PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Farokhi F. 2024 Barycentric and pairwise Rényi quantum leakage with application to privacy-utility trade-off. *Proc. R. Soc. A* **480**: 20240319.
https://doi.org/10.1098/rspa.2024.0319

Received: 29 April 2024
Accepted: 10 October 2024

Subject Category:
Physics

Subject Areas:
quantum physics, algorithmic information theory, quantum computing

Keywords:
quantum information leakage, data privacy, information theory

Author for correspondence:
Farhad Farokhi
e-mail: farhad.farokhi@unimelb.edu.au

THE ROYAL SOCIETY PUBLISHING

Barycentric and pairwise Rényi quantum leakage with application to privacy-utility trade-off

Farhad Farokhi

Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia

FF, 0000-0002-5102-7073

Barycentric and pairwise quantum Rényi leakages are proposed as two measures of information leakage for privacy and security analysis in quantum computing and communication systems. These quantities both require minimal assumptions on the eavesdropper, i.e. they do not make any assumptions on the eavesdropper's attack strategy or the statistical prior on the secret or private classical data encoded in the quantum system. They also satisfy important properties of positivity, independence, post-processing inequality and unitary invariance. The barycentric quantum Rényi leakage can be computed by solving a semi-definite program; the pairwise quantum Rényi leakage possesses an explicit formula. The barycentric and pairwise quantum Rényi leakages form upper bounds on the maximal quantum leakage, the sandwiched quantum α-mutual information, the accessible information and the Holevo's information. Furthermore, differentially private quantum channels are shown to bound these measures of information leakage. Global and local depolarizing channels, that are common models of noise in quantum computing and communication, restrict private or secure information leakage. Finally, a privacy-utility trade-off formula in quantum machine learning using variational circuits is developed. The privacy guarantees can only be strengthened, i.e. information leakage can only be reduced, if the performance degradation grows larger and *vice versa*.

1. Introduction

Quantum computing provides various improvements over classical counterparts, such as speed up [1], security [2] and robustness [3]. These advantages have motivated considerable attention towards the theory and practice of quantum computing systems. A particularly fruitful direction is quantum machine learning [4]. However, machine learning and data analysis can result in unintended or undesired information leakage [5]. As quantum computing hardware becomes more commercially available and quantum algorithms move to the public domain, these privacy and security threats can prove to be detrimental in the adoption of quantum technologies, particularly for real-world sensitive, private or proprietary datasets. Therefore, we need to develop rigorous frameworks for understanding information leakage in quantum systems and construct secure and private algorithms by minimizing unintended information leakage. Note that the applications of these measures of information leakage are not entirely restricted to quantum machine learning or data privacy. Even in quantum communication, there is a need to understand how much information an eavesdropper can extract from the underlying quantum systems [6].

A common drawback of current notions of information leakage, such as quantum mutual information, accessible information and Holevo's information, is an implicit assumption that the intention of the eavesdropper or the adversary is known. It is assumed that the eavesdropper is interested in extracting the entirety of the classical data that is encoded in the state of the quantum system (for communication or analysis). While perfectly reasonable in computing the capacity of quantum channels and developing information storage or compression strategies, this assumption can be problematic in security or privacy analysis. In practice, we may not know the intention of the eavesdropper. Imposing extra assumptions on the eavesdropper is akin to underestimating its capabilities, which can be a lethal flaw in security or privacy analysis. Furthermore, in the classical setting, it is shown that mutual information and its derivatives are not suitable measures of information leakage for security and privacy analysis [7]. These observations motivated the development of a *maximal* notion of information leakage that is more suited to the task at hand [7,8].

Earlier attempts in developing the corresponding notion of maximal information leakage in quantum systems resulted in maximal quantum leakage [6], which was shown to satisfy important properties of positivity (i.e. information leakage is always greater than or equal to zero), independence property (i.e. information leakage is zero if the quantum state is independent of the classical data) and post-processing inequality (i.e. information leakage can be reduced if an arbitrary quantum channel is applied to the quantum state and therefore additional processing cannot increase information leakage). These properties are cornerstones of axiomatic frameworks for measuring information leakage in classical security analysis [7,8]. However, the maximal quantum leakage proposed in [6] did not possess an explicit formula and an iterative algorithm was required to compute it in general, cf accessible information [9].

In this paper, we propose two new measures of information leakage, namely, barycentric and pairwise quantum Rényi leakages, based on the quantum Rényi divergence or the sandwiched quantum Rényi relative entropy [10] of order ∞. These two quantities form upper bounds for the maximal quantum leakage in [6]. They both require minimal assumptions on the eavesdropper, i.e. they do not assume the eavesdropper's attack strategy is known and they do not require priors on the secret or private classical data encoded in the states of the quantum system for communication or analysis. Both barycentric and pairwise quantum Rényi leakages satisfy important properties of positivity, independence and post-processing inequality. They also satisfy unitary invariance, i.e. application of a unitary on the quantum state does not change the information leakage. Unitary invariance in postulated to be important for a quantum measure of information [10,11]. These measures are computationally superior to quantum maximal leakage. The barycentric quantum Rényi leakage can be computed by solving a semi-definite program, while the pairwise quantum Rényi leakage possesses an explicit formula. The barycentric and pairwise quantum Rényi leakages form upper bounds for the sandwiched quantum α-mutual information, the accessible information and the Holevo's information. Finally, we show that differentially private quantum channels [12] bound the barycentric and pairwise quantum Rényi leakages. Therefore, global and local depolarizing channels, that are common models of noise in quantum computing devices and quantum communication systems, are effective quantum channels for bounding private or secure information leakage. Using this result, we develop a privacy-utility trade-off in quantum machine learning using variational circuits. This fundamental trade-off demonstrates that the privacy guarantees can only be strengthened, i.e. information leakage is reduced, if the performance degradation becomes larger and *vice versa*. This is a novel characterization of the trade-off between privacy and utility in quantum machine learning. An information-theoretic analysis of privacy-utility trade-off in quantum machine learning has been missing. The closest analysis in this domain is the privacy-utility trade-off in [13], based on a variant of differential privacy, known as pufferfish privacy. Here, we relate utility to privacy via an operational notion of information leakage, namely, maximal quantum leakage, by investigating barycentric and pairwise quantum Rényi leakages. The effect of differential privacy in hypothesis testing is another interesting related problem [14,15], but its set-up differs from quantum machine learning significantly. There are also some limited results on numerical analysis of utility-privacy trade-offs [16,17], which are often tied to specific structures and training algorithms used in the underlying quantum machine learning model.

The rest of the paper is organized as follows. We first present some preliminary material on classical and quantum information theory in §2. We then formalize the barycentric and pairwise quantum Rényi leakages in §3. We establish the relationship between these measures of information leakage and quantum differential privacy in §4. We investigate privacy-utility trade-off for quantum machine learning in §5. We finally conclude the paper in §6.

2. Preliminary information

In this section, we review some basic concepts from classical and quantum information theory. A reader with this knowledge may benefit from directly jumping to §3.

(a) Random variables and classical information

Random variables are denoted by capital Roman letters, such as $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$. A random variable X is discrete if the set of all its possible outcomes $\mathbb{X}$ is finite. Any discrete random variable X is fully described by its probability mass function $p_X(x) := \mathbb{P}\{X = x\} > 0$ for all $x \in \mathbb{X}$. The support set of the random variable X is defined as $\mathrm{supp}(X) = \mathrm{supp}(p_X) := \{x \in \mathbb{X} \mid p_X(x) > 0\} \subseteq \mathbb{X}$. The set of all probability mass functions with domain $\mathbb{X}$ is $\Delta(\mathbb{X}) := \{\pi : \mathbb{X} \to \mathbb{R}_{\geq 0} \mid \sum_{x \in \mathbb{X}} \pi(x) = 1\}$.

For all $\alpha \in (0,1) \cup (1,\infty)$ and probability mass functions p, q such that $\mathrm{supp}(p) \subseteq \mathrm{supp}(q)$, the Rényi divergence (or relative entropy) of order α [18,19] is

$$d_\alpha(p\|q) := \frac{1}{\alpha - 1}\log\left(\sum_{x \in \mathrm{supp(q)}} p^\alpha(x) q^{1-\alpha}(x)\right).$$

All logarithms, in this paper, are in base 2 and, therefore, all the information quantities are measured in bits. On a few occasions, we require a logarithm in the natural basis, which is denoted by $\ln(\cdot)$ instead of $\log(\cdot)$. By convention, $d_\alpha(p\|q) = \infty$ if $\mathrm{supp}(p) \nsubseteq \mathrm{supp}(q)$ and $\alpha \geq 1$. For $\alpha < 1$, the same definition holds even if $\mathrm{supp}(p) \nsubseteq \mathrm{supp}(q)$. For $\alpha = 1, \infty$ (and also $\alpha = 0$ which is not used in this paper), we define the Rényi divergence by extension:

$$d_1(p\|q) := \lim_{\alpha \to 1} d_\alpha(p\|q) = \sum_{x \in \mathrm{supp}(q)} p(x) \log\left(\frac{p(x)}{q(x)}\right),$$
$$d_\infty(p\|q) := \lim_{\alpha \to \infty} d_\alpha(p\|q) = \log\left(\max_{x \in \mathrm{supp}(q)} \frac{p(x)}{q(x)}\right).$$

Note that $d_1(p\|q)$ is the usual Kullback–Leibler divergence [20, §2.3]. Sibson's α-mutual information, an extension of mutual information in information theory [21], between random variables $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ is

$$I_\alpha(X;Y) := \inf_{\tilde{q}} d_\alpha(P_{XY}\|P_X \times \tilde{q}), \tag{2.1}$$

where $P_{XY} \in \Delta(\mathbb{X} \times \mathbb{Y})$ is the joint probability mass function for the joint random variable (X, Y), $P_X \in \Delta(\mathbb{X})$ and $P_Y \in \Delta(\mathbb{Y})$ are the marginal probability mass functions for the random variable X and Y separately, and $\tilde{q} \in \Delta(\mathbb{Y})$ is any general probability mass function. By continuity [21], we get

$$I_1(X;Y) := \lim_{\alpha \to 1} I_\alpha(X;Y) = \sum_{(x,y) \in \mathrm{supp}(p_X) \times \mathrm{supp}(p_Y)} P_{X,Y}(x,y) \log\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right),$$
$$I_\infty(X;Y) := \lim_{\alpha \to \infty} I_\alpha(X;Y) = \log\left(\sum_{y \in \mathbb{Y}} \max_{x \in \mathrm{supp}(X)} P_{Y|X}(y|x)\right),$$

where $I_1(X;Y)$ is the common mutual information [20, §2.3]. For a more thorough treatment of the Rényi divergence and α-mutual information, see [21].

(b) Quantum states and information

A finite-dimensional Hilbert space is denoted by $\mathcal{H}$ while the set of linear operators from $\mathcal{H}$ to $\mathcal{H}$ is denoted by $\mathcal{L}(\mathcal{H})$. Further, $\mathcal{P}(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$ is the set of positive semi-definite operators on Hilbert space $\mathcal{H}$ and $\mathcal{S}(\mathcal{H}) \subset \mathcal{P}(\mathcal{H})$ is the set of density operators on $\mathcal{H}$, i.e. the set of positive semi-definite operators with unit trace. The state of a quantum system is modelled by a density operator in $\mathcal{S}(\mathcal{H})$. Lowercase Greek letters, such as ρ and σ, are often used to denote density operators or quantum states. A general formalism to model quantum measurements is the positive operator-valued measure (POVM), i.e. a set of positive semi-definite operators $F = \{F_i\}_i$ such that $\sum_i F_i = I$. For a quantum system with state ρ and POVM $F = \{F_i\}_i$, $\mathrm{tr}(\rho F_i) = \mathrm{tr}(F_i \rho)$ is the probability of obtaining output i when taking a measurement. This is typically called Born's rule. A quantum channel is a completely positive and trace-preserving mapping from any $\mathcal{S}(\mathcal{H}_A)$ to another $\mathcal{S}(\mathcal{H}_B)$. Calligraphic capital Roman letters, such as $\mathcal{E}$ and $\mathcal{N}$, are used to denote quantum channels. For a more detailed treatment of basic definitions and properties in quantum information theory, see [22].

For all $\alpha \in (0,1) \cup (1,\infty)$, and arbitrary $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$, the quantum Rényi relative entropy [23] is

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log\left(\mathrm{tr}\left(\rho^\alpha \sigma^{1-\alpha}\right)\right),$$

if the support set of ρ is contained within the support set of σ, denoted by $\rho \ll \sigma$, i.e. the kernel of operator σ lines within the kernel of operator ρ. By convention, $D_\alpha(\rho\|\sigma) = \infty$ if $\rho \not\ll \sigma$ and $\alpha \geq 1$. For $\alpha < 1$, the same definition holds even if $\rho \not\ll \sigma$. For $\alpha = 1, \infty$, we can define the quantum Rényi relative entropy by extension:

$$D_1(\rho\|\sigma) := \lim_{\alpha \to 1} D_\alpha(\rho\|\sigma) = \mathrm{tr}(\rho(\log(\rho) - \log(\sigma))),$$
$$D_\infty(\rho\|\sigma) := \lim_{\alpha \to \infty} D_\alpha(\rho\|\sigma) = \log\left(\max_{i,j:\langle i\|j\rangle \neq 0} \frac{v_i}{\lambda_j}\right),$$

where $\rho = \sum_i \upsilon_i |i\rangle\langle i|$ and $\sigma = \sum_j \lambda_j |\bar{j}\rangle\langle\bar{j}|$. Note that $D_1(\rho\|\sigma)$ is the usual quantum relative entropy [24]. Similarly, for all $\alpha \in (0,1) \cup (1,\infty)$, and arbitrary $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ such that $\rho \ll \sigma$, the quantum Rényi divergence or the sandwiched quantum Rényi relative entropy [10] is

$$\begin{aligned}\widetilde{D}_\alpha(\rho\|\sigma) &:= \frac{1}{\alpha-1}\log\left(\mathrm{tr}\left(\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right)\right)\\ &= \frac{1}{\alpha-1}\log\left(\mathrm{tr}\left(\left(\sigma^{\frac{1-\alpha}{\alpha}}\rho\right)^\alpha\right)\right)\\ &= \frac{1}{\alpha-1}\log\left(\mathrm{tr}\left(\left(\rho\sigma^{\frac{1-\alpha}{\alpha}}\right)^\alpha\right)\right). \end{aligned} \tag{2.2}$$

Again, by convention, $\widetilde{D}_\alpha(\rho\|\sigma) = \infty$ if $\rho \not\ll \sigma$ and $\alpha \geq 1$. For $\alpha < 1$, the same definition holds even if $\rho \not\ll \sigma$. Note that these alternative formulations stem from the fact that, for any two operators $\rho, \sigma \in \mathcal{L}(\mathcal{H})$, $\rho\sigma$ and $\sigma\rho$ have the same eigenvalues [25, exercise I.3.7]. For $\alpha = 1, \infty$, we can define the sandwiched quantum Rényi relative entropy by extension:

$$\begin{aligned}\widetilde{D}_1(\rho \parallel \sigma) &:= \lim_{\alpha\to 1}\widetilde{D}_\alpha(\rho \parallel \sigma) = \mathrm{tr}(\rho(\log(\rho) - \log(\sigma))),\\ \widetilde{D}_\infty(\rho \parallel \sigma) &:= \lim_{\alpha\to\infty}\widetilde{D}_\alpha(\rho \parallel \sigma) = \log(\inf\{\mu \in \mathbb{R}: \rho \leq \mu\sigma\}),\end{aligned}$$

where $\widetilde{D}_1(\rho\|\sigma) = D_1(\rho\|\sigma)$ is the usual quantum relative entropy [24] and $\widetilde{D}_\infty(\rho\|\sigma)$ is the max-relative entropy [26].

Lemma 2.1. *The following results hold for the sandwiched quantum Rényi relative entropy*:

(a) *Post-Processing Inequality*: *For any $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$, such that $\rho \ll \sigma$, and any quantum channel $\mathcal{E}$, $\widetilde{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq \widetilde{D}_\alpha(\rho\|\sigma)$ for all $\alpha \in [1,\infty]$;*
(b) *Order Axiom*: *$\widetilde{D}_\alpha(\rho\|\sigma) \leq 0$ if $\rho \leq \sigma$ and $\widetilde{D}_\alpha(\rho\|\sigma) \geq 0$ if $\rho \geq \sigma$ for all $\alpha \in [1,\infty]$;*
(c) *Unitary Invariance*: *$\widetilde{D}_\alpha(U\rho U^\dagger\|U\sigma U^\dagger) = \widetilde{D}_\alpha(\rho\|\sigma)$ for all $\alpha \in [1,\infty]$.*

Proof. The data-processing inequality follows by setting $\alpha \geq 1$ and $z = \alpha$ [11, theorem 1], where the case of $\alpha = \infty$ can be inferred by taking the limit $\alpha \to \infty$. The order axiom and unitary invariance follow from [10, theorem 2], where the cases of $\alpha = 1, \infty$ follow taking the limit. ∎

Lemma 2.2. *For any two mixture of states $\rho = \sum_{x\in\mathbb{X}}\pi(x)\rho^x$ and $\sigma = \sum_{x\in\mathbb{X}}\pi(x)\sigma^x$ with $\rho^x, \sigma^x \in \mathcal{S}(\mathcal{H})$ for all $x \in \mathbb{X}$ and $\pi \in \Delta(\mathbb{X})$, the max-relative entropy admits the quasi-convexity relationship*:

$$\widetilde{D}_\infty(\rho\|\sigma) \leq \max_{x\in\mathbb{X}} \widetilde{D}_\infty(\rho^x\|\sigma^x).$$

Proof. The proof follows from [26, lemma 9]. ∎

The quantum α-mutual information can be defined by expanding relation (2.1) to quantum states as

$$I_\alpha(A;B)_{\rho_{AB}} := \inf_{\sigma_B} D_\alpha(\rho_{AB}\|\rho_A \otimes \sigma_B),$$

where ρ_{AB} denotes the bipartite quantum state in $\mathcal{S}(\mathcal{H}_A \times \mathcal{H}_B)$ and $\rho_A = \mathrm{tr}_B(\rho_{AB}) \in \mathcal{S}(\mathcal{H}_A)$. Similarly, we can define the sandwiched quantum α-mutual information as

$$\widetilde{I}_\alpha(A;B)_{\rho_{AB}} := \inf_{\sigma_B} \widetilde{D}_\alpha(\rho_{AB}\|\rho_A \otimes \sigma_B).$$

3. Information leakage to arbitrary eavesdropper

In this section, we first need to present the definition of maximal quantum leakage, originally defined and studied in [6], to quantify information leakage to arbitrary eavesdroppers. To do so, we need to borrow the following concepts from [6].

Let discrete random variable $X \in \mathbb{X}$ model, the private or secure classical data. For any $x \in \mathbb{X}$, quantum system A with state $\rho_A^x \in \mathcal{S}(\mathcal{H}_A)$ is prepared. The ensemble of the states $\mathcal{E} := \{p_X(x), \rho_A^x\}_{x \in \mathbb{X}}$ captures the quantum encoding of the classical data. The average or expected density operator is $\rho_A = \mathbb{E}\{\rho_A^X\} = \sum_{x \in \mathbb{X}} p_X(x)\rho_A^x$, which models the state of the system A without knowing the realization of a random variable X. We assume that an arbitrary eavesdropper, who does not know the realization of the random variable X, wants to reliably guess or estimate the realization of a (possibly randomized) discrete function of the random variable X, denoted by a random variable Z, based on measurements from a single copy of the quantum state of system A. The security analyst is not aware of the intention of the eavesdropper (i.e. the nature of the random variable Z that is of interest to the eavesdropper). This set-up also covers the case that the eavesdropper searches for the random variable Z, i.e. an attack strategy, that results in the maximal information leakage. For a given POVM $F = \{F_y\}_{y \in \mathbb{Y}}$, discrete random variable $Y \in \mathbb{Y}$ denotes the outcome of the measurement such that the probability of obtaining measurement outcome $Y = y \in \mathbb{Y}$ when taking a measurement on quantum state ρ_A^x is $\mathbb{P}\{Y = y \mid X = x\} := \mathrm{tr}(\rho_A^x F_y)$. The eavesdropper makes a one-shot[1] guess of the random variable Z denoted by the random variable $\widehat{Z}$. Maximal quantum leakage measures the multiplicative increase in the probability of correctly guessing the realization of the random variable Z based on access to the quantum encoding of the data via ensemble $\mathcal{E} = \{p_X(x), \rho_A^x\}_{x \in \mathbb{X}}$.

Definition 3.1. (Maximal quantum leakage). *The maximal quantum leakage from random variable X through the quantum encoding of the data via ensemble $\mathcal{E} = \{p_X(x), \rho_A^x\}_{x \in \mathbb{X}}$ is*

$$\mathcal{Q}(X \to A)_{\rho_A} := \sup_{\{F_y\}_y} \sup_{Z,\widehat{Z}} \log\left(\frac{\mathbb{P}\{Z = \widehat{Z}\}}{\max_{z \in \mathbb{Z}} \mathbb{P}\{Z = z\}}\right) \tag{3.1a}$$

$$= \sup_{\{F_y\}_y} I_\infty(X; Y) \tag{3.1b}$$

$$= \sup_{\{F_y\}_y} \log\left(\sum_{y \in \mathbb{Y}} \max_{x \in \mathbb{X}} \mathrm{tr}\,(\rho_A^x F_y)\right), \tag{3.1c}$$

where, in relation (3.1a), the inner supremum is taken over all random variables Z and $\widehat{Z}$ with equal arbitrary finite support sets and the outer supremum is taken over all POVMs $F = \{F_y\}_{y \in \mathbb{Y}}$ with arbitrary finite set of outcomes $\mathbb{Y}$.

As noted in [6], there is no explicit formula for the maximal quantum leakage and an iterative algorithm must be used to compute this quantity for various quantum encoding methods. This motivates developing upper bounds for maximal quantum leakage that are easier to compute. Here, we develop two upper bounds for the maximal quantum leakage and show that these novel quantities also satisfy important properties or axioms for measures of information leakage; they are, however, considerably simpler to compute. They can be either reformulated as a semi-definite program or possess an explicit form.

Proposition 3.1. *The maximal quantum leakage from random variable X through the quantum encoding of the data via ensemble $\mathcal{E} = \{p_X(x), \rho_A^x\}_{x \in \mathbb{X}}$ is upper bounded by*

$$\mathcal{Q}(X \to A)_{\rho_A} \leq \min_{\pi \in \Delta(\mathbb{X})} \max_{x \in \mathbb{X}} \widetilde{D}_\infty\left(\rho_A^x \middle\| \sum_{x' \in \mathbb{X}} \pi(x')\rho_A^{x'}\right).$$

[1]It is shown that number of guesses is immaterial in evaluating the maximal information leakage [6].

royalsocietypublishing.org/journal/rspa *Proc. R. Soc. A* **480**: 20240319

Proof. For all $\alpha \geq 1$, we have

$$\begin{aligned} I_\alpha(X;Y) &\leq \frac{1}{\alpha-1}\log\left(\sum_{x\in\mathbb{X}} p_X(x)\mathrm{tr}\left(\left(\rho_A^x\rho_A^{\frac{1-\alpha}{\alpha}}\right)^\alpha\right)\right) \\ &\leq \frac{1}{\alpha-1}\log\left(|\mathbb{X}|\max_{x\in\mathbb{X}}\mathrm{tr}\left(\left(\rho_A^x\rho_A^{\frac{1-\alpha}{\alpha}}\right)^\alpha\right)\right) \\ &= \frac{\log(|\mathbb{X}|)}{\alpha-1} + \max_{x\in\mathbb{X}}\frac{1}{\alpha-1}\log\left(\mathrm{tr}\left(\left(\rho_A^x\rho_A^{\frac{1-\alpha}{\alpha}}\right)^\alpha\right)\right) \\ &= \frac{\log(|\mathbb{X}|)}{\alpha-1} + \max_{x\in\mathbb{X}}\widetilde{D}_\alpha(\rho_A^x\|\rho_A), \end{aligned}$$

where the first inequality follows from [27, eqns (17) and (23)]. Therefore,

$$\begin{aligned} I_\infty(X;Y) &= \lim_{\alpha\to\infty} I_\alpha(X;Y) \\ &\leq \lim_{\alpha\to\infty}\max_{x\in\mathbb{X}}\widetilde{D}_\alpha(\rho_A^x\|\rho_A) \\ &= \max_{x\in\mathbb{X}}\lim_{\alpha\to\infty}\widetilde{D}_\alpha(\rho_A^x\|\rho_A) \qquad (3.2) \\ &= \max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A), \qquad (3.3) \end{aligned}$$

where relation (3.2) follows from lemma A.1 in Appendix A. Finally,

$$\begin{aligned} \mathcal{Q}(X\to A)_{\rho_A} &= \min_{p_X\in\Delta(\mathbb{X})}\mathcal{Q}(X\to A)_{\rho_A} \\ &= \min_{p_X\in\Delta(\mathbb{X})}\sup_{\{F_y\}_y} I_\infty(X;Y) \\ &\leq \min_{p_X\in\Delta(\mathbb{X})}\sup_{\{F_y\}_y}\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A) \\ &= \min_{p_X\in\Delta(\mathbb{X})}\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A) \\ &= \min_{p_X\in\Delta(\mathbb{X})}\max_{x\in\mathbb{X}}\widetilde{D}_\infty\left(\rho_A^x\middle\|\sum_{x'\in\mathbb{X}} p_X(x')\rho_A^{x'}\right), \end{aligned}$$

where the first equality follows from that $\mathcal{Q}(X\to A)_{\rho_A}$ is not a function of p_X, the second equality follows from the definition of $\mathcal{Q}(X\to A)_{\rho_A}$, the inequality follows from relation (3.3), the third equality follows from that $\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A)$ is not a function of $\{F_y\}_y$ and the last equality is based on rewriting ρ_A as $\sum_{x'\in\mathbb{X}} p_X(x')\rho_A^{x'}$. ∎

This upper bound suggests introducing a new measure of information leakage for quantum encoding of classical data, referred to as barycentric quantum Rényi leakage. Barycenter, a term popular in astrophysics, refers to the centre of mass of two or more bodies that orbit one another. In this instance, barycenter refers to a quantum states that minimizes the worst-case distance for all quantum encoding $(\rho_A^x)_{x\in\mathbb{X}}$. Barycentric quantum divergences have been used in the past to measure the information content of a quantum encoding, albeit when the distance cost is in a sum form [28].

Definition 3.2. (Barycentric quantum Rényi leakage). *The barycentric quantum Rényi leakage from random variable X through the quantum encoding of the data via ensemble* $\mathcal{E} = \{p_X(x), \rho_A^x\}_{x\in\mathbb{X}}$ *is*

$$\mathcal{B}(X\to A)_{\rho_A} := \min_{\pi\in\Delta(\mathbb{X})}\max_{x\in\mathbb{X}}\widetilde{D}_\infty\left(\rho_A^x\middle\|\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'}\right). \qquad (3.4)$$

The following corollary immediately follows from proposition 3.1 and definition 3.2.

Corollary 3.1. $\mathcal{Q}(X\to A)_{\rho_A} \leq \mathcal{B}(X\to A)_{\rho_A}$.

royalsocietypublishing.org/journal/rspa *Proc. R. Soc. A* **480:** 20240319

Remark 3.1. (Quantum Rényi divergence radius). *Note the similarity between barycentric quantum Rényi leakage in relation (3.4) and quantum Rényi divergence radius in* [29, *eqn* (85)]:

$$\mathrm{Rad}(\{\rho_A^x\}_{x\in\mathbb{X}}) := \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{x\in\mathbb{X}} \widetilde{D}_\infty(\rho_A^x\|\sigma). \tag{3.5}$$

The difference between these two measures is that, in relation (3.4), the minimization is over the convex hull of $\{\rho_A^x\}_{x\in\mathbb{X}}$ *while, in relation (3.5), the minimization is over the set of all density operators. In general, it is easy to see that*

$$\mathrm{Rad}(\{\rho_A^x\}_{x\in\mathbb{X}}) \leq \mathcal{B}(X\to A)_{\rho_A}.$$

Understanding whether these two notions of information are identical remains an open question.

The barycentric quantum Rényi leakage satisfies important properties or axioms for information leakage: positivity (the information leakage is greater than or equal to zero), independence property (the information leakage is zero if the quantum state is independent of the classical data), post-processing inequality (the information leakage can be reduced if a quantum channel is applied to the state) and unitary invariance (the information remains constant by application of a unitary operator on the quantum state). The corresponding classical properties of positivity, independence and post-processing have been postulated to be of utmost importance in security analysis [7].

Theorem 3.1. *The following properties hold for the barycentric quantum Rényi leakage*:

(a) *Positivity and independence*: $\mathcal{B}(X\to A)_{\rho_A}\geq 0$ *with equality if and only if* $\rho_A^x=\rho_A^{x'}$ *for all* $x,x'\in\mathbb{X}$;
(b) *Unitary invariance*: $\mathcal{B}(X\to A)_{U\rho_A U^\dagger}=\mathcal{B}(X\to A)_{\rho_A}$ *for any unitary* U;
(c) *Data-processing inequality*: $\mathcal{B}(X\to A)_{\mathcal{E}(\rho_A)}\leq\mathcal{B}(X\to A)_{\rho_A}$ *for any quantum channel* $\mathcal{E}$.

Proof. We start by proving part (a) by contradiction. First, note that $\mathcal{B}(X\to A)_{\rho_A}\geq 0$ because $\mathcal{Q}(X\to A)_{\rho_A}\leq\mathcal{B}(X\to A)_{\rho_A}$ (see corollary 3.1) and $\mathcal{Q}(X\to A)_{\rho_A}\geq 0$ [6, proposition 1]. If $\rho_A^x=\rho_A^{x'}$ for all $x,x'\in\mathbb{X}$, we get

$$\begin{aligned}\mathcal{B}(X\to A)_{\rho_A} &= \min_{\pi\in\Delta(\mathbb{X})}\max_{x\in\mathbb{X}}\widetilde{D}_\infty\left(\rho_A^x\middle\|\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'}\right)\\ &= \min_{\pi\in\Delta(\mathbb{X})}\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A^x)\\ &= 0,\end{aligned}$$

where the second equality follows from $\sum_{x\in\mathbb{X}}\pi(x)=1$. On the other hand, $\mathcal{B}(X\to A)_{\rho_A}=0$ implies that, for all $x\in\mathbb{X}$ and $\pi\in\Delta(\mathbb{X})$, $\rho_A^x=\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'}$. This is only possible if $\rho_A^x=\rho_A^{x'}$ for all $x,x'\in\mathbb{X}$.

Now, we can prove part (b). The proof of this part follows from that

$$\begin{aligned}\widetilde{D}_\infty\left(U\rho_A^xU^\dagger\middle\|\sum_{x'\in\mathbb{X}}\pi(x')U\rho_A^{x'}U^\dagger\right) &= \widetilde{D}_\infty\left(U\rho_A^xU^\dagger\middle\|U\left(\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'}\right)U^\dagger\right)\\ &= \widetilde{D}_\infty\left(\rho_A^x\middle\|\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'}\right),\end{aligned}$$

where the second equality follows from unitary invariance in lemma 2.1.

Finally, we can prove part (c). Note that

$$\widetilde{D}_\infty\left(\mathcal{E}(\rho_A^x)\middle\|\sum_{x'\in\mathbb{X}}\pi(x')\mathcal{E}(\rho_A^{x'})\right)=\widetilde{D}_\infty\left(\mathcal{E}(\rho_A^x)\middle\|\mathcal{E}\left(\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'}\right)\right)$$
$$\leq \widetilde{D}_\infty\left(\rho_A^x\middle\|\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'}\right),$$

where the equality stems from the linearity of $\mathcal{E}$ and the inequality follows from the data-processing inequality in lemma 2.1. ∎

Proposition 3.2. (Semi-definite programming for barycentric quantum Rényi leakage). *The barycentric quantum Rényi leakage from random variable X through the quantum encoding of the data via ensemble* $\mathcal{E}=\{p_X(x),\rho_A^x\}_{x\in\mathbb{X}}$ *can be computed using the semi-definite programming*:

$$\min_{\pi\in\Delta(\mathbb{X}),\mu\in\mathbb{R}} \quad \mu,$$
$$\text{s.t.} \quad \rho_A^x\leq\mu\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'},\ \forall x\in\mathbb{X}.$$

Proof. The proof immediately follows from the definition of $\widetilde{D}_\infty$. ∎

Proposition 3.2 shows that, although the barycentric quantum Rényi leakage still does not possess an explicit formula, its computation is far simpler than the maximal quantum leakage (cf [6]). It should however be noted that the computational cost of solving the semi-definite program in proposition 3.2 grows in a polynomial manner with the dimension of the matrices [30], which in turn grows exponentially with the number of qubits used to encode the classical data.

Example 3.1. (Basis/index encoding). *Consider a random variable X with a support set* $\mathbb{X}=\{1,\ldots,2^n\}$ *for some positive integer n. Assume that a basis or index encoding strategy is used, that is,* $\rho_A^x=|x\rangle\langle x|$ *for all* $x\in\mathbb{X}$. *In* [6], *it was shown that* $\mathcal{Q}(X\to A)_{\rho_A}=n$. *To compute the barycentric quantum Rényi leakage, we can use theorem 3.2 to demonstrate that* $\mathcal{B}(X\to A)_{\rho_A}=\min_{\pi\in\Delta(\mathbb{X})}\max_{x\in\mathbb{X}}(1/\pi(x))=n$. *This demonstrates that the upper bound in corollary 3.1 can be tight in some cases.*

In what follows, we further simplify the upper bound in proposition 3.1 to drive a simpler measure of information leakage with an explicit form.

Proposition 3.3. *The barycentric quantum Rényi leakage from random variable X through the quantum encoding of the data via ensemble* $\mathcal{E}=\{p_X(x),\rho_A^x\}_{x\in\mathbb{X}}$ *is upper bounded by*

$$\mathcal{B}(X\to A)_{\rho_A}\leq\max_{x,x'\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A^{x'}).$$

Proof. The proof follows from the quasi-convexity of the max-relative entropy because $\widetilde{D}_\infty(\rho_A^x\|\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'})=\widetilde{D}_\infty(\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^x\|\sum_{x'\in\mathbb{X}}\pi(x')\rho_A^{x'})\leq\max_{x'\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A^{x'})$; see lemma 2.2. ∎

This upper bound suggests introducing another measure of information leakage for quantum encoding of classical data, referred to as pairwise quantum Rényi leakage. This notion of information leakage possesses an explicit formula in terms of the max-relative entropy in quantum information theory and is hence more amenable to theoretical analysis (see the subsequent two sections for examples of such analysis); however, as we demonstrate later, it is more conservative.

Definition 3.3. (Pairwise quantum Rényi leakage). *The pairwise quantum Rényi leakage from random variable X through the quantum encoding of the data via ensemble* $\mathcal{E}=\{p_X(x),\rho_A^x\}_{x\in\mathbb{X}}$ *is*

$$\mathcal{R}(X\to A)_{\rho_A}:=\max_{x,x'\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A^{x'}). \tag{3.6}$$

The following corollary immediately follows from corollary 3.1, proposition 3.3 and definition 3.3.

Corollary 3.2. $\mathcal{Q}(X \to A)_{\rho_A} \le \mathcal{B}(X \to A)_{\rho_A} \le \mathcal{R}(X \to A)_{\rho_A}$.

Similarly, the pairwise quantum Rényi leakage satisfies important properties of positivity, independence, post-processing inequality and unitary invariance. These properties are established in the following theorem.

Theorem 3.2. *The following properties hold for the pairwise quantum Rényi leakage*:

(a) *Positivity and Independence*: $\mathcal{R}(X \to A)_{\rho_A} \ge 0$ *with equality if and only if* $\rho_A^x = \rho_A^{x'}$ *for all* $x, x' \in \mathbb{X}$;
(b) *Unitary Invariance*: $\mathcal{R}(X \to A)_{U\rho_A U^\dagger} = \mathcal{R}(X \to A)_{\rho_A}$ *for any unitary* U;
(c) *Data-Processing Inequality*: $\mathcal{R}(X \to A)_{\mathcal{E}(\rho_A)} \le \mathcal{R}(X \to A)_{\rho_A}$ *for any quantum channel* $\mathcal{E}$.

Proof. We start by proving part (a). Note that $\mathcal{R}(X \to A)_{\rho_A} = \max_{x,x' \in \mathbb{X}} \widetilde{D}_\infty(\rho_A^x \| \rho_A^{x'}) \ge \widetilde{D}_\infty(\rho_A^x \| \rho_A^x) = 0$. If $\rho_A^x = \rho_A^{x'}$ for all $x, x' \in \mathbb{X}$, $\widetilde{D}_\infty(\rho_A^x \| \rho_A^{x'}) = 0$ for all $x', x \in \mathbb{X}$. Therefore, $\mathcal{R}(X \to A)_{\rho_A} = 0$. On the other hand, if $\mathcal{R}(X \to A)_{\rho_A} = 0$, it means that, for all $x, x' \in \mathbb{X}$, $\widetilde{D}_\infty(\rho_A^x \| \rho_A^{x'}) = 0$, which is only possible if $\rho_A^x = \rho_A^{x'}$.

The proof for part (b) is similar to the proof of part (b) of theorem 3.1 because lemma 2.1 results in $\widetilde{D}_\infty(U\rho_A^x U^\dagger \| U\rho_A^{x'} U^\dagger) = \widetilde{D}_\infty(\rho_A^x \| \rho_A^{x'})$ for all $x, x' \in \mathbb{X}$.

The proof of part (c) follows from that $\widetilde{D}_\infty(\mathcal{E}(\rho_A^x) \| \mathcal{E}(\rho_A^{x'})) \le \widetilde{D}_\infty(\rho_A^x \| \rho_A^{x'})$, where the inequality is a consequence of lemma 2.1. ∎

Example 3.1. (Basis/index encoding (continued)). *Note that, because* $\rho_A^x \not\ll \rho_A^{x'}$ *if* $x \ne x'$, *we can easily see that* $\mathcal{R}(X \to A)_{\rho_A} = \max_{x,x' \in \mathbb{X}} \widetilde{D}_\infty(\rho_A^x \| \rho_A^{x'}) = \infty$. *Therefore, the upper bound relating to* $\mathcal{R}(X \to A)_{\rho_A}$ *in corollary 3.2 can be loose in this instance. This measure of information leakage can be conservative in general.*

We finish this section by investigating the relationship between the barycentric and pairwise quantum Rényi leakage with the sandwiched quantum α-mutual information and accessible information. Consider the following classical-quantum state representing the ensemble $\mathcal{E} = \{p_X(x), \rho_A^x\}_{x \in \mathbb{X}}$:

$$\rho_{XA} = \sum_{x \in \mathbb{X}} p_X(x) |x\rangle\langle x| \otimes \rho_A^x.$$

Recall that the sandwiched quantum α-mutual information of the classical-quantum state ρ_{XA} is given by $\widetilde{I}_\alpha(X; A)_{\rho_{XA}} = \inf_\sigma \widetilde{D}_\alpha(\rho_{XA} \| \rho_X \otimes \sigma)$, where $\rho_X = \mathrm{tr}_A(\rho_{XA})$.

Proposition 3.4. *The sandwiched quantum α-mutual information of the classical-quantum state* ρ_{XA} *is upper bounded by*

$$\widetilde{I}_\alpha(X; A)_{\rho_{XA}} \le \widetilde{I}_\infty(X; A)_{\rho_{XA}} \le \mathcal{B}(X \to A)_{\rho_A} \le \mathcal{R}(X \to A)_{\rho_A}.$$

Proof. First, $\widetilde{I}_\alpha(X; A)_{\rho_{XA}} \le \widetilde{I}_\infty(X; A)_{\rho_{XA}}$ is a direct consequence of in [10, theorem 7]. Now, note that

$$\begin{aligned}
\widetilde{D}_\infty(\rho_{XA} \| \rho_X \otimes \sigma) &= \widetilde{D}_\infty\left(\sum_{x \in \mathbb{X}} p_X(x)|x\rangle\langle x| \otimes \rho_A^x \Big\| \sum_{x \in \mathbb{X}} p_X(x)|x\rangle\langle x| \otimes \sigma\right) \\
&\le \max_{x \in \mathbb{X}} \widetilde{D}_\infty(|x\rangle\langle x| \otimes \rho_A^x \| |x\rangle\langle x| \otimes \sigma) \\
&\le \max_{x \in \mathbb{X}} \widetilde{D}_\infty(\rho_A^x \| \sigma),
\end{aligned}$$

where the first inequality follows from the quasi convexity of the max-relative entropy in lemma 2.2 and the second inequality is the post-processing inequality in lemma 2.1, i.e. $\widetilde{D}_\infty(\mathcal{N}(\rho_A^x)\|\mathcal{N}(\sigma)) \leq \widetilde{D}_\infty(\rho_A^x\|\sigma)$ with quantum channel $\mathcal{N}(\rho) = |x\rangle\langle x| \otimes \rho$ for any $x \in \mathbb{X}$. An alternative proof for the second inequality can be derived by using the additivity axiom [10, theorem 2]. Therefore,

$$\begin{aligned}\widetilde{I}_\infty(X;A)_{\rho_{XA}} &= \inf_\sigma \widetilde{D}_\infty(\rho_{XA}\|\rho_X \otimes \sigma)\\ &\leq \inf_\sigma \max_{x\in\mathbb{X}} \widetilde{D}_\infty(\rho_A^x\|\sigma)\\ &\leq \inf_{\pi\in\Delta(\mathbb{X})} \max_{x\in\mathbb{X}} \widetilde{D}_\infty\left(\rho_A^x \middle\| \sum_{x'\in\mathbb{X}} \pi(x')\rho_A^{x'}\right).\end{aligned}$$

∎

This concludes the proof.

An important notion of information in quantum information theory is accessible information [22, p. 298]. For ensemble $\mathcal{E} = \{p_X(x), \rho_A^x\}_{x\in\mathbb{X}}$, the accessible information is

$$I_{\mathrm{acc}}(\mathcal{E}) := \sup_{\{F_y\}_y} I_1(X;Y).$$

In addition, the Holevo's information [31] (also see [22, p. 318]) is

$$\chi(\mathcal{E}) := I_1(X;A)_{\rho_A} = H(\rho_A) - \sum_{x\in\mathbb{X}} p_X(x)H(\rho_A^x),$$

where $H(\rho) = -\mathrm{tr}(\rho\log(\rho))$ is the von Neumann quantum entropy. The next proposition provides a relationship between accessible information, and the barycentric and pairwise quantum Rényi leakage.

Proposition 3.5. $I_{acc}(\mathcal{E}) \leq \chi(\mathcal{E}) \leq \mathcal{B}(X \to A)_\rho \leq \mathcal{R}(X \to A)_\rho$.

Proof. First, note that the Holevo bound demonstrates that $I_{\mathrm{acc}}(\mathcal{E}) \leq I_1(X;A)_{\rho_{XA}}$ [31]. Furthermore, $I_1(X;A)_{\rho_{XA}} = \widetilde{I}_1(X;A)_{\rho_{XA}}$. In addition, $\widetilde{I}_1(X;A)_{\rho_{XA}} \leq \widetilde{I}_\infty(X;A)_{\rho_{XA}}$ [10, theorem 7]. The rest follows from proposition 3.4. ∎

4. Quantum differential privacy and depolarizing channels

Differential privacy is the gold standard of data privacy analysis in the computer science literature [32]. Differential privacy has been recently extended to quantum computing [12,33,34]. The aim of differential privacy in quantum computing is to ensure that an adversary e.g. eavesdropper, cannot distinguish between two 'similar' datasets based on measurements of the underlying quantum system, cf hypothesis-testing privacy [35]. Similarity is modelled using a neighbourhood relationship.

Definition 4.1. (Neighbouring relationship). *A neighbouring relationship (over the set of density operators) is a mathematical relation $\sim$ that is both reflective ($\rho \sim \rho$ for all density operators ρ) and symmetric ($\rho \sim \sigma$ implies $\sigma \sim \rho$ for any two density operators ρ, σ).*

An example of a neighbouring relationship can be defined using the trace distance over quantum density operators, i.e. $\rho \sim \sigma$ if $\|\rho - \sigma\|_1 \leq \kappa$ for some constant $\kappa > 0$ [33]. This is formalized in the following definition.

Definition 4.2. (Closeness neighbouring relationship). $\rho \sim \sigma$ if $\|\rho - \sigma\|_1 \leq \kappa$.

Note that other neighbouring relationships can be embedded in differential privacy. For instance, two quantum density operators can be neighbouring if they encode two private datasets that differ in the data of one individual [32].

royalsocietypublishing.org/journal/rspa *Proc. R. Soc. A* **480:** 20240319

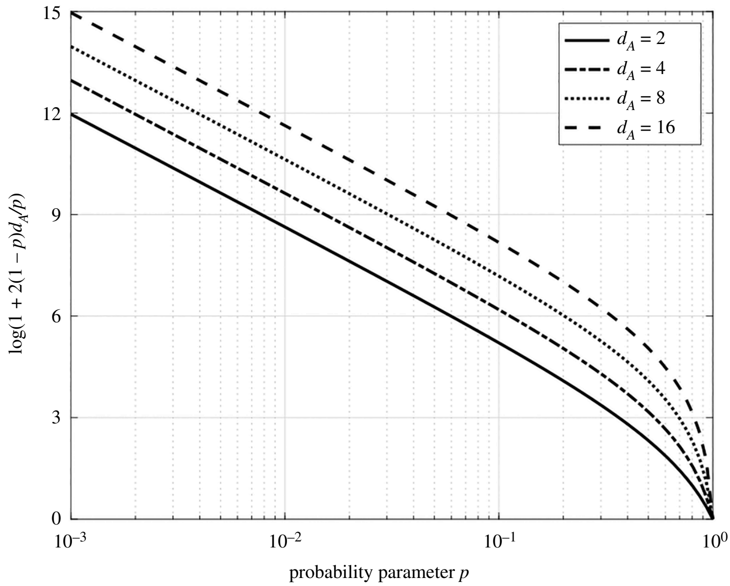

Figure 1. The upper bound for $\mathcal{B}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)}$ versus probability parameter p.

Definition 4.3. *For any $\epsilon, \delta \geq 0$, a quantum channel $\mathcal{E}$ is (ϵ, δ)-differentially private if*

$$\mathrm{tr}(M\mathcal{E}(\rho)) \leq \exp(\epsilon)\mathrm{tr}(M\mathcal{E}(\sigma)) + \delta, \tag{4.1}$$

for all measurements $0 \leq M \leq I$ and neighbouring density operators $\rho \sim \sigma$.

An interesting concept is to measure the effect of the quantum differential privacy on information leakage.

Proposition 4.1. *$\mathcal{B}(X \to A)_{\mathcal{E}(\rho_A)} \leq \mathcal{R}(X \to A)_{\mathcal{E}(\rho_A)} \leq \epsilon/\ln(2)$ if $\rho_A^x \sim \rho_A^{x'}$ for all $x, x' \in \mathbb{X}$ under any neighbouring relationship and the quantum channel $\mathcal{E}$ is $(\epsilon, 0)$-differentially private.*

Proof. From [12, lemma III.2] (with $\delta = 0$), we get $\widetilde{D}_\infty(\mathcal{E}(\rho_A^x)\|\mathcal{E}(\rho_A^{x'})) \leq \epsilon/\ln(2)$. Note that the division by $\ln(2)$ is caused by the use of natural basis in the definition of the differential privacy and logarithms in [12]. This shows that $\mathcal{R}(X \to A)_{\rho_A} \leq \epsilon/\ln(2)$. The rest follows from corollary 3.2. ∎

A physical noise model for quantum systems is the global depolarizing channel:

$$\mathcal{D}_{p,d_A}(\rho) := \frac{p}{d_A}I + (1-p)\rho, \tag{4.2}$$

where d_A is the dimension of the Hilbert space $\mathcal{H}_A$ to which the system belongs and $p \in [0,1]$ is a probability parameter. The larger the probability parameter p, the noisier the global depolarizing channel $\mathcal{D}_{p,d_A}$.

Proposition 4.2. $\mathcal{B}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)} \leq \mathcal{R}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)} \leq \log(1 + 2(1-p)d_A/p)$.

Proof. Reference [12, lemma IV.2] demonstrates that $\mathcal{D}_{p,d_A}$ is $(\epsilon, 0)$-differentially private with $\epsilon = \ln(1 + 2(1-p)d_A/p)$. Note that, here, we set $\varkappa = 2$ as always $\|\rho_A^x - \rho_A^{x'}\| \leq 2$ for any two density operators ρ_A^x and $\rho_A^{x'}$. Finally, note that $\ln(1 + 2(1-p)d_A/p)/\ln(2) = \log(1 + 2(1-p)d_A/p)$. The rest follows from propositions 3.5 and 4.1. ∎

Figure 1 illustrates the upper bound in Proposition 4.2 versus probability parameter p. This clearly demonstrates that by increasing p, the amount of leaked information, measured by both the barycentric and pairwise quantum Rényi leakage, tends to zero. This is established for the global depolarizing channels. However, in quantum computing devices, each qubit can be affected by local noise. To model this case, assume that the Hilbert space $\mathcal{H}_A$ is composed of k qubits so that $d_A = 2^k$. The local depolarizing noise channel is defined as

$$\mathcal{D}_{p,2}^{\otimes k} := \mathcal{D}_{p,2} \otimes \cdots \otimes \mathcal{D}_{p,2}, \tag{4.3}$$

where a depolarizing channel $\mathcal{D}_{p,2}$ acts locally on each qubit.

Proposition 4.3. $\mathcal{B}(X \to A)_{\mathcal{D}_{p,2}^{\otimes k}(\rho_A)} \leq \mathcal{R}(X \to A)_{\mathcal{D}_{p,2}^{\otimes k}(\rho_A)} \leq \log(1 + 2(1 - p^k)d_A/p^k)$.

Proof. The proof is similar to that of proposition 4.2 with the exception of relying on corollary IV.3 in [12]. ∎

5. Quantum machine learning

In this section, we consider variational circuits for implementing quantum machine learning and analysing privacy-utility trade-off in these models. The first step in a variational circuit is an encoding layer that transforms the classical data, i.e. input of the quantum machine learning model, into a quantum state. In the notation of §3, the ensemble $\mathcal{E} := \{p_X(x), \rho_A^x\}_{x \in \mathbb{X}}$ is used to model this layer. The next layer is a variational unitary U_θ with tunable parameter θ. After this layer, the state of the quantum system is $U_\theta \rho_A^x U_\theta^\dagger$. Finally, measurements are taken to determine the output label. The measurement can be modelled by POVM $O = \{O_c\}$, where the outcome c denotes the class to which the input belongs. Figure 2 illustrates a variational quantum machine learning model with encoding unitary V_x and variational circuits U_θ. Here, the quantum states are initialized at $|0\rangle \otimes \cdots \otimes |0\rangle$ and, therefore, we have $\rho_A^x = V_x|0\rangle \otimes \cdots \otimes |0\rangle\langle 0| \otimes \cdots \otimes \langle 0|V_x^\dagger$. Training the quantum machine learning model entails finding parameters θ, e.g. by gradient descent, to minimize the prediction error based on a training dataset. To achieve private machine learning, we can add a global or local depolarizing channel to ensure differential privacy or to bound the information leakage (in the language of this paper). The performance degradation caused by the quantum channel $\mathcal{E}$ is

$$\Gamma(\mathcal{E}) := \max_{x \in \mathbb{X}} \sum_c |\mathrm{tr}(O_c U_\theta \rho_A^x U_\theta^\dagger) - \mathrm{tr}\,(O_c \mathcal{E}\,(U_\theta \rho_A^x U_\theta^\dagger))\,|, \tag{5.1}$$

which captures the changes in the probability of reporting each class by the addition of the quantum channel $\mathcal{E}$. The following proposition provides an upper bound for the performance degradation caused by the global depolarizing channel.

Proposition 5.1. $\Gamma(\mathcal{D}_{p,d_A}) \leq 2p(d_A - 1)/d_A$.

Proof. Note that

$$\begin{aligned}\sum_c |\,\mathrm{tr}\,(O_c U_\theta \rho_A^x U_\theta^\dagger) - \mathrm{tr}\,(O_c \mathcal{D}_{p,d_A}(U_\theta \rho_A^x U_\theta^\dagger))| &\leq \|U_\theta \rho_A^x U_\theta^\dagger - \mathcal{D}_{p,d_A}(U_\theta \rho_A^x U_\theta^\dagger)\|_1 \\ &\leq \sup_{\sigma \in \mathcal{S}(\mathcal{H}_A)} \|\sigma - \mathcal{D}_{p,d_A}(\sigma)\|_1,\end{aligned}$$

where the first inequality follows from [22, exercise 9.1.10, pp. 239] and the second inequality is a consequence of maximizing over the state. For any $\sigma \in \mathcal{S}(\mathcal{H}_A)$, spectral decomposition can be used to write $\sigma = \sum_{z \in \mathbb{Z}} \lambda(z)|z\rangle\langle z|$ with $\lambda \in \Delta(\mathbb{Z})$ and orthonormal basis $\{|z\rangle\}_{z \in \mathbb{Z}}$ for $\mathcal{H}_A$ (so $|\mathbb{Z}| = d_A$). Therefore, we have

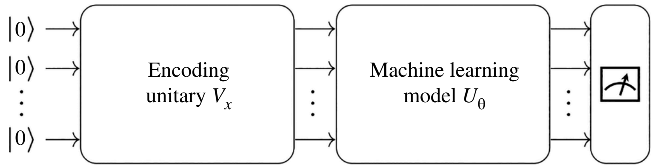

Figure 2. A variational quantum machine learning model with encoding unitary V_x and variational circuit U_θ.

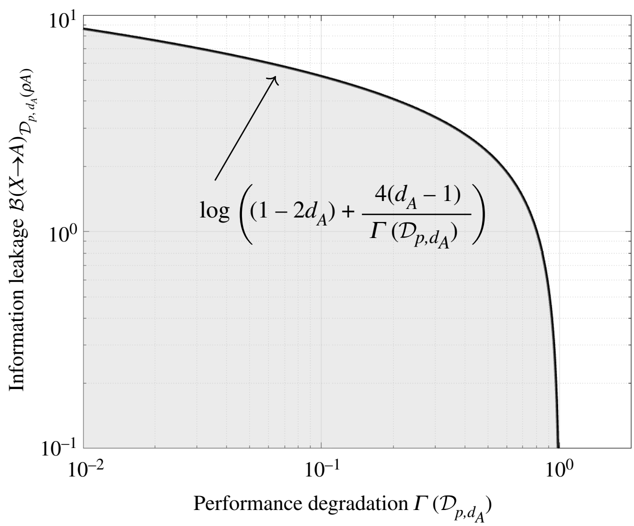

Figure 3. Privacy-utility trade-off region when using the global depolarizing channel for $d_A = 2$. The solid curve depicts the upper bound in corollary 5.1.

$$
\begin{aligned}
\|\sigma - \mathcal{D}_{p,d_A}(\sigma)\|_1 &= \left\|\sigma - \frac{p}{d_A}I - (1-p)\sigma\right\|_1 \\
&= p\left\|\sigma - \frac{1}{d_A}I\right\|_1 \\
&= p\left\|\sum_{z\in\mathbb{Z}}\lambda(z)|z\rangle\langle z| - \sum_{z\in\mathbb{Z}}\lambda(z)\frac{1}{d_A}I\right\|_1 \\
&\le p\sum_{z\in\mathbb{Z}}\lambda(z)\left\||z\rangle\langle z| - \frac{1}{d_A}I\right\|_1 \\
&= p\sum_{z\in\mathbb{Z}}\lambda(z)\left\|\left(1-\frac{1}{d_A}\right)|z\rangle\langle z| - \frac{1}{d_A}\sum_{z'\in\mathbb{Z}\setminus\{z\}}|z'\rangle\langle z'|\right\|_1 \\
&= p\sum_{z\in\mathbb{Z}}\lambda(z)\left(\left(1-\frac{1}{d_A}\right) + \frac{d_A-1}{d_A}\right) \\
&= p\sum_{z\in\mathbb{Z}}\lambda(z)\left(1+\frac{d_A-2}{d_A}\right) \\
&= \frac{2p(d_A-1)}{d_A},
\end{aligned}
$$

where the inequality stems from the convexity of the trace norm [22, exercise 9.1.11, p. 239]. This concludes the proof. ∎

Corollary 5.1. *The following privacy-utility trade-off holds when using the global depolarizing channel*:

$$\mathcal{B}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)} \le \mathcal{R}(X \to A)_{\mathcal{D}_{p,d_A}(\rho_A)} \le \log\left((1-2d_A) + \frac{4(d_A-1)}{\Gamma(\mathcal{D}_{p,d_A})}\right).$$

Proof. Proposition 5.1 shows that $1/p \le 2p(d_A-1)/(d_A\Gamma(\mathcal{D}_{p,d_A}))$. Combining this with the inequality in proposition 4.2 finishes the proof. ∎

Figure 3 illustrates the privacy-utility trade-off region when using the global depolarizing channel for $d_A = 2$. The solid curve depicts the upper bound in corollary 5.1. Note that, as expected, the privacy guarantees can only be straightened, i.e. information leakage is reduced, if the performance degradation is larger. Recently, the utility-privacy trade-off in quantum machine learning was also considered in [13]; however, in that paper, different notions of utility (i.e. diamond distance) and privacy (i.e. quantum pufferfish privacy) were considered. Note that here we are relating utility to an operational notion of information leakage, that is, maximal quantum leakage, through barycentric and pairwise quantum Rényi leakages.

6. Conclusions and future work

Two new measures of information leakage for security and privacy analysis against arbitrary eavesdroppers, i.e. adversaries whose intention is not known by the analyst, are proposed. They satisfy important properties of positivity, independence, post-processing inequality and unitary invariance. They can also be computed easily. Differentially private quantum channels are shown to bound these new notions of information leakage. Finally, the fundamental problem of privacy-utility trade-off in quantum machine learning models was analysed using the proposed notions of information leakage. Future work can focus on developing optimal privacy-preserving policies by minimizing the information leakage subject to a constraint on the utility. This approach is widely used in classical data privacy literature when using information-theoretic notions for security and privacy analysis. Furthermore, these measures can be used for the analysis of wiretap or cipher channels, where the eavesdropper is generalized (not interested in estimating the entire secret data but can also focus on part data recovery).

Data accessibility. This article has no additional data.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. F.F.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing.

Conflict of interest declaration. We declare we have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. The author is thankful to Theshani Nuradha for bringing [13] to his attention and providing useful comments and suggestions for improving the current paper. The author would like to also thank the anonymous reviewers for finding a minor mistake in the proof of theorem 3.1, shortening the proofs of propositions 3.3 and 3.4 by appealing to quasi-convexity and post-processing inequality, and tightening the results of proposition 5.1 and corollary 5.1 by developing a bound that relies on the dimension of the quantum system.

Appendix

A. Technical lemma

Lemma A.1. $\lim_{\alpha\to\infty} \max_{x\in\mathbb{X}} \widetilde{D}_\alpha(\rho_A^x \| \rho_A) = \max_{x\in\mathbb{X}} \lim_{\alpha\to\infty} \widetilde{D}_\alpha(\rho_A^x \| \rho_A)$.

Proof. Note that $\lim_{\alpha\to\infty}\widetilde{D}_\alpha(\rho_A^x\|\rho_A)=\widetilde{D}_\infty(\rho_A^x\|\rho_A)$, which is finite because $\rho_A^x\ll\rho_A$ by definition. Therefore, for all $\alpha>0$, there exists $\bar{\alpha}_{x,\epsilon}$ such that $|\widetilde{D}_\alpha(\rho_A^x\|\rho_A)-\widetilde{D}_\infty(\rho_A^x\|\rho_A)|\leq\epsilon$ for all $\alpha\geq\bar{\alpha}_{x,\epsilon}$. This implies that

$$\widetilde{D}_\infty(\rho_A^x\|\rho_A)-\epsilon\leq\widetilde{D}_\alpha(\rho_A^x\|\rho_A)\leq\widetilde{D}_\infty(\rho_A^x\|\rho_A)+\epsilon,\forall\alpha\geq\bar{\alpha}_{x,\epsilon},\forall x\in\mathbb{X}. \tag{A 1}$$

Based on this inequality, we can prove that

$$\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A)-\epsilon\leq\max_{x\in\mathbb{X}}\widetilde{D}_\alpha(\rho_A^x\|\rho_A)\leq\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A)+\epsilon,\forall\alpha\geq\max_{x\in\mathbb{X}}\bar{\alpha}_{x,\epsilon}. \tag{A 2}$$

The proof for this come from a contrapositive argument. For proving the upper bound, assume that there exists $\alpha\geq\max_{x\in\mathbb{X}}\bar{\alpha}_{x,\epsilon}$ such that $\max_{x\in\mathbb{X}}\widetilde{D}_\alpha(\rho_A^x\|\rho_A)>\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A)+\epsilon$. Therefore, there exists $x'\in\mathbb{X}$ such that $\widetilde{D}_\alpha(\rho_A^{x'}\|\rho_A)>\widetilde{D}_\infty(\rho_A^x\|\rho_A)+\epsilon$ for all $x\in\mathbb{X}$. This implies that $\widetilde{D}_\alpha(\rho_A^{x'}\|\rho_A)>\widetilde{D}_\infty(\rho_A^{x'}\|\rho_A)+\epsilon$, which is in contradiction with equation (A 1). For proving the lower bound, assume that there exists $\alpha\geq\max_{x\in\mathbb{X}}\bar{\alpha}_{x,\epsilon}$ such that $\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A)-\epsilon>\max_{x\in\mathbb{X}}\widetilde{D}_\alpha(\rho_A^x\|\rho_A)$. Therefore, there exists $x'\in\mathbb{X}$ such that $\widetilde{D}_\infty(\rho_A^{x'}\|\rho_A)-\epsilon>\widetilde{D}_\alpha(\rho_A^x\|\rho_A)$ for all $x\in\mathbb{X}$. This implies that $\widetilde{D}_\infty(\rho_A^{x'}\|\rho_A)-\epsilon>\widetilde{D}_\alpha(\rho_A^{x'}\|\rho_A)$, which is in contradiction with equation (A 1). Equation (A 2) shows that, for all $\epsilon>0$, there exists $\bar{\alpha}_\epsilon:=\max_{x\in\mathbb{X}}\bar{\alpha}_{x,\epsilon}$ such that

$$|\max_{x\in\mathbb{X}}\widetilde{D}_\infty(\rho_A^x\|\rho_A)-\max_{x\in\mathbb{X}}\widetilde{D}_\alpha(\rho_A^x\|\rho_A)|\leq\epsilon,\forall\alpha\geq\bar{\alpha}_\epsilon.$$

In addition, $\bar{\alpha}_\epsilon=\max_{x\in\mathbb{X}}\bar{\alpha}_{x,\epsilon}<\infty$ because $\mathbb{X}$ is finite. This concludes the proof. ∎

References

1. Shor PW. Algorithms for quantum computation: discrete logarithms and factoring. In *35th Annual Symposium on Foundations of Computer Science*, pp. 124–134. IEEE. (doi:10.1109/SFCS.1994.365700)
2. Shor PW, Preskill J. 2000 Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.* **85**, 441–444. (doi:10.1103/PhysRevLett.85.441)
3. West MT, Tsang SL, Low JS, Hill CD, Leckie C, Hollenberg LCL, Erfani SM, Usman M. 2023 Towards quantum enhanced adversarial robustness in machine learning. *Nat. Mach. Intell.* **5**, 581–589. (doi:10.1038/s42256-023-00661-1)
4. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. 2017 Quantum machine learning. *Nature* **549**, 195–202. (doi:10.1038/nature23474)
5. Kearns M, Roth A. 2019 *The ethical algorithm: the science of socially aware algorithm design*. New York, NY: Oxford University Press.
6. Farokhi F. 2024 Maximal information leakage from quantum encoding of classical data. *Phys. Rev. A* **109**, 022608. (doi:10.1103/PhysRevA.109.022608)
7. Issa I, Wagner AB, Kamath S. 2019 An operational approach to information leakage. *IEEE Trans. Inf. Theory* **66**, 1625–1657. (doi:10.1109/TIT.2019.2962804)
8. Farokhi F, Ding N. 2021 Measuring information leakage in non-stochastic brute-force guessing. In *2020 IEEE Information Theory Workshop (ITW)*, pp. 1–5. (doi:10.1109/ITW46852.2021.9457645)
9. Řeháček J, Englert BG, Kaszlikowski D. 2005 Iterative procedure for computing accessible information in quantum communication. *Phys. Rev. A* **71**, 054303. (doi:10.1103/PhysRevA.71.054303)
10. Müller-Lennert M, Dupuis F, Szehr O, Fehr S, Tomamichel M. 2013 On quantum rényi entropies: a new generalization and some properties. *J. Math. Phys.* **54**, 12. (doi:10.1063/1.4838856)
11. Audenaert KMR, Datta N. 2015 α-z-Rényi relative entropies. *J. Math. Phys.* **56**. (doi:10.1063/1.4906367)

12. Hirche C, Rouzé C, França DS. 2023 Quantum differential privacy: an information theory perspective. *IEEE Trans. Inf. Theory* **69**, 5771–5787. (doi:10.1109/TIT.2023.3272904)
13. Nuradha T, Goldfeld Z, Wilde MM. 2023 Quantum pufferfish privacy: a flexible privacy framework for quantum systems. *IEEE Trans. Inf. Theory* **70**, 5731–5762. (doi:10.1109/TIT.2024.3404927)
14. Nuradha T, Wilde MM. 2024 Contraction of private quantum channels and private quantum hypothesis testing. *arXiv*. (doi:10.48550/arXiv.2406.18651)
15. Cheng HC, Hirche C, Rouzé C. 2024 Sample complexity of locally differentially private quantum hypothesis testing. *arXiv*. (doi:10.48550/arXiv.2406.18658)
16. Watkins WM, Chen SYC, Yoo S. 2023 Quantum machine learning with differential privacy. *Sci. Rep.* **13**, 2453. (doi:10.1038/s41598-022-24082-z)
17. Watkins W, Wang H, Bae S, Tseng HH, Cha J, Chen SYC, Yoo S. 2024 Quantum Privacy Aggregation of Teacher Ensembles (QPATE) for Privacy Preserving Quantum Machine Learning. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6875–6879. (doi:10.1109/ICASSP48485.2024.10447786)
18. Rényi A. 1961 On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, vol. **4**, pp. 547–562,
19. van Erven T, Harremoes P. 2014 Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. Inf. Theory* **60**, 3797–3820. (doi:10.1109/TIT.2014.2320500)
20. Thomas JA, Cover TM. 2012 *Elements of information theory*. Hoboken, NJ: Wiley.
21. Verdú S. 2015 α-mutual information. In *2015 Information Theory and Applications Workshop (ITA)*, pp. 1–6. IEEEE. (doi:10.1109/ITA.2015.7308959)
22. Wilde M. 2013 *Quantum information theory*. Cambridge, UK: Cambridge University Press.
23. Petz D. 1986 Quasi-entropies for finite quantum systems. *Rep. Math. Phys.* **23**, 57–65. (doi:10.1016/0034-4877(86)90067-4)
24. Umegaki H. 1962 Conditional expectation in an operator algebra, IV (entropy and information). In *Kodai Mathematical Seminar Reports*, vol. **14**, pp. 59–85, Department of Mathematics, Tokyo Institute of Technology. (doi:10.2996/kmj/1138844604)
25. Bhatia R. 2013 *Matrix analysis. graduate texts in mathematics*. New York, NY: Springer New York.
26. Datta N. 2009 Min- and max-relative entropies and a new entanglement monotone. *IEEE Trans. Inf. Theory* **55**, 2816–2826. (doi:10.1109/TIT.2009.2018325)
27. Bussandri DG, Rajchel-Mieldzioć G, Lamberti PW, Życzkowski K. 2024 Rényi–Holevo inequality from α-z-Rényi relative entropies. *Int. J. Quantum Inform.* **22**. (doi:10.1142/S021974992440015X)
28. Mosonyi M, Bunth G, Vrana P. 2022 Geometric relative entropies and barycentric Rényi divergences. *arXiv*. (doi:10.48550/arXiv.2207.14282)
29. Mosonyi M, Ogawa T. 2017 Strong converse exponent for classical-quantum channel coding. *Commun. Math. Phys.* **355**, 373–426. (doi:10.1007/s00220-017-2928-4)
30. Vandenberghe L, Balakrishnan VR, Wallin R, Hansson A, Roh T. 2005 Interior-point algorithms for semidefinite programming problems derived from the KYP lemma. *Posit. Polynom. Control* 195–238. (doi:10.1007/10997703_12)
31. Holevo AS. 1973 Bounds for the quantity of information transmitted by a quantum communication channel. *Probl. Pered. Inform.* **9**, 3–11.
32. Dwork C. 2008 Differential privacy: a survey of results. In *International Conference on Theory and Applications of Models of Computation*, pp. 1–19. Springer. (doi:10.1007/978-3-540-79228-4_1)
33. Zhou L, Ying M. 2017 Differential privacy in quantum computation. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, IEEE, pp. 249–262. (doi:10.1109/CSF.2017.23)
34. Aaronson S, Rothblum GN. 2019 Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 322–333. New York. (doi:10.1145/3313276.3316378)
35. Farokhi F. 2023 Quantum privacy and hypothesis-testing. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 2841–2846. New York, NY: IEEE. (doi:10.1109/CDC49753.2023.10384192)