i.e. (ε, δ)-DP, under the assumption that $\delta \ll 1$.

Consider a quantum channel $\mathcal{A}(\cdot)$ and assume for the sake of simplicity that $\mathcal{A}$ is unital, i.e. $\mathcal{A}(\mathbb{1}) = \mathbb{1}$. We show that different neighbouring relationships $\overset{Q}{\sim}$ have a disparate impact on the accuracy. The first result is devoted to states with bounded trace distances.

Theorem 7.1 (Concentration inequality for bounded trace distance)**.** *Consider the observable $O = \sum_{i=1}^{n} Z_i$ and let $\mathcal{A}$ be a unital quantum channel satisfying ε-DP with respect to τ-neighbouring states, i.e. $D_\infty(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leqslant \varepsilon$ if $\frac{1}{2}\|\rho - \sigma\|_1 \leqslant \tau$. Assume $\tau = \Theta(1)$. Then, for any input state ρ, the output $\mathcal{A}(\rho)$ satisfies the following concentration inequality:*

$$\Pr_{\mathcal{A}(\rho)}(|O| \geqslant an) \leqslant K'e^{-a^2 n},$$

where $K' = e^{O(\varepsilon)}$.

Proof. For two arbitrary quantum states, we have

$$\forall \rho, \sigma : D_\infty(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leqslant \varepsilon/\tau. \tag{12}$$

This can be seen by building the following chain :

$$\rho_i = \rho \max(0, 1 - i\tau) + \sigma \min(1, i\tau)$$

We note that $\frac{1}{2}\|\rho_i - \rho_{i+1}\|_1 \leqslant \tau$ which implies $D_\infty(\mathcal{A}(\rho_i)\|\mathcal{A}(\rho_{i+1})) \leqslant \varepsilon$. Then Eq. (12) can be deduced by iterating the triangle inequality. Combining it with Eq. (11), we obtain

$$\forall \rho : \Pr_{\mathcal{A}(\rho)}(|O| \geqslant an) \leqslant Ke^{-a^2 n},$$

where $K = e^{\varepsilon/\tau} = e^{O(\varepsilon)}$. □

To showcase the implications of the Theorem 7.1, we set $\tau = 0.1$ and we consider $\rho := |1^n\rangle\langle 1^n|$. We remark that ρ is an eigenvector of O, with eigenvalue n. However, instead of measuring O directly, we can post-process ρ with a ε-DP channel $\mathcal{A}$ as defined in the statement of the theorem. Set $\varepsilon = 1$. In order to achieve an error smaller than, say, $0.5n$, we need to ensure that the outcome is larger than $0.9n$. Then Theorem 7.1 implies that the error is larger than $0.5n$ with high probability:

$$\Pr_{\mathcal{A}(\rho)}(|n - O| \leqslant 0.5n) = \Pr_{\mathcal{A}(\rho)}(O \geqslant 0.5n) \leqslant \Pr_{\mathcal{A}(\rho)}(|O| \geqslant 0.5n) \leqslant e^{10-0.25n}$$

and hence setting $n = 100$ we obtain

$$\Pr_{\mathcal{A}(\rho)}(|n - O| \leqslant 0.5n) \leqslant 3 \times 10^{-7}.$$

Now, we provide a similar result for another neighbouring definition. In [28], the authors extend the Wasserstein distance of order 1 (or W_1 distance) to quantum states and suggest quantum differential privacy as a potential application of their work. Recall that the W_1 distance between the quantum states ρ and σ of $\mathcal{H}_n$ is defined as

$$W_1(\rho,\sigma) = \min\Bigg(\sum_{i=1}^{n} c_i : c_i \geqslant 0, \rho - \sigma = \sum_{i=1}^{n} c_i\left(\rho^{(i)} - \sigma^{(i)}\right), \\ \rho^{(i)}, \sigma^{(i)} \in \mathcal{S}_n, \mathrm{Tr}_i\rho^{(i)} = \mathrm{Tr}_i\sigma^{(i)}\Bigg).$$

The following theorem shows that the W_1 distance leads to the following undesired concentration inequality.

Theorem 7.2 (Concentration inequality for bounded W_1 distance)**.** *Consider the observable $O = \sum_{i=1}^{n} Z_i$ and let $\mathcal{A}$ be a unital quantum channel satisfying ε-DP with respect stated with W_1 distance bounded by 1, i.e. $D_\infty(\mathcal{A}(\rho_1)\|\mathcal{A}(\rho_2)) \leqslant \varepsilon$ if $W_1(\rho_1,\rho_2) \leqslant 1$. Then, for any input state ρ, the output $\mathcal{A}(\rho)$ satisfies the following concentration inequality:*

$$\Pr_{\mathcal{A}(\rho)}(|O| \geqslant an) \leqslant K' e^{-a^2 n},$$

where $K' = e^{\varepsilon}(n - e^{-\varepsilon}(n-1))$.

Proof. Quantum differential privacy with respect to bounded Wasserstein distance of order 1 can be expressed as:

$$W_1(\rho_1,\rho_2) \leqslant 1 \implies D_\infty(\mathcal{A}(\rho_1)\|\mathcal{A}(\rho_2)) \leqslant \varepsilon.$$

We show that even this definition causes the output state to be highly concentrated around zero, independent of the input state. In particular, we show that for two arbitrary quantum states ρ and σ, we have

$$\forall \rho,\sigma : D_\infty(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leqslant \varepsilon', \tag{13}$$

where $\varepsilon' = \varepsilon + \log(n - ne^{-\varepsilon} + e^{-\varepsilon})$. This can be seen considering the mixture $\rho' := \left(1 - \frac{1}{n}\right)\rho + \frac{\sigma}{n}$ and noting that $W_1(\rho,\rho') \leqslant 1$. Then, by the definition of ε-differential privacy,

$$\left(1 - \frac{1}{n}\right)\mathrm{Tr}[M_m\mathcal{A}(\rho)] + \frac{1}{n}\mathrm{Tr}[M_m\mathcal{A}(\sigma)] \\ = \mathrm{Tr}[M_m\mathcal{A}(\rho')] \leqslant e^{\varepsilon}\mathrm{Tr}[M_m\mathcal{A}(\rho)]$$

And thus

$$\begin{aligned}\mathrm{Tr}[M_m\mathcal{A}(\sigma)] &\leqslant e^{\varepsilon}(n-e^{-\varepsilon}(n-1))\mathrm{Tr}[M_m\mathcal{A}(\rho)]\\ &= e^{\varepsilon'}\mathrm{Tr}[M_m\mathcal{A}(\rho)],\end{aligned}$$

which implies Eq. (13). Then, for any input ρ, $\mathcal{A}(\rho)$ is ε-close to the maximally mixed state in quantum max-relative entropy, up to additive logarithmic factors. Applying Eq. (11) yields

$$\Pr_{\mathcal{A}(\rho)}(|O| \geqslant an) \leqslant K' e^{-a^2 n},$$

where where $K = e^{\varepsilon'} = e^{\varepsilon}(n - e^{-\varepsilon}(n-1))$. □

Proceeding similarly as for the trace distance, set $\rho := |1^n\rangle\langle 1^n|$ and $\varepsilon = 1$. Theorem 7.2 implies that

$$\Pr_{\mathcal{A}(\rho)}(|n-O| \leqslant 0.5n) = \Pr_{\mathcal{A}(\rho)}(O \geqslant 0.5n) \leqslant \Pr_{\mathcal{A}(\rho)}(|O| \geqslant 0.5n) \leqslant (en-(n-1))e^{-0.25n}$$

and hence setting $n = 100$ we obtain

$$\Pr_{\mathcal{A}(\rho)}(|n-O| \leqslant 0.5n) \leqslant 2.4 \times 10^{-9}.$$

Then the above example can be considered as a no-go result concerning $(\varepsilon, 0)$-DP under Wasserstein distance of order 1. We emphasise that the main argument of Theorem 7.2 is based on the construction of a classical mixed state, and then it holds both for the classical and the quantum W_1 distance. On the other hand, one could define the neighbouring relationship solely on pure states and hence overcome our no-go result. However, it is not obvious whether this definition can lead to a good privacy-utility tradeoff. We leave this possibility as an open problem for future explorations.

We also remark that $(0,\delta)$-DP under the W_1 distance is equivalent to $(0,\delta)$-DP with respect to $(1,1)$-neighbouring quantum states. Assume that a channel $\mathcal{A}$ is $(0,\delta)$-DP with respect to $(1,1)$-neighbouring quantum states and let $M = (M_1, \ldots, M_k)$ be a POVM measurement

$$\forall \rho_1 \overset{(1,1)}{\sim} \rho_2 \forall S \subseteq [k] \sum_{j\in S} \mathrm{Tr}\left[M_j(\mathcal{A}(\rho_1) - \mathcal{A}(\rho_2))\right] \leqslant \delta.$$

Then,

$$\begin{aligned}\sum_{j\in S}\mathrm{Tr}\left[M_j(\mathcal{A}(\rho)-\mathcal{A}(\sigma))\right] &\leqslant \sum_{j\in S}\sum_{i=1}^{n} c_i \mathrm{Tr}\left[M_j(\mathcal{A}(\rho^{(i)})-\mathcal{A}(\sigma^{(i)}))\right]\\ &= \sum_{i=1}^{n} c_i \sum_{j\in S}\mathrm{Tr}\left[M_j(\mathcal{A}(\rho^{(i)})-\mathcal{A}(\sigma^{(i)}))\right] \leqslant \sum_{i=1}^{n} c_i\delta = W_1(\rho,\sigma)\delta,\end{aligned}$$

where the last inequality follows from $\rho^{(i)} \overset{(1,1)}{\sim} \sigma^{(i)}$. Since $(1,1)$-neighbouring states satisfies $W_1(\rho,\sigma) \leqslant 1$, the equivalence follows.

7.2 A positive result for (ℓ, τ)-neighbouring states

We conclude this section with a positive result: adopting the definition introduced in Section 6, we can privately sample from an observable that approximates $O = \sum_{i=1}^{n} Z_i$, with a small loss in accuracy. We remark that the special case $\ell = \tau = 1$ has already been studied in [23].

Theorem 7.3 (Efficient private measurement for (ℓ, τ)-neighbouring states)**.** *Let $\mathcal{O}$ be the quantum to classical channel implementing a measurement of the observable $O = \sum_{i=1}^{n} Z_i$. Assume that a state ρ satisfies*

$$\Pr_{\rho}[|O - \langle O \rangle_{\rho}| > a] \leqslant b.$$

and let $\alpha := \frac{2\ell}{\log((e^{\varepsilon}-1)\tau^{-1}+1)} \approx 2\ell\tau\varepsilon^{-1}$. Then there exits a a quantum-to-classical channel $\mathcal{O}_{\varepsilon}$ such that:

1. *$\mathcal{O}_{\varepsilon}$ is ε-DP with respect to (ℓ, τ)-neighbouring states.*
2. *The following concentration inequality holds:*

$$\Pr[|\mathcal{O}_{\varepsilon}(\rho) - \langle O \rangle_{\rho}| > a + t\alpha] \leqslant b + e^{-t}.$$

Proof. Let $\Lambda_{\mathcal{L}}$ be the Laplace noise of magnitude α. The first part of the theorem follows directly from Theorem 5.3, by choosing $\mathcal{O}_{\varepsilon} = \Lambda_{\mathcal{L}} \circ \mathcal{O}$. Moreover, if $Y \sim \mathrm{Lap}(\alpha)$, then

$$\Pr[|Y| > t \cdot \alpha] = e^{-t}.$$

Define the event $E = \{|Y| \leqslant t\alpha\}$. Then we have

$$\begin{aligned} &\Pr[|\mathcal{O}_{\varepsilon}(\rho) - \langle O \rangle_{\rho}| > a + t\alpha] \\ &\leqslant \Pr[|\mathcal{O}_{\varepsilon}(\rho) - \langle O \rangle_{\rho}| > a + t\alpha | E] \Pr[E] + \Pr[|\mathcal{O}_{\varepsilon}(\rho) - \langle O \rangle_{\rho}| > a + t\alpha | \overline{E}] \Pr[\overline{E}] \\ &\leqslant \Pr_{\rho}[|O - \langle O \rangle_{\rho}| > a] + \Pr[|Y| > t \cdot \alpha] \leqslant b + e^{-t}. \end{aligned}$$

□

So, in particular, $\rho = |1^n\rangle\langle 1^n|$, we have that $\Pr_{\rho}[O = \langle O \rangle_{\rho}] = 1$ since ρ is an eigenvector of O. Then Theorem 7.3 yields

$$\Pr[|\mathcal{O}_{\mathcal{L}}(\rho) - n| < t \cdot \alpha] \geqslant 1 - e^{-t}.$$

Finally, we plot the upper bounds derived in this section in Fig. 3 and Fig. 4.

8 Privacy-preserving estimation of expected values

In this section, we provide differentially private mechanisms for estimating the expected values of observables given m copies of a quantum state. Despite their similarities, performing private measurements on a single state and privately estimating the expected value of these measurements given many copies are inherently different tasks. In principle, we could perform an ε-DP measurement on each copy and then average the results. Then the overall algorithm satisfies (ε', δ')-DP with $\varepsilon' \approx \varepsilon\sqrt{m \log(1/\delta')}$ by advanced composition (Theorem 6 in [22]).

However, this approach is highly suboptimal as the privacy loss (i.e. the parameter ε) grows as $\sqrt{m}$. We present here a simpler and more efficient approach based on the concentration

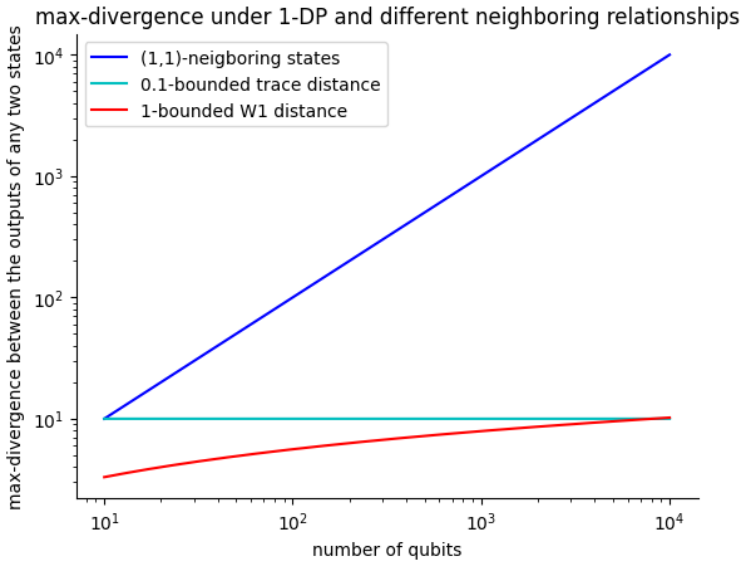

Figure 3: Upper bounds on the quantum max-relative entropy between any two states under 1-DP for several neighbouring relationships and various values of n.

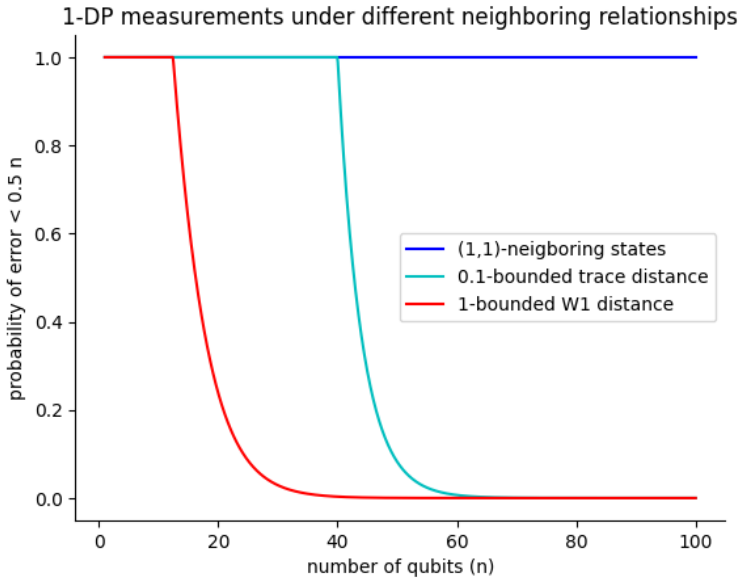

Figure 4: Upper bounds on the probability of achieving error lower than $0.5n$ for a measurement of $\frac{1}{n}\sum_{i=1}^{n} Z_i$ on the state $|1^n\rangle$, for several neighbouring relationships and various values of n. We assumed the input state undergoes a 1-DP channel.

of measure, whose privacy loss decreases as m increases. Given an observable O and set of quantum states equipped with a relationship denoted as $\overset{Q}{\sim}$, we'll define the *average quantum sensitivity* of O as follows:

$$\overline{\Delta}(O) = \max_{\rho \overset{Q}{\sim} \sigma} \mathrm{Tr}\{O(\rho - \sigma)\}.$$

Notably, we will present a simple technique whose privacy loss is proportional to $\overline{\Delta}(O) + \sqrt{1/m}$. This newly defined quantity is closely related to other notions introduced in prior work. Remark that the Lipschitz constant [28] can be recovered as a special case by considering as Q-neighbouring the states with W_1 distance at most one, i.e. $\rho \overset{Q}{\sim} \sigma \iff W_1(\rho, \sigma) = 1$. Moreover, if a quantum encoding $\rho(\cdot)$ is Q-neighbouring-preserving, then the above can be related to the classical definition of sensitivity introduced in Eq. (1). Consider the function $f(x) = \mathrm{Tr}\{O\rho(x)\}$, then

$$\Delta_f = \max_{x \sim x'} |f(x) - f(x')| \leqslant \max_{\rho(x) \overset{Q}{\sim} \rho(x')} |\mathrm{Tr}(O(\rho(x) - \rho(x'))| \leqslant \max_{\rho \overset{Q}{\sim} \sigma} |\mathrm{Tr}(O(\rho - \sigma))| := \overline{\Delta}(O)$$

We now prove that there exists a simple differentially private algorithm consisting of measurements and classical post-processing that gives a suitable tradeoff between sensitivity and privacy. We first consider a general post-processing channel and then we provide more concrete bounds for the Laplace and Gaussian noises.

Theorem 8.1. *Consider a neighbouring relationship $\overset{Q}{\sim}$ over the set of quantum states $\mathcal{S}_n$. Let $\rho^{\otimes m}$ be a collection of m copies of a quantum state $\rho \in S_n$ and O an observable. Let $\Lambda(\cdot)$ be a classical channel with the following property. For $\delta' \in (0, 1]$ and $x, x' \in \mathbb{R}$,*

$$|x - x'| \leqslant \overline{\Delta}(O) + \sqrt{m^{-1} \log(4/\delta')} \implies E_{e^\varepsilon}(\Lambda(x) \| \Lambda(x')) \leqslant \delta.$$

Consider the following algorithm $\mathcal{A}$:

1. *Measure O on each copy of ρ and collect the outcomes $y_1, \ldots, y_m$.*
2. *Compute the average $\hat{\mu} = \frac{1}{m} \sum_{i=1}^m y_i$ and output $\Lambda(\hat{\mu})$.*

Then the algorithm $\mathcal{A}$ is $(\varepsilon, \delta + \delta')$-DP.

Proof. Consider two neighbouring quantum states $\rho \overset{Q}{\sim} \sigma$. For $X \in \{\rho, \sigma\}$, let $\hat{\mu}_X$ the average obtained on input $X^{\otimes m}$. By Chernoff-Hoeffding's bound,

$$\Pr\left[|\hat{\mu}_X - \mathrm{Tr}[OX]| \geqslant \frac{t}{2}\right] \leqslant 2e^{-mt^2}.$$

Hence, by union bound,

$$\Pr[E] \leqslant \delta' := 4e^{-mt^2},$$

where E is the following event:

$$E := \left\{ \left(|\hat{\mu}_\rho - \mathrm{Tr}[O\rho]| \geqslant \frac{t}{2} \right) \vee \left(|\hat{\mu}_\sigma - \mathrm{Tr}[O\sigma]| \geqslant \frac{t}{2} \right) \right\}.$$

Conditioning on the complementary event $\overline{E}$ and observing that $t = \sqrt{m^{-1}\log(4/\delta')}$, we have,

$$\begin{aligned}|\hat{\mu}_\rho - \hat{\mu}_\sigma| &\leqslant |\hat{\mu}_\rho - \mathrm{Tr}[O\rho]| + |\mathrm{Tr}[O\rho] - \mathrm{Tr}[O\sigma]| + |\mathrm{Tr}[O\sigma] - \hat{\mu}_\sigma| \\ &\leqslant \Delta + t = \Delta + \sqrt{m^{-1}\log(4/\delta')}.\end{aligned}$$

This implies that, conditioning on $\overline{E}$,

$$E_{e^\varepsilon}(\Lambda(\hat{\mu}_\rho)\|\Lambda(\hat{\mu}_\sigma)) \leqslant \delta,$$

equivalently, we have

$$\forall S : \Pr[\Lambda(\hat{\mu}_\rho) \in S|\overline{E}] \leqslant e^\varepsilon \Pr[\Lambda(\hat{\mu}_\sigma) \in S|\overline{E}] + \delta.$$

Then we also have that, for all S

$$\begin{aligned}\Pr[\Lambda(\hat{\mu}_\rho) \in S] &= \Pr[\Lambda(\hat{\mu}_\rho) \in S|E]\Pr[E] + \Pr[\Lambda(\hat{\mu}_\rho) \in S|\overline{E}]\Pr[\overline{E}] \\ &\leqslant \Pr[\Lambda(\hat{\mu}_\rho) \in S|\overline{E}] + \delta' \leqslant e^\varepsilon \Pr[\Lambda(\hat{\mu}_\sigma) \in S|\overline{E}] + \delta + \delta' \\ &\leqslant e^\varepsilon \Pr[\Lambda(\hat{\mu}_\sigma) \in S] + \delta + \delta'.\end{aligned}$$

□

Finally, plugging the Laplace and the Gaussian channels in Theorem 8.1, we obtain the following corollary.

Corollary 8.1. *Let $\mathcal{A}, \rho^{\otimes m}$ and O as in Theorem 8.1 and let $\Delta := \overline{\Delta}(O)$. The following privacy guarantees hold.*

- *(Laplace noise) Let $\Lambda_{\mathcal{L},b}$ the Laplace channel of scale $b := (\Delta + \sqrt{m^{-1}\log(4/\delta')})/\varepsilon$. Then the algorithm $\mathcal{A}$ is (ε, δ')-DP.*
- *(Gaussian noise) Let $\Lambda_{\mathcal{G},\sigma}$ the Gaussian channel of variance*

$$\sigma^2 \geqslant 2\log(1.25/\delta)(\Delta + \sqrt{m^{-1}\log(4/\delta')})^2/\varepsilon^2.$$

  Then the algorithm $\mathcal{A}$ is $(\varepsilon, \delta + \delta')$-DP.

8.1 Bounding the average quantum sensitivity

Here we provide several bounds for the quantum sensitivity based on different neighbouring relationships. The first bound is based on Hölder's inequality, i.e. $|\mathrm{Tr}(LR)| \leqslant \|L\|_p\|R\|_q$ for $p^{-1} + q^{-1} = 1$, where $\|\cdot\|_p$ is the Schatten p-norm. Say that $\rho \overset{Q}{\sim} \sigma$ if $\|\rho - \sigma\|_p \leqslant \tau$. Then applying Hölder's inequality yields

$$\Delta(O) \leqslant \|O\|_q\tau.$$

For the special case of $p = 1$ (which corresponds to the trace distance) a stronger bound holds:

$$\Delta(O) = \max_{\rho,\sigma:\|\rho-\sigma\|_1\leqslant\tau} \mathrm{Tr}[O(\rho - \sigma)] \leqslant \frac{1}{2}\|O\|_\infty\|\rho - \sigma\|_1 \leqslant \frac{\tau}{2}\|O\|_\infty.$$

Table 2: Here we summarize the results of Section 8.1. For each neighbouring relationship over quantum states, we list the corresponding average quantum sensitivity $\Delta(O)$ of an observable O.

$\rho \overset{Q}{\sim} \sigma$	$\Delta(O)$
$\|\rho-\sigma\|_p \leqslant \tau$	$\tau\|O\|_q$
$\frac{1}{2}\|\rho-\sigma\|_1 \leqslant \tau$	$\tau\|O\|_1$
$W_1(\rho,\sigma) \leqslant \tau$	$\|O\|_{Lip}\tau$
$\rho \overset{(\Xi,\tau)}{\sim} \sigma$	$\min\{\frac{3}{2}\|O\|_{Lip}\max_{\mathcal{I}\in\Xi}\|\mathcal{I}\|\tau, \|O\|_{Lip}n\tau\}$

We can also consider a neighbouring relationship based on the Wasserstein distance of order 1, i.e. $\rho \overset{Q}{\sim} \sigma$ if $W_1(\rho,\sigma) \leqslant \tau$. Then the quantum sensitivity is proportional to the Lipschitz constant.

$$\Delta(O) = \max_{\rho,\sigma:W_1(\rho,\sigma)\leqslant\tau} \mathrm{Tr}\{O(\rho-\sigma)\} \leqslant \|O\|_{Lip}\tau.$$

By Lemma A.2, we also have that if $\rho \overset{(\Xi,\tau)}{\sim} \sigma$, then $W_1(\rho,\sigma) \leqslant \frac{3}{2}\max_{\mathcal{I}\in\Xi}|\mathcal{I}|\tau$. This implies

$$\Delta(O) = \max_{\rho,\sigma:\rho\overset{(\Xi,\tau)}{\sim}\sigma} \mathrm{Tr}\{O(\rho-\sigma)\} \leqslant \frac{3}{2}\|O\|_{Lip}\max_{\mathcal{I}\in\Xi}|\mathcal{I}|\tau.$$

The above bounds for $\Delta(O)$ are listed concisely in Table (2).

9 Private quantum machine learning

In this section, we demonstrate the applications of the results and tools we derived so far to variational quantum algorithms for machine learning. Let $\rho(\boldsymbol{\theta};\boldsymbol{x})$ be the output of a variational quantum circuit. We will assume that the parameters $\boldsymbol{\theta}$ are trained using a suitable (classical) dataset $S = (\boldsymbol{s}^{(1)},\ldots,\boldsymbol{s}^{(m)})$. Given a test set $\mathcal{X}$, we're asked to approximate a function $f : \mathbb{R}^d \to \mathbb{R}$. Thus, we can use variational quantum algorithms to find a set of parameters $\boldsymbol{\theta}$ that satisfy

$$\forall \boldsymbol{x} \in \mathcal{X} : f(\boldsymbol{x}) \simeq \mathrm{Tr}(O\rho(\boldsymbol{\theta};\boldsymbol{x})),$$

where O is a suitable observable. Given this simple scenario, differential privacy can come in different flavours.

- Let $\boldsymbol{x} = (x_1,\ldots,x_d) \in \mathcal{X}$ be the input vector. Given a neighbouring relationship $\boldsymbol{x} \sim \boldsymbol{x}'$, we can ensure differential privacy with respect to the input $\boldsymbol{x}$. This is particularly useful when $\boldsymbol{x}$ contains the sensitive information of multiple individuals or when $\boldsymbol{x}$ might be corrupted by an *adversarial attack*.
- In the alternative, we can require differential privacy with respect to the training set $S = (\boldsymbol{s}^{(1)},\ldots,\boldsymbol{s}^{(m)})$, where $S \sim S'$ if they differ only in a single entry $\boldsymbol{s}^{(j)}$. This notion of privacy is meant to protect the sensitive information of the individuals who compose the training set. Furthermore, it also enhances *generalisation*, i.e. it allows to upper bound of the discrepancy between the error on the training set and the generalisation error.

9.1 Private evaluation with respect to the input x

Given a suitable notion of neighbouring inputs $x \sim x'$, we want to find a neighbouring relationship over quantum states $\overset{Q}{\sim}$ such that $\rho(\cdot, \theta)$ is Q-neighbouring-preserving. In other terms, we need to ensure that

$$x \sim x' \implies \rho(x, \theta) \overset{Q}{\sim} \rho(x', \theta).$$

First, we select the relationship Q according to Table (1). If a single copy of $\rho(x, \theta)$ is available, we can make the measurement differentially private either by adding a final quantum noisy channel (Theorem 6.1) or by classical post-processing (Theorem 6.2). If, instead, we're able to prepare multiple copies of $\rho(x, \theta)$, it's convenient to post-process the average outcome with classical noise. Then differential privacy is guaranteed by Corollary 8.1.

Certified adversarial robustness

Now, we outline the connection between differential privacy and adversarial robustness, which has been previously established in [27] and extended to the quantum setting in [39, 51, 52]. We consider a slightly different setting, known as *k-class classification*, where a classification algorithm $\mathcal{A}$ outputs a label $y \in [k]$ on input x. For instance, for $k = 2$, we can consider an algorithm that outputs label 1 if x represents a dog and 2 if x represents a cat. Consider k observables $O_1, \ldots, O_k$, and assume, for simplicity, that their spectrum lies in $[0, 1]$. The algorithm $\mathcal{A}$ works as follows.

1. On input x, for each $i \in [k]$, the algorithm measures the observable O_i on the state $\rho(x, \theta)$ m times and stores the outcomes in $y_1^{(i)}, \ldots, y_m^{(i)}$.
2. For each $i \in [k]$, let $y^{(i)} = \sum_{j=1}^{m} y_j^{(i)}$.
3. $\mathcal{A}$ returns the index $i^* \in [k]$ such that $i^* = \arg\max y^{(i)}$.

We adopt Proposition 1 in [27] to the quantum setting.

Proposition 9.1 (Robustness condition)**.** *Let $\beta \in (0, 1]$. Let $\rho(\cdot, \theta)$ be Q-neighbouring-preserving and assume that each of the m measurements in step (1) satisfies (ε, δ)-DP with respect to Q-neighbouring quantum states. For any input x, if for some $i \in [k]$,*

$$y^{(i)} > e^{2\varepsilon} \max_{j \neq i} y^{(j)} + (1 + e^{\varepsilon})\delta + \sqrt{\frac{2}{m} \log\left(\frac{4k}{\beta}\right)}, \tag{14}$$

then the algorithm $\mathcal{A}$ satisfies, for all $x \sim x'$

$$\Pr[\mathcal{A}(x) = \mathcal{A}(x')] \geqslant 1 - \beta.$$

In this case, we say that the classifier $\mathcal{A}$ is β-robust to adversarial attacks.

Proof. Let $x \sim x'$. Since $\rho(\cdot, \theta)$ is Q-neighbouring-preserving, $\rho(x, \theta) \overset{Q}{\sim} \rho(x', \theta)$. The assumption that each measurement satisfies (ε, δ)-DP implies

$$\forall i \in [k], \forall F \subseteq \mathsf{range}(O_i) : \Pr_{\rho(x,\theta)}[O_i \in F] \leqslant e^{\varepsilon} \Pr_{\rho(x',\theta)}[O_i \in F] + \delta.$$

We first need to prove the following inequality. For all i,

$$\mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})] \leqslant e^{\varepsilon}\mathrm{Tr}[O_i\rho(\boldsymbol{x}',\boldsymbol{\theta})] + \delta. \tag{15}$$

Recall that the expectation of a non-negative random variable X can be expressed as

$$\mathbb{E}(X) = \int_{t\geqslant 0} \Pr[X > t]dt.$$

Combining this with differential privacy, we obtain

$$\begin{aligned}\mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})] &= \int_{t\geqslant 0} \Pr_{\rho(\boldsymbol{x},\boldsymbol{\theta})}[O_i > t]dt \\ \leqslant e^{\varepsilon}\int_{t\geqslant 0} \Pr_{\rho(\boldsymbol{x}',\boldsymbol{\theta})}[O_i > t]dt + \delta &= e^{\varepsilon}\mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})] + \delta,\end{aligned}$$

which proves Eq. (15). It remains to show that the discrepancy between $y^{(i)} = \frac{1}{m}\sum_{j=1}^{m} y_j^{(i)}$ and of $\mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})]$ is small enough with high probability. To this end, we can use concentration of measure. By Chernoff-Hoeffding's bound,

$$\Pr\left[\left|\frac{1}{m}\sum_{j=1}^{m} y_j^{(i)} - \mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})]\right| \geqslant t\right] \leqslant 2e^{-2mt^2}.$$

and thus $y^{(i)} = \mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})] \pm t$ with probability at least $1 - 2e^{-2mt^2}$. Denote by $\widetilde{y}^{(1)},\ldots,\widetilde{y}^{(k)}$ the average of the measurements on the state $\rho(\boldsymbol{x}',\boldsymbol{\theta})$. By union bound, with probability at least $1 - 4ke^{-2mt^2} = 1-\beta$ we have that

$$\forall i \in [k] : \left(y^{(i)} = \mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})] \pm t\right) \wedge \left(\widetilde{y}^{(i)} = \mathrm{Tr}[O_i\rho(\boldsymbol{x}',\boldsymbol{\theta})] \pm t\right). \tag{16}$$

Assume, by contradiction, that $\mathcal{A}(x) \neq \mathcal{A}(x')$ and Eq. (16) hold simultaneously Since $\mathcal{A}(x) \neq \mathcal{A}(x')$, there exists $i \neq i'$ such that

$$y^{(i)} > \max_{j\neq i} y^{(j)} \quad \text{and} \quad \widetilde{y}^{(i')} > \max_{j\neq i} \widetilde{y}^{(j)}.$$

Putting them all together, we have

$$\begin{aligned}\widetilde{y}^{(i)} \geqslant \mathrm{Tr}[O_i\rho(\boldsymbol{x}',\boldsymbol{\theta})] - t &\geqslant e^{-\varepsilon}(\mathrm{Tr}[O_i\rho(\boldsymbol{x},\boldsymbol{\theta})] - t) - e^{-\varepsilon}\delta \\ &\geqslant e^{-\varepsilon}(y^{(i)} - 2t) - e^{-\varepsilon}\delta > \max_{j\neq i} e^{\varepsilon}y^{(j)} + \delta \\ &\geqslant \max_{j\neq i} \widetilde{y}^{(j)} \geqslant \widetilde{y}^{(i')}\end{aligned}$$

Thus we obtained $\widetilde{y}^{(i)} > \widetilde{y}^{(i')}$ contradicting the assumptions $\mathcal{A}(x) \neq \mathcal{A}(x')$. This proves that $\mathcal{A}(x) = \mathcal{A}(x')$ with probability at least $1-\beta$. □

It's easy to see how the above proposition is related to adversarial attacks. Assume that an adversary has the capabilities of tampering with the input $\boldsymbol{x}$ by replacing it with $\boldsymbol{x}'$ such that $\boldsymbol{x} \sim \boldsymbol{x}'$. We remark that there's no unique way of choosing the neighbouring relationship in this

context, as it is closely related to the capabilities of the adversary. Under the same assumptions of Proposition 9.1, the adversarial attack doesn't alter the output with high probability. The condition expressed in Eq. (14) can be interpreted as the classifier being "fairly confident" about its prediction. We also remark that Proposition 9.1 can be applied to virtually any algorithm $\mathcal{A}$, even in the absence of an explicit private mechanism, since all algorithms are by default $(0, \tau)$-DP with respect to τ-neighbouring states. This can be easily checked from the properties of the trace distance.

Following [27], given a distribution $\mathcal{D}$ over labeled inputs of the form $(\boldsymbol{x}, f(\boldsymbol{x}))$, we can define the *certified accuracy* $\mathcal{R}(\mathcal{A})$ of an (ε, δ)-DP algorithm $\mathcal{A}$ as follows

$$\mathcal{R}(\mathcal{A}) := \Pr_{(\boldsymbol{x}, f(\boldsymbol{x}))\sim\mathcal{D}} \left[(i^* = f(\boldsymbol{x})) \wedge \left(\delta < \frac{y^{(i^*)} - e^{2\varepsilon} \max_{j\neq i^*} y^{(j)} - g(k, \beta, m)}{1 + e^{\varepsilon}} \right) \right],$$

where $g(k, \beta, m) := \sqrt{2m^{-1} \log(4k/\beta)}$ and $i^* = \arg\max y^{(i)}$. In other terms, $\mathcal{R}$ is a lower bound on the probability that an instance is classified correctly and the classification is β-robust to adversarial attacks. We remark that $\mathcal{R}$ can be easily estimated by computing the fraction of the test set that is classified correctly and, simultaneously, satisfies Eq. (14).

Numerical results. Finally, we complement our theoretical analysis with a numerical simulation implemented in PennyLane. We consider a classification task based on the first two classes of the famous IRIS dataset and each input $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ is susceptible to be perturbed by an adversarial attack. We assume that the adversary can select a single entry x_i and map it to x_i' with $|x_i - x_i'| \leqslant \tau$, for some threshold $0 \leqslant \tau \leqslant 1$. We trained a simple 4-qubit binary classifier, based on the variational circuit depicted in Fig. 5, whose gates are parametrised by a trainable vector $\boldsymbol{\theta}$ and the input vector $\boldsymbol{x}$. Hence, the output is measured according to $O = \frac{1}{8}\sum_{i=1}^{4}(Z_i + 1)$ and the classifier outputs 0 if the outcome is larger than 0.5 and 1 otherwise. It's easy to see that this encoding is $(1, \tau)$-privacy-preserving with respect to the neighbouring definition induced by the adversarial attack. The circuit is ended by a final layer of local depolarising noise $\mathcal{N}_p^{\otimes n}$, which ensures (ε, δ_1)-differential privacy with respect to $(1, \tau)$-neighbouring states, with δ_1 defined as in Theorem 6.1. We trained the model with the `Adam` optimiser [53] with several noise levels p and then we used the test set to estimate the certified accuracy for each p, and we plotted it against the threshold τ in Fig. 6. The results show that the noise level should be set according to attack threshold τ, as for $\tau \leqslant 0.2$ the circuit with $p = 0.1$ outperforms the others, while for $\tau \geqslant 0.2$ the circuit with $p = 0.3$ achieves the best certified accuracy.

Our simulation differs from previous experiments in multiple ways. First, we remark that our simulation combines local noisy channels with the novel neighbouring relationship we introduced in the present paper. In contrast to this, the simulation in [39] is based on τ-neighbouring states and ensures privacy via multiple layers of *global* depolarizing noise. On the other hand, [52] combines local noisy channels with τ-neighbouring states, resulting in privacy guarantees that degrade exponentially fast as the number of qubits increases. This stems from the fact that in Lemma 3 in [52], the authors show quantum differential privacy with $\varepsilon = \log(1 + \tau/p^n) \simeq \tau/p^n$. In addition, both [39] and [52] are based on ε-differential privacy while Proposition 9.1 is stated in terms of (ε, δ)-differential privacy. This is particularly useful to assess the certified accuracy of various noise regimes, including the case with no noise at all $(p = 0)$.

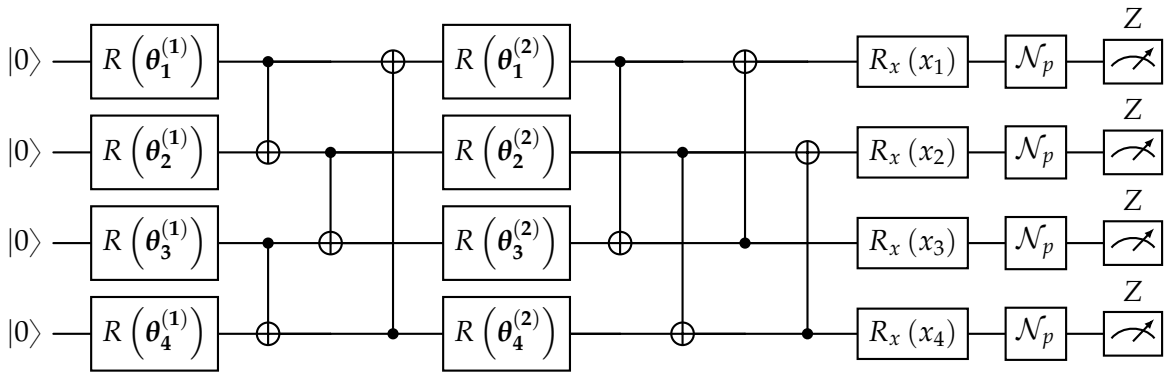

Figure 5: The parametric quantum circuit used in the simulation. We placed the encoding gates after the trainable gates in order to produce a $(1, \tau)$-neighbouring-preserving encoding. The output state is measured according to the observable $O = \frac{1}{8}\sum_{i=1}^{4}(Z_i + 1)$.

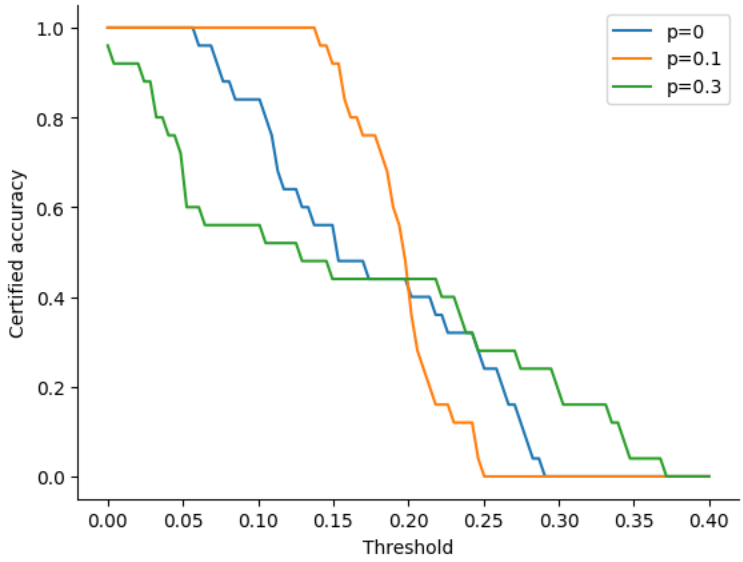

Figure 6: This plot contains the values of the certified accuracy estimated for various noise levels p and various attack thresholds τ.

9.2 Private prediction with respect to the training set S

Training a variational quantum algorithm involves finding a set of parameters $\boldsymbol{\theta}^*$ that minimizes a loss function $\mathcal{L}(\boldsymbol{\theta}, S) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Tr}\{O(y_i)\rho(\boldsymbol{\theta}; \boldsymbol{x_i})\} = \frac{1}{m}\sum_{i=1}^{m} \ell(\boldsymbol{\theta}, \boldsymbol{s_i})$ with respect to a given training set $S = (\boldsymbol{s_1}, \ldots, \boldsymbol{s_m})$ where $\boldsymbol{s_i} = (\boldsymbol{x_i}, y_i)$. In this setting, we let S and S' be neighbouring if $\exists i \in [m], \forall j \neq i : \boldsymbol{s_j} = \boldsymbol{s_j}$, i.e. if they differ in at most one element. Despite the existence of quantum algorithms for optimising a loss function, they're often not suitable for near-term devices. In most near-term applications, a variational quantum circuit is paired with a classical optimiser. Thus, standard techniques for differentially private (classical) optimisation can be adapted [54, 6]. For instance, Watkins et al. [20] implements the algorithm for private stochastic gradient descent (SGD) provided in [6] to optimize the parameters of a variational quantum circuit, achieving good empirical performance. The technique provided in [6] involves a procedure known as *gradient clipping*, which consists in rescaling the gradient $\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{s}_i)$ to ensure that its ℓ_2 norm is bounded by a suitable constant C, i.e. $\|\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{s}_i)\|_2 \leqslant C$. Then, privacy is ensured by the addition of Gaussian noise with variance proportional to C^2 on each estimate of the gradient. Instead of clipping the gradient, alternative techniques such as [54], estimates an upper bounds UB, where

$$\forall \boldsymbol{\theta} : \|\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{s_i})\|_2 \leqslant UB.$$

and add Gaussian noise proportional to UB^2 on each estimate of the gradient.

Here we show that UB can be easily estimated for some classes of variational quantum circuits. Assuming ℓ is differentiable with respect to $\boldsymbol{\theta}$ we have

$$|\ell(\boldsymbol{\theta}, \boldsymbol{s_i}) - \ell(\boldsymbol{\theta}', \boldsymbol{s_i})| \leqslant UB\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_{\ell_2} \implies \|\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{s_i})\|_2 \leqslant UB.$$

For $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, assume that each coordinate θ_j is encoded via a single gate Hamiltonian encoding, i.e. $e^{-i\theta_j H_j}$ with $\|H_i\|_2 \leqslant 1$. Moreover, assume that the output state is produced by a $1D$ circuit with bounded depth L (and thus the light-cone of each single qubit gate is upper bounded by $2L$). As shown in Appendix C, the Hamiltonian encoding $\rho(\cdot, \boldsymbol{s}_i)$ is (Ξ, τ)-neighbouring-preserving, where

$$\tau \leqslant \sqrt{\frac{d}{2}}\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2 \quad \text{and} \quad \max_{\mathcal{I} \in \Xi} |\mathcal{I}| \leqslant 2L.$$

Hence, we have

$$|\ell(\boldsymbol{\theta}, \boldsymbol{s_i}) - \ell(\boldsymbol{\theta}', \boldsymbol{s_i})| \leqslant |\mathrm{Tr}\{O(y_i)\rho(\boldsymbol{\theta}; \boldsymbol{x_i}) - \mathrm{Tr}\{O(y_i)\rho(\boldsymbol{\theta}'; \boldsymbol{x_i})\}|$$

$$\leqslant \|O(y_i)\|_{Lip} W_1(\rho(\boldsymbol{\theta}; \boldsymbol{x_i}), \rho(\boldsymbol{\theta}'; \boldsymbol{x_i})) \leqslant 3L\sqrt{\frac{d}{2}}\|O(y_i)\|_{Lip}\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2.$$

And then

$$\forall \boldsymbol{\theta} : \|\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{s_i})\|_2 \leqslant 3L\sqrt{\frac{d}{2}}\|O(y_i)\|_{Lip}.$$

Generalisation

We conclude by recalling the connection between differential privacy and generalisation. Given a randomised algorithm $M : \mathcal{X}^m \times \mathcal{X} \to [0, B]$ and two datasets $S, S' \in \mathcal{X}^m$ we define the following quantity:

$$\mathcal{E}_S[M(S)] := \frac{1}{m}\sum_{z \in S} \mathbb{E}_M[M(S, z)], \quad \mathcal{E}_{S'}[M(S)] := \frac{1}{m}\sum_{z' \in S'} \mathbb{E}_M[M(S, z')].$$

Lemma 9.1 (Lemma 6.4, [15]). *Let $S \in \mathcal{X}^m$ and $\boldsymbol{x} \in \mathcal{X}$. Let M be an algorithm that on input $(S, \boldsymbol{x})$ outputs a value $y \in [0, B]$. Assume that M is (ε, δ)-differentially private with respect to S, where $S \sim S'$ if they differ in at most one entry. Let $\mathcal{P}$ be an arbitrary distribution over $\mathcal{X}$. Then:*

$$\mathbb{E}_{S,S' \sim \mathcal{P}^m}[(\mathcal{E}_{S'}[M(S)])^k] \leqslant e^{k^2\varepsilon}\mathbb{E}_{S \sim \mathcal{P}^m}[(\mathcal{E}_{S'}[M(S)] + k\delta B)^k].$$

We also define $\mathcal{E}_{\mathcal{P}}[M(S)] := \mathbb{E}_{z \sim \mathcal{P},\, M}[M(S, z)]$. Clearly,

$$\mathbb{E}_{S' \sim \mathcal{P}^n}[\mathcal{E}_{S'}[M(S)]] = \mathcal{E}_{\mathcal{P}}[M(S)].$$

Moreover, as noted in [15], standard concentration inequalities implies that $\mathcal{E}_{S'}[M(S)]$ is strongly concentrated around $\mathcal{E}_{\mathcal{P}}[M(S)]$. Note that for $M(S, (x, y)) = \ell(M'(S, x), y)$, $\mathcal{E}_S[M(S)] = \mathcal{E}_S[\ell(M'(S))]$ and $\mathcal{E}_{\mathcal{P}}[M(S)] = \mathcal{E}_{\mathcal{P}}[\ell(M'(S))]$, in other words these are exactly the empirical and the expected loss of the predictor given by M'.

Acknowledgments

The authors thank Daniel Stilck-França, Christoph Hirche, Yihui Quek and Chirag Wadhwa for helpful discussions at different phases of this project. AA acknowledges financial support from the QICS (Quantum Information Center Sorbonne) and the H2020-FETOPEN Grant PHOQUSING (GA no.: 899544).

References

[1] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the netflix prize dataset, 2007.

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, page 265–284, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3540327312. doi: 10.1007/11681878_14. URL https://doi.org/10.1007/11681878_14.

[3] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. 9 (3–4):211–407, August 2014. ISSN 1551-305X. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

[4] Rachel Cummings, Damien Desfontaines, David Evans, Roxana Geambasu, Matthew Jagielski, Yangsibo Huang, Peter Kairouz, Gautam Kamath, Sewoong Oh, Olga Ohrimenko, et al. Challenges towards the next frontier in privacy. *arXiv preprint arXiv:2304.06929*, 2023.

[5] Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(29):1069–1109, 2011. URL http://jmlr.org/papers/v12/chaudhuri11a.html.

[6] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Oct 2016. doi: 10.1145/2976749.2978318. URL http://dx.doi.org/10.1145/2976749.2978318.

[7] Nicolas Papernot, Martín Abadi, Úlfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data, 2017.

[8] Raef Bassily, Om Thakkar, and Abhradeep Thakurta. Model-agnostic private learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 7102–7112, Red Hook, NY, USA, 2018. Curran Associates Inc.

[9] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM J. Comput.*, 40(3):793–826, June 2011. ISSN 0097-5397. doi: 10.1137/090756090. URL https://doi.org/10.1137/090756090.

[10] Yu-Xiang Wang, Jing Lei, and Stephen E. Fienberg. Learning with differential privacy: Stability, learnability and the sufficiency and necessity of erm principle. *Journal of Machine Learning Research*, 17(183):1–40, 2016. URL http://jmlr.org/papers/v17/15-313.html.

[11] M. Bun, R. Livni, and S. Moran. An equivalence between private classification and online prediction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 389–402, Los Alamitos, CA, USA, nov 2020. IEEE Computer Society. doi: 10.1109/FOCS46700.2020.00044. URL https://doi.ieeecomputersociety.org/10.1109/FOCS46700.2020.00044.

[12] Srinivasan Arunachalam, Yihui Quek, and John Smolin. Private learning implies quantum stability. In *Advances in Neural Information Processing Systems 34 pre-proceedings (NeurIPS 2021)*, NIPS'21, 2021.

[13] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth. Preserving statistical validity in adaptive data analysis. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 117–126, 2015.

[14] Raef Bassily, Kobbi Nissim, Adam Smith, Thomas Steinke, Uri Stemmer, and Jonathan Ullman. Algorithmic stability for adaptive data analysis. *SIAM Journal on Computing*, 50 (3):STOC16–377, 2021.

[15] Vitaly Feldman and Thomas Steinke. Generalization for adaptively-chosen estimators via stable median. In *Conference on Learning Theory*, pages 728–757. PMLR, 2017.

[16] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103, 2007. doi: 10.1109/FOCS.2007.66.

[17] Makhamisa Senekane, Mhlambululi Mafu, and Benedict Molibeli Taele. Privacy-preserving quantum machine learning using differential privacy. In *2017 IEEE AFRICON*, pages 1432–1435. IEEE, 2017.

[18] Weikang Li, Sirui Lu, and Dong-Ling Deng. Quantum federated learning through blind quantum computing. *Science China Physics, Mechanics & Astronomy*, 64(10), sep 2021. doi: 10.1007/s11433-021-1753-3. URL https://doi.org/10.1007%2Fs11433-021-1753-3.

[19] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, Shan You, and Dacheng Tao. Quantum differentially private sparse regression learning. *IEEE Transactions on Information Theory*, 68 (8):5217–5233, aug 2022. doi: 10.1109/tit.2022.3164726. URL https://doi.org/10.1109%2Ftit.2022.3164726.

[20] William M Watkins, Samuel Yen-Chi Chen, and Shinjae Yoo. Quantum machine learning with differential privacy. *Scientific Reports*, 13(1):2453, 2023.

[21] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. doi: 10.22331/q-2018-08-06-79. URL https://quantum-journal.org/papers/q-2018-08-06-79/. Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.

[22] Li Zhou and Mingsheng Ying. Differential privacy in quantum computation. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 249–262, 2017. doi: 10.1109/CSF.2017.23.

[23] Scott Aaronson and Guy N. Rothblum. Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 322–333, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367059. doi: 10.1145/3313276.3316378. URL https://doi.org/10.1145/3313276.3316378.

[24] Christoph Hirche, Cambyse Rouzé, and Daniel Stilck França. Quantum differential privacy: An information theory perspective, 2022. URL https://arxiv.org/abs/2202.10717.

[25] Farhad Farokhi. Privacy against hypothesis-testing adversaries for quantum computing, 2023.

[26] Theshani Nuradha, Ziv Goldfeld, and Mark M. Wilde. Quantum pufferfish privacy: A flexible privacy framework for quantum systems, 2023.

[27] Mathias Lecuyer, Vaggelis Atlidakis, Roxana Geambasu, Daniel Hsu, and Suman Jana. Certified robustness to adversarial examples with differential privacy, 2019.

[28] Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd. The quantum wasserstein distance of order 1. *IEEE Transactions on Information Theory*, 67(10):6627–6643, Oct 2021. ISSN 1557-9654. doi: 10.1109/tit.2021.3076442. URL http://dx.doi.org/10.1109/TIT.2021.3076442.

[29] Srinivasan Arunachalam, Alex B. Grilo, and Henry Yuen. Quantum statistical query learning, 2020. URL https://arxiv.org/abs/2002.08240.

[30] Armando Angrisani and Elham Kashefi. Quantum local differential privacy and quantum statistical query model. *ArXiv*, abs/2203.03591, 2022.

[31] Yuuya Yoshida and Masahito Hayashi. Classical mechanism is optimal in classical-quantum differentially private mechanisms. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1973–1977. IEEE, 2020.

[32] Yuuya Yoshida. Mathematical comparison of classical and quantum mechanisms in optimization under local differential privacy, 2021.

[33] Salil Vadhan. *The Complexity of Differential Privacy*, pages 347–450. Springer, Yehuda Lindell, ed., 2017. URL https://link.springer.com/chapter/10.1007/978-3-319-57048-8_7.

[34] Yury Polyanskiy, H. Vincent Poor, and Sergio Verdu. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010. doi: 10.1109/TIT.2010.2043769.

[35] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography: 14th International Conference, TCC 2016-B, Beijing, China, October 31-November 3, 2016, Proceedings, Part I*, pages 635–658. Springer, 2016.

[36] Sebastian Meiser. Approximate and probabilistic differential privacy definitions. *Cryptology ePrint Archive*, 2018.

[37] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE, aug 2017. doi: 10.1109/csf.2017.11. URL https://doi.org/10.1109%2Fcsf.2017.11.

[38] Marco Tomamichel. *Quantum information processing with finite resources: mathematical foundations*, volume 5. Springer, 2015.

[39] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, Dacheng Tao, and Nana Liu. Quantum noise protects quantum classifiers against adversaries. *Physical Review Research*, 3(2), May 2021. ISSN 2643-1564. doi: 10.1103/physrevresearch.3.023153. URL http://dx.doi.org/10.1103/PhysRevResearch.3.023153.

[40] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications*, 12(1):1791, 2021.

[41] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12(1), nov 2021. doi: 10.1038/s41467-021-27045-6. URL https://doi.org/10.1038%2Fs41467-021-27045-6.

[42] Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 6280–6290, Red Hook, NY, USA, 2018. Curran Associates Inc.

[43] Ewin Tang. A quantum-inspired classical algorithm for recommendation systems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 217–228, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367059. doi: 10.1145/3313276.3316310. URL https://doi.org/10.1145/3313276.3316310.

[44] Ewin Tang. Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions. *Physical Review Letters*, 127(6), Aug 2021. ISSN 1079-7114. doi: 10.1103/physrevlett.127.060503. URL http://dx.doi.org/10.1103/PhysRevLett.127.060503.

[45] András Gilyén, Seth Lloyd, and Ewin Tang. Quantum-inspired low-rank stochastic regression with logarithmic dependence on the dimension, 2018.

[46] Nai-Hui Chia, Han-Hsuan Lin, and Chunhao Wang. Quantum-inspired sublinear classical algorithms for solving low-rank linear systems, 2018. URL https://arxiv.org/abs/1811.04852.

[47] Nai-Hui Chia, András Gilyén, Tongyang Li, Han-Hsuan Lin, Ewin Tang, and Chunhao Wang. *Sampling-Based Sublinear Low-Rank Matrix Arithmetic Framework for Dequantizing Quantum Machine Learning*, page 387–400. Association for Computing Machinery, New York, NY, USA, 2020. ISBN 9781450369794. URL https://doi.org/10.1145/3357713.3384314.

[48] Jonathan Ullman. Cs7880: Rigorous approaches to data privacy, 2017. URL https://www.ccs.neu.edu/home/jullman/cs7880s17/HW1sol.pdf.

[49] Daniel Stilck França and Raul García-Patrón. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, oct 2021. doi: 10.1038/s41567-021-01356-3. URL https://doi.org/10.1038%2Fs41567-021-01356-3.

[50] Giacomo De Palma, Milad Marvian, Cambyse Rouzé, and Daniel Stilck França. Limitations of variational quantum algorithms: a quantum optimal transport approach. *PRX Quantum*, 4(1):010309, 2023.

[51] Christoph Hirche. Benefits and detriments of noise in quantum classification. 2023.

[52] Jhih-Cing Huang, Yu-Lin Tsai, Chao-Han Huck Yang, Cheng-Fang Su, Chia-Mu Yu, Pin-Yu Chen, and Sy-Yen Kuo. Certified robustness of quantum classifiers against adversarial examples through quantum noise. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[53] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[54] Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science*, pages 464–473. IEEE, 2014.

[55] Milán Mosonyi and Fumio Hiai. On the quantum rényi relative entropies and related capacity formulas. *IEEE Transactions on Information Theory*, 57(4):2474–2487, 2011.

[56] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, dec 2013. doi: 10.1063/1.4838856. URL https://doi.org/10.1063%2F1.4838856.

[57] Christoph Hirche, Cambyse Rouzé, and Daniel Stilck França. On contraction coefficients, partial orders and approximation of capacities for quantum channels. *Quantum*, 6:862, 2022.

[58] Tim van Erven and Peter Harremoes. Ré nyi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, jul 2014. doi: 10.1109/tit.2014.2320500. URL https://doi.org/10.1109%2Ftit.2014.2320500.

[59] Naresh Sharma and Naqueeb Ahmad Warsi. On the strong converses for the quantum channel capacity theorems. *arXiv preprint arXiv:1205.1712*, 2012.

[60] Christoph Hirche and Marco Tomamichel. Quantum r\'enyi and f-divergences from integral representations. *arXiv preprint arXiv:2306.12343*, 2023.

[61] Jean Bretagnolle and Catherine Huber. Estimation des densités: risque minimax. *Séminaire de probabilités de Strasbourg*, 12:342–363, 1978.

[62] Clément L. Canonne. A short note on an inequality between kl and tv, 2022.

[63] Chae-Yeun Park and Jaeyoon Cho. Correlations in local measurements and entanglement in many-body systems. *Physical Review A*, 98(1), jul 2018. doi: 10.1103/physreva.98.012107. URL https://doi.org/10.1103%2Fphysreva.98.012107.

[64] Maria Schuld. Supervised quantum machine learning models are kernel methods, 2021. URL https://arxiv.org/abs/2101.11020.

[65] Rupak Chatterjee and Ting Yu. Generalized coherent states, reproducing kernels, and quantum support vector machines. *arXiv preprint arXiv:1612.03713*, 2016.

[66] J. Berberich, D. Fink, and C. Holm. Robustness of quantum algorithms against coherent control errors, 2023.

A Preliminaries

We present here several technical tools that are used throughout the paper.

A.1 Schatten p-norms

Schatten p-norm can be used to define distances between linear operators. The Schatten p-norm of an operator $A \in \mathcal{L}(\mathcal{H}_n)$ is given by

$$\|A\|_p := [\mathrm{Tr}\{|A|^p\}]^{1/p},$$

where $|A| := \sqrt{A^\dagger A}$ and $p \geqslant 1$. For each $p \in [1, \infty]$, we consider the dual index q such that $\frac{1}{p} + \frac{1}{q} = 1$. The Hölder inequality gives:

$$\mathrm{Tr}\{A^\dagger B\} \leqslant \|A\|_p \|B\|_q. \tag{17}$$

A.2 Rényi divergences

Differential privacy, both in the classical and the quantum settings, can be expressed in terms of information-theoretic divergences. For two probability measures P, Q the Rényi divergences of order $\alpha \in (1, \infty)$ are defined as

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_{x\sim Q} \left(\frac{P(x)}{Q(x)}\right)^\alpha,$$

where we adopt the conventions that $0/0 = 0$ and $z/0 = \infty$ for $z > 0$. In the limit $\alpha \to 1$, the Rényi divergence reduces to the relative entropy, also known as the Kullback-Leibler divergence, i.e. $\lim_{\alpha\to 1} D_\alpha(P\|Q) = D(P\|Q) = \mathbb{E}_{x\sim P} \log \frac{P(x)}{Q(x)}$. Moreover, by taking the limit $\alpha \to \infty$, we obtain the max-divergence

$$D_\infty(P\|Q) = \sup_{S \subseteq \mathsf{supp}(Q)} \log \frac{P(S)}{Q(S)}.$$

We will also need the related smooth max-divergence,

$$D_\infty^\delta(P\|Q) = \sup_{S \subseteq \mathsf{supp}(Q) : P(S) \geqslant \delta} \log \frac{P(S) - \delta}{Q(S)}.$$

We emphasise that $D_\infty^\delta(P\|Q) \leqslant \varepsilon$ if and only if for every subset S,

$$P(S) \leqslant e^\varepsilon Q(S) + \delta.$$

Notably, the (smooth) max-divergence occurs in the definition of differential privacy.

Now we introduce divergences for quantum states. We make use of the quantum Petz-Rényi divergences [55, 56] of order $\alpha \in (1, \infty)$. For two states ρ, σ such that the support of ρ is included in the support of σ, they are defined as

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}].$$

In case the support of ρ is not contained in that of σ, all the divergences above are defined to be $+\infty$. In the limit $\alpha \to 1$, the quantum Petz-Rényi divergence reduces to the quantum relative entropy, i.e., $\lim_{\alpha\to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma) = \mathrm{Tr}[\rho(\log\rho - \log\sigma)]$. We also consider the divergence obtained by taking the limit $\alpha \to \infty$, known as quantum max-divergence,

$$D_\infty(\rho\|\sigma) = \inf\{\lambda : \rho \leqslant e^\lambda \sigma\},$$

and the related quantum smooth max-divergence [24],

$$D_\infty^\delta(\rho\|\sigma) = \inf_{\overline{\rho}\in B_\delta(\rho)} D_\infty(\overline{\rho}\|\sigma),$$

where $B^\delta(\rho) = \{\overline{\rho} : \overline{\rho}^\dagger = \overline{\rho} \geqslant 0 \wedge \|\rho - \overline{\rho}\|_1 < 2\delta\}$. Similarly to its classical counterpart, the quantum (smooth) max-divergence is at the heart of our work as it occurs in the definition of differentially private quantum channels.

Rényi divergences also play a key role in the analysis of quantum algorithms on noisy devices, as shown by the following result, which follows from Corollary 5.6 in [57].

Lemma A.1 (Supplementary Lemma 6, [41])**.** *Consider a single instance of the noise channel $\mathcal{N} = \mathcal{N}_1 \otimes \cdots \otimes \mathcal{N}_n$ where each local noise channel $\mathcal{N}_j$ is a Pauli noise channel that satisfies*

$$\mathcal{N}_j(\sigma) = q_\sigma \sigma$$

for $\sigma \in \{X, Y, Z\}$ and $|q_\sigma| < 1$. Let $q = \sqrt{\max\{|q_X|, |q_Y|, |q_Z|\}}$. Then, we have

$$D_2(\mathcal{N}(\rho)\|\mathbb{1}/2^n) \leqslant q^2 D_2(\rho\|\mathbb{1}/2^n). \tag{18}$$

The (standard) joint convexity of the Rényi divergence for $\alpha \in [0, \infty]$ is proven in [58] (Theorem 13). For the max divergence have

$$D_\infty(\sum_i \lambda_i P_i \| \sum_i \lambda_i Q_i) \leqslant \max_i D_\infty(P_i\|Q_i).$$

For the smooth max divergence, we can easily prove the statement from scratch. Assume $P_i(x) \leqslant e^\varepsilon Q_i(x) + \delta$:

$$\sum_i \lambda_i P_i(x) \leqslant \sum_i \lambda_i (e^\varepsilon Q_i(x) + \delta) = e^\varepsilon \left(\sum_i \lambda_i Q_i(x)\right) + \delta.$$

A.3 The quantum hockey-stick divergence

The quantum hockey-stick divergence was first introduced in [59], in the context of exploring strong converse bounds for the quantum capacity, and further investigate in [24, 60] in the context of quantum differential privacy. It is defined as

$$E_\gamma(\rho\|\sigma) := \mathrm{Tr}(\rho - \gamma\sigma)^+, \tag{19}$$

for $\gamma \geqslant 1$. Here X^+ denotes the positive part of the eigendecomposition of a Hermitian matrix $X = X^+ - X^-$. In [59] it was noted that this quantity is closely related to the trace norm via

$$E_\gamma(\rho\|\sigma) = \frac{1}{2}\|\rho - \gamma\sigma\|_1 + \frac{1}{2}(\mathrm{Tr}(\rho) - \gamma\mathrm{Tr}(\sigma)), \tag{20}$$

so for ρ, σ quantum states, $E_1(\rho\|\sigma) = \frac{1}{2}\|\rho - \sigma\|_1$ equals the trace distance. We also state some useful properties of the hockey-stick divergence proven in ([24], Proposition II.5).

- (Triangle inequality) For $\gamma_1, \gamma_2 \geqslant 1$ and $\rho, \sigma \in \mathcal{S}_n$, we have

$$E_{\gamma_1\gamma_2}(\rho\|\sigma) \leqslant E_{\gamma_1}(\rho\|\tau) + \gamma_1 E_{\gamma_2}(\tau\|\sigma). \tag{21}$$

- (Convexity) Let $\gamma_1, \gamma_2 \geqslant 1$, $\rho = \sum_x p(x)\rho_x$ and $\sigma = \sum_x q(x)\sigma_x$ with $\rho_x, \sigma_x \in \mathcal{S}_n$, we have

$$E_{\gamma_1\gamma_2}(\rho\|\sigma) \leqslant \sum_x p(x) E_{\gamma_1}(\rho_x\|\sigma_x) + \gamma_1 E_{\gamma_2}(\tilde{p}\|\tilde{q}), \tag{22}$$

  where $\tilde{p}$ and $\tilde{q}$ are non-normalised distributions $\tilde{p}(x) = p(x)\mathrm{tr}\sigma_x$ and $\tilde{q}(x) = q(x)\mathrm{tr}\sigma_x$, respectively. This also implies convexity and joint convexity.

- (Stability) For $\gamma \geqslant 1$ and $\rho, \sigma, \tau \in \mathcal{S}_n$, we have

$$E_\gamma(\rho \otimes \tau\|\sigma \otimes \tau) = \mathrm{tr}\,[\tau]\, E_\gamma(\rho\|\sigma). \tag{23}$$

A.4 The Wasserstein distance of order 1

We adopt the definition of quantum Wasserstein distance of order 1 proposed in [28]. This is based on the following notion of neighbouring quantum states, which also arises in the context of differentially private measurements [23]. We say that ρ and $\sigma \in \mathcal{S}_n$ are neighbouring if they coincide after discarding one qudit, i.e., if $\mathrm{Tr}_i\rho = \mathrm{Tr}_i\sigma$ for some $i \in [n]$. The quantum W_1 distance between the quantum states ρ and σ of $\mathcal{H}_n$ is defined as

$$W_1(\rho,\sigma) = \min\Bigg(\sum_{i=1}^n c_i : c_i \geqslant 0, \rho - \sigma = \sum_{i=1}^n c_i\left(\rho^{(i)} - \sigma^{(i)}\right), \\ \rho^{(i)}, \sigma^{(i)} \in \mathcal{S}_n, \mathrm{Tr}_i\rho^{(i)} = \mathrm{Tr}_i\sigma^{(i)}\Bigg).$$

The Wasserstein distance of order 1 and the trace distance are within a multiplicative factor n,

$$\frac{1}{2}\|\rho - \sigma\|_1 \leqslant W_1(\rho,\sigma) \leqslant \frac{n}{2}\|\rho - \sigma\|_1. \tag{24}$$

We also define the quantum Lipschitz constant of a self-adjoint linear operator $H \in \mathcal{O}_n$:

$$\|H\|_L = \max_{i\in[n]}(\max(\mathrm{Tr}[H(\rho - \sigma)] : \rho, \sigma \in \mathcal{S}_n, \mathrm{Tr}_i\rho = \mathrm{Tr}_i\sigma)).$$

From the definition of Wasserstein distance, we can readily derive that

$$\mathrm{Tr}[H(\rho - \sigma)] \leqslant \|H\|_L W_1(\rho,\sigma).$$

We also need the following technical lemma that can be used to upper bound the quantum W_1 distance under the action of a local evolution.

Lemma A.2 (Proposition 5, [28]). *Let $\mathcal{I} \subseteq [n]$, and let $\rho, \sigma \in \mathcal{S}_n$ such that $\mathrm{Tr}_\mathcal{I}\rho = \mathrm{Tr}_\mathcal{I}\sigma$,*

$$W_1(\rho,\sigma) \leqslant |\mathcal{I}|\frac{d^2-1}{d^2}\|\rho - \sigma\|_1.$$

Lemma A.3 (Proposition IV.8, [24]). *Given a noisy circuit $\mathcal{A}$ over n qubits, consisting in L layers interspersed by local depolarising noise of parameter $0 \leqslant p \leqslant 1$, we assume that each layer of the circuit is a quantum channel of light-cone $\mathcal{I}$. Then, we have that for any two input states ρ, σ we have*

$$W_1(\mathcal{A}(\rho), \mathcal{A}(\sigma)) \leqslant (2|\mathcal{I}|(1-p))^L W_1(\rho, \sigma),$$

and hence,

$$\frac{1}{2}\|\mathcal{A}(\rho) - \mathcal{A}(\sigma)\|_1 \leqslant \frac{n}{2}(2|\mathcal{I}|(1-p))^L \|\rho - \sigma\|_1,$$

In other words, the trace distance between any two output states vanishes in logarithmic depth as soon as p satisfies $2|I|(1-p) < 1$.

B Improved bounds for quantum divergences

We present two technical contributions that establish tighter bounds for quantum divergences. First, we prove here a quantum version of the Bretagnolle-Huber (BH) inequality [61, 62]. The proof closely follows the one of the classical BH inequality, and for this reason the quantum BH can be regarded as a folklore result. However, we include here the complete proof since, to the best of our knowledge, it doesn't appear in any previous reference. We remark that a different quantum generalisation of the BH inequality result was provided in [63] in the context of local measurements.

Lemma B.1 (Quantum Bretagnolle-Huber inequality)**.** *For every ρ, σ we have*

$$\frac{1}{2}\|\rho - \sigma\|_1 \leqslant \sqrt{1 - e^{-D(\rho\|\sigma)}}$$

Proof. We define the following quantity

$$\begin{aligned} U &:= \rho^{-1}\sigma, \\ V &:= (U - \mathbb{1})^+, \\ W &:= \mathbb{1} + V - U = (U - \mathbb{1})^-. \end{aligned}$$

It's well known that

$$\begin{aligned} \mathrm{Tr}(\rho V) &= \mathrm{Tr}(\sigma - \rho)^+ = \frac{1}{2}\|\rho - \sigma\|_1, \\ \mathrm{Tr}(\rho W) &= \mathrm{Tr}(\sigma - \rho)^- = \frac{1}{2}\|\rho - \sigma\|_1. \end{aligned}$$

Moreover, remark that $(1+V)(1-W) = U$ and hence $\log U = \log(\mathbb{1} + V) + \log(\mathbb{1} - W)$. Applying the Jensen's inequality, we obtain

$$\begin{aligned} -D(\rho\|\sigma) \leqslant \mathrm{Tr}[\rho \log(\rho^{-1}\sigma)] = \mathrm{Tr}[\rho \log U] \\ = \mathrm{Tr}[\rho \log(\mathbb{1} + V)] + \mathrm{Tr}[\rho \log(\mathbb{1} - W)] \leqslant \log \mathrm{Tr}[\rho(\mathbb{1} + V)] + \log \mathrm{Tr}[\rho(\mathbb{1} - W)] \\ = \log(1 - \mathrm{Tr}[\rho V]) + \log(1 - \mathrm{Tr}[\rho W]) = \log\left(1 - \frac{1}{2}\|\rho - \sigma\|_1^2\right). \end{aligned}$$

Exponentiating both sides, rearranging and taking the square root, proves the lemma. □

Following [42], we prove a quantum version of the *advanced joint convexity* of the hockey-stick divergence.

Lemma B.2 (Advanced joint convexity of the quantum hockey-stick divergence)**.** *For all states ρ_0, ρ_1, ρ_2 and $\gamma' = 1 + (1-p)(\gamma - 1)$, we have*

$$E_{\gamma'}(p\rho_0 + (1-p)\rho_1 \| p\rho_0 + (1-p)\rho_2) \leqslant (1-p)(1-\beta)E_\gamma(\rho_1\|\rho_0) + (1-p)\beta E_\gamma(\rho_1\|\rho_2),$$

where $\beta = \gamma'/\gamma$.

Proof. Recall that

$$E_\gamma(\rho\|\sigma) := \mathrm{Tr}(\rho - \gamma\sigma)^+ = \frac{1}{2}\|\rho - \gamma\sigma\|_1 + \frac{1}{2}(1-\gamma).$$

We have

$$\begin{aligned} E_{\gamma'}(p\rho_0 + (1-p)\rho_1 \| p\rho_0 + (1-p)\rho_2) &= \mathrm{Tr}[p\rho_0 + (1-p)\rho_1 - \gamma'(p\rho_0 + (1-p)\rho_2)]^+ \\ &= \mathrm{Tr}[p\rho_0 + (1-p)\rho_1 - (1 + (1-p)(\gamma-1))(p\rho_0 + (1-p)\rho_2)]^+ \\ &= (1-p)\mathrm{Tr}[\rho_1 - \gamma(\rho_0(1-\beta) + \beta\rho_2)]^+ = (1-p)E_\gamma(\rho_1\|\rho_0(1-\beta) + \beta\rho_2) \\ &\leqslant (1-p)(1-\beta)E_\gamma(\rho_1\|\rho_0) + (1-p)\beta E_\gamma(\rho_1\|\rho_2), \end{aligned}$$

where the inequality follows from the (standard) joint-convexity of the quantum hockey-stick divergence. □

C Quantum encodings

Quantum encodings, also known as quantum feature maps or quantum embedding, are classical-to-quantum functions mapping vectors to quantum states. In this section, we review some popular encodings and highlight their connection with various quantum distances and neighbouring relationships. We refer to [64] for more details about the encodings and their corresponding kernel (i.e. the value of $|\langle\psi_x|\psi_{x'}\rangle|^2$ for two vectors x, x'). Throughout this section, we will show that encoding vectors close in various p-distance leads to states that are either close in trace distance or that can be mapped one into the other by a local operation. The results of this section are summarised in Table (1).

Amplitude encoding. A normalised vector $x = (x_1, \ldots, x_{2^n}) \in \mathbb{C}^{2^n}$, $\|x\|_2 = 1$ can be represented by the amplitudes of a quantum state $|\psi_x\rangle$ via

$$x \mapsto |\psi_x\rangle = \sum_{j=1}^{2^n} x_j |j\rangle.$$

For two normalised vectors x, x' we have

$$|\langle\psi_x|\psi_{x'}\rangle| = |x^\dagger x'| = \left|1 - \frac{1}{2}\|x - x'\|_2^2\right|,$$

where the second identity holds for any pair of normalised vectors. Hence,

$$\begin{aligned}\frac{1}{2}\| \left|\psi_{\boldsymbol{x}}\right\rangle\!\left\langle\psi_{\boldsymbol{x}}\right| - \left|\psi_{\boldsymbol{x}'}\right\rangle\!\left\langle\psi_{\boldsymbol{x}'}\right| \|_1 &= \sqrt{1 - |\langle\psi_{\boldsymbol{x}}|\psi_{\boldsymbol{x}'}\rangle|^2} \\ &= \sqrt{1 - |\boldsymbol{x}^\dagger \boldsymbol{x}'|^2} = \sqrt{1 - \left(1 - \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2\right)^2} \\ &\leqslant \|\boldsymbol{x} - \boldsymbol{x}'\|_2.\end{aligned}$$

Rotation encoding. Rotation encoding is a qubit-based embedding without any normalisation condition. Given a vector $\boldsymbol{x}$ in the hypercube $[0, 2\pi]^{\otimes n}$, the i^{th} feature x_i is encoded into the i^{th} qubit via a Pauli rotation. For example, a Pauli-Y rotation puts the qubit into state $|q_i(x_i)\rangle = \cos(x_i)\,|0\rangle + \sin(x_i)\,|1\rangle$. The data-encoding feature map is therefore given by

$$\phi : \boldsymbol{x} \to \rho(\boldsymbol{x}) := |\phi(\boldsymbol{x})\rangle\!\langle\phi(\boldsymbol{x})| \text{ with } |\phi(\boldsymbol{x})\rangle = \sum_{q_1,\ldots,q_n=0}^{1} \prod_{k=1}^{n} \cos(x_k)^{q_k} \sin(x_k)^{1-q_k} |q_1, \ldots, q_n\rangle.$$

Let $\mathcal{I} = \{i : x_i \neq x_i'\}$. We have that $|\mathcal{I}| = \|\boldsymbol{x} - \boldsymbol{x}'\|_0$. We immediately see that

$$\mathrm{Tr}_{\mathcal{I}}\rho(\boldsymbol{x}) = \mathrm{Tr}_{\mathcal{I}}\rho(\boldsymbol{x}').$$

Coherent-state encoding. Coherent states are known in the field of quantum optics as a description of light modes. Formally, they are superpositions of so-called *Fock states*, which are basis states from an infinite-dimensional discrete basis $\{|0\rangle, |1\rangle, |2\rangle, ...\}$, instead of the binary basis of qubits. A coherent state has the form

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{k=0}^{\infty} \frac{\alpha^k}{\sqrt{k!}} |k\rangle,$$

for $\alpha \in \mathbb{C}$. Encoding a real scalar input $x_i \in \mathbb{R}$ into a coherent state $|\alpha_{x_i}\rangle$ corresponds to a data-encoding feature map with an infinite-dimensional feature space,

$$\phi : x_i \to |\alpha_{x_i}\rangle\!\langle\alpha_{x_i}|, \text{ with } |\alpha_{x_i}\rangle = e^{-\frac{|x_i|^2}{2}} \sum_{k=0}^{\infty} \frac{x_i^k}{\sqrt{k!}} |k\rangle.$$

We can encode a real vector $\boldsymbol{x} = (x_1, ..., x_n)$ into n joint coherent states,

$$|\alpha_{\boldsymbol{x}}\rangle = |\alpha_{x_1}\rangle \otimes \cdots \otimes |\alpha_{x_n}\rangle.$$

Following [64, 65], we have:

$$|\langle\alpha_{\boldsymbol{x}}|\alpha_{\boldsymbol{x}}\rangle|^2 = \left| e^{-\left(\frac{\|\boldsymbol{x}\|_2^2}{2} + \frac{\|\boldsymbol{x}'\|_2^2}{2} - \boldsymbol{x}^\dagger \boldsymbol{x}'\right)} \right|^2 = e^{-\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2},$$

and hence,

$$\frac{1}{2}\| \left|\phi_{\boldsymbol{x}}\right\rangle\!\left\langle\phi_{\boldsymbol{x}}\right| - \left|\phi_{\boldsymbol{x}'}\right\rangle\!\left\langle\phi_{\boldsymbol{x}'}\right| \|_1 = \sqrt{1 - |\langle\alpha_{\boldsymbol{x}}|\alpha_{\boldsymbol{x}}\rangle|^2} = \sqrt{1 - e^{-\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2}}.$$

Moreover, let $\mathcal{I} = \{i : x_i \neq x_i'\}$, where $|\mathcal{I}| = \|\boldsymbol{x} - \boldsymbol{x}'\|_0$. Hence,

$$\mathrm{Tr}_{\mathcal{I}}\rho(\boldsymbol{x}) = \mathrm{Tr}_{\mathcal{I}}\rho(\boldsymbol{x}').$$

Hamiltonian encoding. Let $\boldsymbol{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$ be a vector. Following Berberich et al. [66], consider the following parameterised quantum circuit

$$|\psi(\boldsymbol{x})\rangle = U_1(x_1) \cdots U_N(x_N) |\psi_0\rangle, \tag{25}$$

consisting of N parametric unitary operators $U_i(x_i) \in \mathcal{U}_n$ acting on the initial state $|\psi_0\rangle$. Let $\rho(\boldsymbol{x}) := |\psi(\boldsymbol{x})\rangle\langle\psi(\boldsymbol{x})|$. These unitaries can also be written as $U_j(x_j) = e^{-ix_jH_j}$, where the Hamiltonian $H_i = H_i^\dagger$ generates the gate U_i. The following result shows that quantum circuits are robust to slight perturbation of the classical parameters.

Lemma C.1 (Adapted from Theorem 2.2, [66]). *Let $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^N$. $U(\theta) = e^{-i\theta H}$. For any initial state $|\psi_0\rangle$ we have*

$$\| |\psi(\boldsymbol{x})\rangle\langle\psi(\boldsymbol{x})| - |\psi(\boldsymbol{x}')\rangle\langle\psi(\boldsymbol{x}')| \|_2 \leqslant \sum_{i=1}^{N} \|H_i\|_2 |x_i - x_i'| \leqslant \|\boldsymbol{x} - \boldsymbol{x}'\|_1 \max_i \|H_i\|_2.$$

Remark also that for ρ, σ pure states we have $\|\rho - \sigma\|_1 = \sqrt{2}\|\rho - \sigma\|_2$ and for any vectors $x, x' \in \mathbb{R}^N$ we have $\|\boldsymbol{x} - \boldsymbol{x}'\|_1 \leqslant \sqrt{N}\|\boldsymbol{x} - \boldsymbol{x}'\|_2$. Then we have:

$$\frac{1}{2}\|\rho(\boldsymbol{x}) - \rho(\boldsymbol{x}')\|_1 \leqslant \sqrt{\frac{1}{2}}\|\boldsymbol{x} - \boldsymbol{x}'\|_1 \max_i \|H_i\|_2 \tag{26}$$

$$\leqslant \sqrt{\frac{N}{2}}\|\boldsymbol{x} - \boldsymbol{x}'\|_2 \max_i \|H_i\|_2. \tag{27}$$

It's easy to see that the circuits $U(\boldsymbol{x})$ and $U(\boldsymbol{x}')$ coincides excepts for $\|\boldsymbol{x} - \boldsymbol{x}'\|_0$ gates. In order to investigate the local structure of the output, we need to introduce some assumptions on the circuit architecture. For instance, assuming that the circuit has 1-dimensional connectivity and depth L, there exists $\mathcal{I} \subseteq [n]$, $|\mathcal{I}| \leqslant 2L\|\boldsymbol{x} - \boldsymbol{x}'\|_0$, such that

$$\mathrm{Tr}_{\mathcal{I}}\rho(\boldsymbol{x}) = \mathrm{Tr}_{\mathcal{I}}\rho(\boldsymbol{x}').$$

C.1 Noisy encodings

A case of interest is when the circuit $U(\boldsymbol{x})$ is interspersed of L layers of local Pauli noise $\mathcal{P}_q$. Let $\mathcal{C}_x$ be the channel describing the composition of unitaries and noise:

$$\mathcal{C}_x(\rho_0) = \mathcal{P}_q^{\otimes n} \circ U_N(x_N)(\cdot)U_N(x_N)^\dagger \circ \mathcal{P}_q^{\otimes n} \circ \cdots \circ \mathcal{P}_q^{\otimes n} \circ U_1(x_1)(\rho_0)U_1^\dagger(x_1)$$

Then by Lemma A.1, we get:

$$D_2(\mathcal{C}_x(\rho_0)\|\mathbb{1}/2^n) \leqslant q^{2L}n.$$

and by Pinsker's inequality,

$$\frac{1}{2}\left\|\mathcal{C}_x(\rho_0) - \frac{\mathbb{1}}{2^n}\right\|_1 \leqslant \sqrt{\frac{q^{2L}n}{2}}. \tag{28}$$

Alternatively, by the quantum Bretagnolle-Huber inequality (Lemma B.1),

$$\frac{1}{2}\left\|\mathcal{C}_x(\rho_0) - \frac{\mathbb{1}}{2^n}\right\|_1 \leqslant \sqrt{1 - \exp\left(-q^{2L}n\right)}.$$

And by the triangle inequality

$$\frac{1}{2}\|\mathcal{C}_x(\rho_0) - \mathcal{C}_{x'}(\rho_0)\|_1 \leqslant 2\min\left\{\sqrt{\frac{q^{2L}n}{2}}, \sqrt{1 - \exp\left(-q^{2L}n\right)}\right\}.$$

High noise regime Now, assume that $\rho(\cdot)$ is an encoding post-processed by a channel $\mathcal{A}$, consisting in L layers such that each of them has light-cone $\mathcal{I}$ and its followed by local depolarising noise with noise parameter p. If p satisfies $2|\mathcal{I}|(1-p) < 1$, we have from Lemma A.3,

$$\begin{aligned}&\frac{1}{2}\|\mathcal{A}(\rho(\boldsymbol{x})) - \mathcal{A}(\rho(\boldsymbol{x}'))\|_1\\ \leqslant\ &(2|\mathcal{I}|(1-p))^L W_1(\rho(\boldsymbol{x}), \rho(\boldsymbol{x}'))\end{aligned}$$

For $\rho(\boldsymbol{x}) \overset{(\Xi,\tau)}{\sim} \rho(\boldsymbol{x}')$ we have

$$\frac{1}{2}\|\rho(\boldsymbol{x}) - \rho(\boldsymbol{x}')\|_1 \leqslant W_1(\rho(\boldsymbol{x}), \rho(\boldsymbol{x}')) \leqslant \min\left\{\max_{\mathcal{I}\in\Xi} |\mathcal{I}|\frac{3}{2}\tau, n\tau\right\}.$$

D Private quantum-inspired sampling

Our argument is similar to the one of (Problem 1.b, [48]) for uniform subsampling, but we include the complete proof here for clarity. Given a normalised vector $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$, let $|x\rangle := \sum_{i=1}^n x_i |i\rangle$ be the amplitude encoding defined in the previous section.

Theorem D.1 (DP amplification by quantum-inspired sampling)**.** *For any $x \in \mathbb{C}^n$, let $s = (s_1, \ldots, s_m)$ be the measurement outcomes in the computational basis of $|x\rangle^{\otimes m}$. Denote $\mathcal{S}$ as the sampling mechanism that maps x into s. Let $\mathcal{A}$ be a (ε, δ)-DP algorithm that takes only s as input. Then $\mathcal{A}' = \mathcal{A} \circ \mathcal{S}$ is (ε', δ')-DP, with $\varepsilon' = \log(1 + (e^\varepsilon - 1)m(\alpha + \beta))$ and $\delta' = \delta m(\alpha + \beta)$.*

Proof. We will use $T \subseteq \{1, \ldots, n\}$ to denote the identities of the m-subsampled elements $s_1, \ldots, s_m$ (i.e. their index, not their actual value). Note that T is a random variable and that the randomness of $\mathcal{A}' := \mathcal{A} \circ \mathcal{S}$ includes both the randomness of the sample T and the random coins of $\mathcal{A}$. Let $x \sim x'$ be adjacent datasets and assume that x and x' differ only on some row t. Let s (or s') be a subsample from x (or x') containing the rows in T . Let F be an arbitrary subset of the range of $\mathcal{A}$). For convenience, define $p = (\alpha + \beta)m$. Note that, by definition of quantum amplitude encoding and by union bound,

$$\Pr[i \in T] \leqslant m \times \Pr[|x\rangle \text{ collapses to state } |i\rangle] \leqslant m(\alpha + \beta) := p$$

To show $(\log(1 + p(e^\varepsilon - 1)), p\delta)$-DP, we have to bound the ratio

$$\frac{\Pr[\mathcal{A}'(x) \in F] - p\delta}{\Pr[\mathcal{A}'(x') \in F]} \leqslant \frac{p\Pr[\mathcal{A}(s) \in F | i \in T] + (1-p)\Pr[\mathcal{A}(s) \in F | i \notin T] - p\delta}{p\Pr[\mathcal{A}(s') \in F | i \in T] + (1-p)\Pr[\mathcal{A}(s') \in F | i \notin T]}$$

by $p(1 + (e^\varepsilon - 1))$. For simplicity, define the quantities

$$\begin{aligned}C &= \Pr[\mathcal{A}(s) \in F | i \in T]\\ C' &= \Pr[\mathcal{A}(s') \in F | i \in T]\\ D = \Pr[\mathcal{A}(s) \in F | i \notin T] &= \Pr[\mathcal{A}(s') \in F | i \notin T].\end{aligned}$$

We can rewrite the ratio as

$$\frac{\Pr[\mathcal{A}'(x) \in F] - p\delta}{\Pr[\mathcal{A}'(x') \in F]} = \frac{pC + (1-p)D - p\delta}{pC' + (1-p)D}.$$

Now we use the fact that, by (ε, δ)-DP, $C \leqslant \min\{C', D\} + \delta$. Plugging all together, we get

$$\begin{aligned} pC + (1-p)D - p\delta &\leqslant p(e^{\varepsilon} \min\{C', D\}) + (1-p)D \\ &\leqslant p(\min\{C', D\} + (e^{\varepsilon} - 1)\min\{C', D\}) + (1-p)D \\ &\leqslant p(C' + (e^{\varepsilon} - 1)(pC' + (1-p)D)) + (1-p)D \\ \leqslant (pC' + (1-p)D) + p(e^{\varepsilon} - 1))(pC' + (1-p)D) &\leqslant (1 + p(e^{\varepsilon} - 1))(pC' + (1-p)D), \end{aligned}$$

where the third-to-last line follow from $\min\{x, y\} \leqslant \alpha x + (1-\alpha)y$ for every $0 \leqslant \alpha \leqslant 1$. To conclude the proof, we rewrite the ratio and get the desired bound.

$$\frac{\Pr[\mathcal{A}'(x) \in F] - p\delta}{\Pr[\mathcal{A}'(x') \in F]} \leqslant 1 + p(e^{\varepsilon} - 1).$$

□