Lipschitz continuity of quantum-classical conditional entropies with respect to angular distance and related properties

Michael Liaofan Liu,[1,2,*] Florian Kanitschar,[1,3] Amir Arqand,[1] and Ernest Y.-Z. Tan[1]

[1] *Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[2] *Department of Mathematics, Amherst College, Amherst, Massachusetts 01002, USA*
[3] *Technische Universität Wien, Faculty of Mathematics and Geoinformation, Wiedner Hauptstraße 8, 1040 Vienna, Austria*

(Received 22 October 2022; accepted 7 February 2023; published 21 February 2023)

We derive a Lipschitz continuity bound for quantum-classical conditional entropies with respect to angular distance, with a Lipschitz constant that is independent of the dimension of the conditioning system. This bound is sharper in some situations than previous continuity bounds, which were either based on trace distance (where Lipschitz continuity is not possible), or based on angular distance but did not include a conditioning system. However, we find that the bound does not directly generalize to fully quantum conditional entropies. To investigate possible counterexamples in that setting, we study the characterization of states which saturate the Fuchs–van de Graaf inequality and thus have angular distance approximately equal to trace distance. We give an exact characterization of such states in the invertible case. For the noninvertible case, we show that the situation appears to be significantly more elaborate, and seems to be strongly connected to the question of characterizing the set of fidelity-preserving measurements.

DOI: 10.1103/PhysRevA.107.022426

I. INTRODUCTION

Given two quantum states ρ and σ on a Hilbert space $\mathcal{H}$, one of the most natural questions to ask is how similar ρ and σ are. Common measures to answer this question include the trace distance,

$$\mathrm{T}(\rho,\sigma) := \tfrac{1}{2}\|\rho-\sigma\|_1, \tag{1}$$

and the (root-)fidelity,

$$\mathrm{F}(\rho,\sigma) := \|\sqrt{\rho}\,\sqrt{\sigma}\|_1. \tag{2}$$

The trace distance is a metric on the set of density operators $\mathcal{D}(\mathcal{H})$, and it has a meaningful interpretation as the distinguishability of two quantum states. In a quantum hypothesis testing scenario, where Bob randomly prepares one of two states ρ and σ (with equal probability) for Alice to distinguish, Alice can correctly identify the incoming state with probability $[1+\mathrm{T}(\rho,\sigma)]/2$. In contrast, the fidelity is not a metric, but it can be interpreted as the probability that a state ρ "passes a test" for being the same as a pure state σ [1].

Another important task in quantum information theory is to quantify the amount of information present in a quantum system. The von Neumann entropy

$$\mathrm{H}(\rho) := -\operatorname{tr}(\rho\ln\rho)$$

is one quantity which fulfills this role[1] because it appears in many fundamental information theoretic tasks such as Schumacher data compression [2] and randomness extraction [3]. This concept can be extended to conditional entropies $\mathrm{H}(A|B)_\rho$ for bipartite states $\rho := \rho_{AB} \in \mathcal{D}(\mathcal{H}^A\otimes\mathcal{H}^B)$, with one of several equivalent definitions being the difference between the joint entropy and the marginal entropy,

$$\mathrm{H}(A|B)_\rho := \mathrm{H}(\rho_{AB}) - \mathrm{H}(\rho_B), \tag{3}$$

where $\rho_B := \operatorname{tr}_A(\rho)$ is the reduced state of ρ on $\mathcal{H}^B$. Further details about quantum distance measures and quantum entropies can be found in, e.g., Refs. [1,4].

A useful property of the von Neumann entropy is that it is continuous for finite-dimensional quantum systems. This motivates the search for so-called entropic continuity bounds, which capture the notion that two states ρ, σ close in some metric d, e.g. $\mathrm{d}(\rho,\sigma)=\delta\gtrsim 0$, are expected to be close in entropy as well, i.e.,

$$|\mathrm{H}(\rho)-\mathrm{H}(\sigma)| \leqslant f(\delta),$$

where f is some function such that $\lim_{\delta\to 0} f(\delta)=0$.

For example, in Ref. [5], Audenaert derived the tightest form of the Fannes-type continuity bound for the von Neumann entropy in terms of trace distance. Specifically, letting $d := \dim\mathcal{H}\in\mathbb{N}$, $\rho,\sigma\in\mathcal{D}(\mathcal{H})$, and $T := \mathrm{T}(\rho,\sigma)$, Audenaert showed that

$$|\mathrm{H}(\rho)-\mathrm{H}(\sigma)| \leqslant T\ln(d-1)+\mathrm{h}(T), \tag{4}$$

where $\mathrm{h}(x) := -x\ln x-(1-x)\ln(1-x)$ is the binary entropy function.

Similar continuity bounds also exist for conditional entropies. As shown by Winter [6], letting $d_A := \dim\mathcal{H}^A\in\mathbb{N}$, $d_B := \dim\mathcal{H}^B\in\mathbb{N}$, $\rho,\sigma\in\mathcal{D}(\mathcal{H}^A\otimes\mathcal{H}^B)$, and $T := \mathrm{T}(\rho,\sigma)$, the following holds:

$$|\mathrm{H}(A|B)_\rho-\mathrm{H}(A|B)_\sigma| \leqslant 2T\ln d_A+(1+T)\,\mathrm{h}\left(\frac{T}{1+T}\right). \tag{5}$$

[*] mliu24@amherst.edu

[1] In this work, we define entropies via the natural logarithm rather than the base-2 logarithm for ease of presentation in the proofs.

Such continuity bounds have been applied in various contexts. For example, in Ref. [7], Upadhyaya *et al.* constructed a finite-dimensional cutoff formulation for a class of infinite-dimensional entropy optimization problems. Qualitatively, that work argues that if an infinite-dimensional state is "close" (under some metric) to a finite-dimensional state, then an entropic continuity bound allows us to replace the former with the latter and compensate for the resulting change in entropy by applying a correction based on the continuity bound. This so-called dimension-reduction method plays an important role in quantum key distribution (QKD) security proofs [8]. However, it relies heavily on the continuity bound in Eq. (5) to compute the required correction term. An improved continuity bound would lead to a smaller correction term in this method and hence a larger secret key rate.

Another application of entropic continuity bounds arises in unstructured entropy optimization problems, as studied in, e.g., Ref. [9]. In that work, the approach is that, in order to minimize the entropy over some set of states, one simply computes the entropy on a sufficiently fine discrete "grid" of states in the set, then uses the continuity bound to ensure that the true minimum does not lie more than $f(\delta)$ away from the minimum over the grid. Again, an improved continuity bound would result in tighter results from such an approach.

In the above contexts, two desirable properties of the continuity bound $f(\delta)$ (for conditional entropies) are as follows:

Condition 1. $f(\delta)$ should be independent of d_B, the dimension of the conditioning system $\mathcal{H}^B$.

Condition 2. $f(\delta)$ should have finite (and ideally small) derivative at $\delta = 0$.

The first property is useful (or in some cases required) for the applications mentioned above, since in those contexts the conditioning system may have large or unbounded dimension. The second property is desirable for obtaining better scaling at small δ, since then we would not require extremely small values of δ in order to force the entropy difference to be small.

While the Winter bound [Eq. (5)] satisfies condition 1, it does not satisfy condition 2 due to the binary entropy term h, which has unbounded derivative as $\delta \to 0$. In fact, such scaling of the conditional entropy with respect to trace distance is in some sense unavoidable, since there is an explicit family of states that saturates the Audenaert bound [Eq. (4)], which has the binary entropy term as well. To work around this issue and obtain a bound that satisfies both conditions 1 and 2, one approach is to consider an alternative distance measure such as the angular distance, defined as

$$\mathrm{A}(\rho, \sigma) := \arccos \mathrm{F}(\rho, \sigma).$$

We remark that this is not simply an arbitrary change of distance measure: in the context of the applications mentioned above, the quantity that arises "naturally" in the analysis is the fidelity rather than the trace distance, hence working with the bound in Eq. (5) is somewhat suboptimal.

This approach is promising in light of the following result: In Ref. [9], Sekatski *et al.* proved Lipschitz continuity of the von Neumann entropy with respect to angular distance. That is, for $d := \dim \mathcal{H} \in \mathbb{N}$, $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, and $x_0 := \exp[\mathrm{W}_0(-\frac{2}{e})] \approx 4.922$, where W_0 is the principal branch of the Lambert W function, it was shown that

$$|\mathrm{H}(\rho) - \mathrm{H}(\sigma)| \leqslant u(d)\, \mathrm{A}(\rho, \sigma), \tag{6}$$

where the Lipschitz constant $u(d)$ is

$$u(d) := \begin{cases} \sqrt{8\frac{\ln x_0}{x_0}}\sqrt{d-1}, & 1 \leqslant d \leqslant 4 \\ 2\ln d, & d \geqslant 5. \end{cases} \tag{7}$$

Now, a naive application of Eq. (6) to conditional entropies, using Eq. (3) and the triangle inequality, would yield

$$\begin{aligned} &|\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma| \\ &\quad \leqslant u(d_A d_B)\, \mathrm{A}(\rho, \sigma) + u(d_B)\, \mathrm{A}(\rho_B, \sigma_B) \\ &\quad \leqslant [u(d_A d_B) + u(d_B)]n\, \mathrm{A}(\rho, \sigma), \end{aligned} \tag{8}$$

where in the last line we used the monotonicity of the angular distance under quantum channels. While this bound satisfies condition 2, it violates condition 1. However, the estimates to obtain Eq. (8) from Eq. (6) are crude and leave room for refinement. Thus, we ask whether it is possible to obtain Lipschitz continuity of the conditional entropy with respect to angular distance, while avoiding dependence on d_B in the final bound.

In this work, we answer this question in the affirmative when ρ and σ are quantum-classical states on $\mathcal{H}^A \otimes \mathcal{H}^B$ i.e., when there exists an orthonormal basis $\{|g_k\rangle\}_k$ for $\mathcal{H}^B$ such that both ρ and σ are of the form $\sum_k \gamma_k \tau_k \otimes |g_k\rangle\langle g_k|$ for some density operators $\tau_k \in \mathcal{D}(\mathcal{H}^A)$ and probabilities $\gamma_k \in [0, 1]$. We present this result in Sec. II. However, we find that our bound does not hold in general for fully quantum states. To further investigate counterexamples in this setting, we study characterizations of states saturating the Fuchs–van de Graaf inequalities. In particular, the states saturating the upper bound in the inequality have $\mathrm{T}(\rho, \sigma) \approx \mathrm{A}(\rho, \sigma)$ when $\mathrm{A}(\rho, \sigma)$ is small, so these states could pose an obstruction to deriving continuity bounds in terms of $\mathrm{A}(\rho, \sigma)$ that scale better than those in terms of $\mathrm{T}(\rho, \sigma)$. While it is well known that any pair of pure states saturate the upper Fuchs–van de Graaf inequality, we show that these are not the only such states. In Sec. III, we provide a characterization of all such pairs (ρ, σ) in the case where both of them are invertible. This result may be of independent interest in other applications such as computing QKD keyrates (we discuss this further in the Appendixes). However, we find that such a characterization in the general case where (ρ, σ) are noninvertible appears significantly more challenging, and we discuss how it relates to identifying the set of measurements that preserve the fidelity between states. Finally, we provide some concluding remarks in Sec. IV.

II. CONTINUITY BOUND

We now state and prove the main result of our paper, a continuity bound for the conditional entropy of quantum-classical states with respect to angular distance. Subsequently, we discuss the tightness of this bound, and we highlight some challenges for generalizing our result to classical-quantum or fully quantum states.

A. Main theorem and proof

Theorem 1. Let $\mathcal{H}^A$ and $\mathcal{H}^B$ be Hilbert spaces of finite dimension d_A and d_B, respectively. Let $\rho, \sigma \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$. Let $u(\cdot)$ and x_0 be defined as in Eq. (7) and the preceding text. Suppose in addition that ρ and σ are both quantum-classical states with respect to $\mathcal{H}^A$ and $\mathcal{H}^B$. Then

$$|\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma| \leqslant u(d_A)\,\mathrm{A}(\rho, \sigma). \tag{9}$$

Proof. Since ρ and σ are quantum-classical states, we can write

$$\rho = \sum_{k=1}^{d_B} \alpha_k \rho_k \otimes |f_k\rangle\langle f_k|,$$
$$\sigma = \sum_{k=1}^{d_B} \beta_k \sigma_k \otimes |f_k\rangle\langle f_k|,$$

for some density operators $\rho_k,\ \sigma_k \in \mathcal{D}(\mathcal{H}^A)$, probabilities $\alpha_k,\ \beta_k \in [0, 1]$ which satisfy $\sum_{k=1}^{d_B} \alpha_k = 1 = \sum_{k=1}^{d_B} \beta_k$, and orthonormal basis $\{|f_k\rangle\}_k$ for $\mathcal{H}^B$. For each k, consider a spectral decomposition of ρ_k and σ_k,

$$\rho_k = \sum_{j=1}^{d_A} p_{jk} |e_{jk}\rangle\langle e_{jk}|,$$
$$\sigma_k = \sum_{j=1}^{d_A} q_{jk} |\tilde{e}_{jk}\rangle\langle \tilde{e}_{jk}|,$$

where the eigenvalues $p_{jk}, q_{jk} \geqslant 0$ satisfy $\sum_{j=1}^{d_A} p_{jk} = 1 = \sum_{j=1}^{d_A} q_{jk}$, and the eigenvectors form orthonormal bases $\{|e_{jk}\rangle\}_j, \{|\tilde{e}_{jk}\rangle\}_j$ for $\mathcal{H}^A$. Defining $\rho_{jk} := p_{jk}\alpha_k$ and $\sigma_{jk} := q_{jk}\beta_k$ for all j and k, ρ and σ can be written as

$$\rho = \sum_{k=1}^{d_B} \sum_{j=1}^{d_A} \rho_{jk} |e_{jk}\rangle\langle e_{jk}| \otimes |f_k\rangle\langle f_k|,$$
$$\sigma = \sum_{k=1}^{d_B} \sum_{j=1}^{d_A} \sigma_{jk} |\tilde{e}_{jk}\rangle\langle \tilde{e}_{jk}| \otimes |f_k\rangle\langle f_k|,$$

and their partial traces can be written as

$$\rho_B := \mathrm{tr}_A(\rho) = \sum_{k=1}^{d_B} \left(\sum_{j=1}^{d_A} \rho_{jk} \right) |f_k\rangle\langle f_k|,$$
$$\sigma_B := \mathrm{tr}_A(\sigma) = \sum_{k=1}^{d_B} \left(\sum_{j=1}^{d_A} \sigma_{jk} \right) |f_k\rangle\langle f_k|.$$

Now, observe that the eigenvalues ρ_{jk} and σ_{jk} of ρ and σ completely determine the eigenvalues of their partial traces ρ_B and σ_B, respectively. This allows us to "map" the problem to $\mathbb{R}^d$, where $d := d_A d_B$, as follows: For each $k \in \{1, \ldots, d_B\}$, let us choose the ordering of the eigenvalues p_{jk} (and corresponding eigenvectors $|e_{jk}\rangle$) to be such that $p_{1k} \geqslant p_{2k} \geqslant \cdots \geqslant p_{d_A k}$; similarly, choose the ordering of the eigenvalues q_{jk} to be such that $q_{1k} \geqslant q_{2k} \geqslant \cdots \geqslant q_{d_A k}$. Now, consider the vectors

$$r := (\sqrt{\rho_{jk}})_{k,j},$$
$$s := (\sqrt{\sigma_{jk}})_{k,j} \tag{10}$$

in $\mathbb{R}^d$, where the entries of r and s are ordered with k as the outer index and j as the inner index. We observe that the angular distance $\mathrm{A}(\rho, \sigma)$ between ρ and σ is always lower bounded by the angular distance $\theta_0 := \arccos(r \cdot s) \in [0, \frac{\pi}{2}]$ between r and s. To see this, we decompose the fidelity as a sum over k using the quantum-classical structure, then apply a variational characterization of the trace norm [1] and the von Neumann trace inequality [10], which yields

$$\begin{aligned}\|\sqrt{\rho}\,\sqrt{\sigma}\|_1 &= \sum_{k=1}^{d_B} \sqrt{\alpha_k \beta_k} \|\sqrt{\rho_k}\sqrt{\sigma_k}\|_1 \\ &= \sum_{k=1}^{d_B} \sqrt{\alpha_k \beta_k} |\,\mathrm{tr}(\sqrt{\rho_k}\sqrt{\sigma_k} U_k)| \\ &\leqslant \sum_{k=1}^{d_B} \sqrt{\alpha_k \beta_k} \sum_{j=1}^{d_A} \sqrt{p_{jk}}\sqrt{q_{jk}} \\ &= \sum_{k=1}^{d_B} \sum_{j=1}^{d_A} \sqrt{\rho_{jk}}\sqrt{\sigma_{jk}} \\ &= r \cdot s,\end{aligned}$$

where the U_k are some unitaries on $\mathcal{H}^A$. Thus, we see that

$$\theta_0 = \arccos(r \cdot s) \leqslant \arccos \|\sqrt{\rho}\,\sqrt{\sigma}\|_1 = \mathrm{A}(\rho, \sigma), \tag{11}$$

as needed.

Next, since the eigenvalues of ρ and σ completely determine the eigenvalues of their partial traces, it is possible to compute the conditional entropy of ρ and σ given only the vectors r and s. To see this, consider the following function:

$$\begin{aligned}\mathrm{H}_c\,(v) := &- \sum_{k=1}^{d_B} \sum_{j=1}^{d_A} v_{jk}^2 \ln v_{jk}^2 \\ &+ \sum_{k=1}^{d_B} \left(\sum_{j=1}^{d_A} v_{jk}^2 \right) \ln \left(\sum_{l=1}^{d_A} v_{lk}^2 \right),\end{aligned}$$

where $v = (v_{jk})_{k,j}$ can be any vector in $\mathbb{R}^d$. Then,

$$\begin{aligned}\mathrm{H}_c\,(r) &= - \sum_{k=1}^{d_B} \sum_{j=1}^{d_A} \rho_{jk} \ln \rho_{jk} + \sum_{k=1}^{d_B} \left(\sum_{j=1}^{d_A} \rho_{jk} \right) \ln \left(\sum_{l=1}^{d_A} \rho_{lk} \right) \\ &= \mathrm{H}(A|B)_\rho, \\ \mathrm{H}_c(s) &= - \sum_{k=1}^{d_B} \sum_{j=1}^{d_A} \sigma_{jk} \ln \sigma_{jk} + \sum_{k=1}^{d_B} \left(\sum_{j=1}^{d_A} \sigma_{jk} \right) \ln \left(\sum_{l=1}^{d_A} \sigma_{lk} \right) \\ &= \mathrm{H}(A|B)_\sigma,\end{aligned} \tag{12}$$

so the vectors r and s are sufficient to determine the conditional entropies $\mathrm{H}(A|B)_\rho$ and $\mathrm{H}(A|B)_\sigma$.

The idea of our proof is now to integrate from r to s in $\mathbb{R}^d$, tracking the infinitesimal changes in the conditional entropy and angular distance. To see this formally, first note that r and s are unit vectors (with respect to the standard inner product on $\mathbb{R}^d$), since $r \cdot r = \operatorname{tr}(\rho) = 1 = \operatorname{tr}(\sigma) = s \cdot s$. Moreover, we have $r, s \geqslant 0$ by definition (10). Now, note that if $r \cdot s = 1$, then $r = s$, so we have $\mathrm{H}_c(r) = \mathrm{H}_c(s)$ i.e., $\mathrm{H}(A|B)_\rho = \mathrm{H}(A|B)_\sigma$. Since $u(d_A) \geqslant 0$ and $\mathrm{A}(\rho, \sigma) \geqslant 0$, Eq. (9) holds trivially in this case. Now consider the remaining case $r \cdot s \in [0, 1)$. Let $\tilde{s}$ be the normalized projection of s onto the orthogonal complement of $\mathrm{Span}\{r\}$,

$$\tilde{s} := \frac{s - (s \cdot r)r}{|s - (s \cdot r)r|}.$$

Using $\tilde{s}$, we define the path

$$v(\theta) := \cos(\theta) r + \sin(\theta)\tilde{s}$$

from r to s, where $\theta \in [0, \theta_0]$. Note that $v(0) = r$, $v(\theta_0) = s$, and $v(\theta)$ traverses the great circle along the $(d-1)$ sphere from r to s. In addition, note that $|v(\theta)| = 1$ for all $\theta \in [0, \theta_0]$. Now, the tangent to the path $v(\theta)$ is

$$w(\theta) := v'(\theta) = -\sin(\theta) r + \cos(\theta)\tilde{s},$$

which satisfies $|w(\theta)| = 1$ and $v(\theta) \cdot w(\theta) = 0$ for all $\theta \in [0, \theta_0]$.

For notational simplicity, we now define $\mathrm{H}_c(\theta) := \mathrm{H}_c(v(\theta))$, so

$$\begin{aligned}\mathrm{H}_c(\theta) = &- \sum_{k=1}^{d_B} \sum_{j=1}^{d_A} v(\theta)_{jk}^2 \ln\left(v(\theta)_{jk}^2\right)\\ &+ \sum_{k=1}^{d_B} \left(\sum_{j=1}^{d_A} v(\theta)_{jk}^2 \right) \ln \left(\sum_{l=1}^{d_A} v(\theta)_{lk}^2 \right).\end{aligned}$$

Observe that $\mathrm{H}_c(\theta)$ is continuous on $[0, \theta_0]$ (under the standard convention for entropy definitions that $0 \ln 0 \equiv 0$). Thus, if we show that $\mathrm{H}_c(\theta)$ is differentiable on $(0, \theta_0)$ and its derivative satisfies $|\mathrm{H}_c'(\theta)| \leqslant u(d_A)$ on that interval, then the desired result follows immediately, since

$$\begin{aligned}&|\mathrm{H}(A|B)_\sigma - \mathrm{H}(A|B)_\rho|\\ &= |\mathrm{H}_c(\theta_0) - \mathrm{H}_c(0)| = \left| \int_0^{\theta_0} \mathrm{H}_c'(\theta)\, d\theta \right|\\ &\leqslant \int_0^{\theta_0} |\mathrm{H}_c'(\theta)|\, d\theta \leqslant u(d_A)\theta_0 \leqslant u(d_A)\,\mathrm{A}(\rho, \sigma),\end{aligned}$$

where the first line follows from Eq. (12) and the last line follows from Eq. (11). Thus, all that remains is to bound $|\mathrm{H}_c'(\theta)|$ by $u(d_A)$.

To do this, we first handle a technicality regarding zero eigenvalues. For each $k \in \{1, \dots, d_B\}$, let S_A^k be the set of $j \in \{1, \dots, d_A\}$ such that at least one of r_{jk}, s_{jk} is nonzero. Furthermore, let S_B be the set of $k \in \{1, \dots, d_B\}$ such that S_A^k is nonempty. Then for any (k, j) with $k \in S_B$ and $j \in S_A^k$, at least one of r_{jk}, s_{jk} is nonzero, which implies that $v(\theta)_{jk}^2 > 0$ for all $\theta \in (0, \theta_0)$. Moreover, for all other (k, j), we have that $r_{jk} = s_{jk} = 0$, so $v(\theta)_{jk}^2 = 0$ for all $\theta \in (0, \theta_0)$, which implies that the value of $\mathrm{H}_c(\theta)$ would not change upon removing the term $v(\theta)_{jk}^2$. Thus, in the remainder of the argument, summations of the form $\sum_{k,j}$ should be understood to mean $\sum_{k \in S_B} \sum_{j \in S_A^k}$ (and analogously, $\sum_l$ means $\sum_{l \in S_A^k}$), which ensures that all terms appearing in the summations satisfy $v(\theta)_{jk}^2 > 0$ and $\sum_{l \in S_A^k} v(\theta)_{lk}^2 > 0$ for all $\theta \in (0, \theta_0)$. With this, we see that $\mathrm{H}_c(\theta)$ is indeed differentiable on $(0, \theta_0)$, and

$$\begin{aligned}|H_c'(\theta)| = \Bigg| &- \sum_{k,j} 2v(\theta)_{jk} w(\theta)_{jk} \ln\left(v(\theta)_{jk}^2\right)\\ &- \sum_{k,j} v(\theta)_{jk}^2 \frac{2v(\theta)_{jk} w(\theta)_{jk}}{v(\theta)_{jk}^2}\\ &+ \sum_{k,j} 2v(\theta)_{jk} w(\theta)_{jk} \ln\left(\sum_l v(\theta)_{lk}^2 \right)\\ &+ \sum_{k,j} v(\theta)_{jk}^2 \frac{\sum_l 2v(\theta)_{lk} w(\theta)_{lk}}{\sum_l v(\theta)_{lk}^2} \Bigg|.\end{aligned}$$

Using $v(\theta) \cdot w(\theta) = 0$, this simplifies to

$$\begin{aligned}|H_c'(\theta)| &= 2 \left| \sum_{k,j} v(\theta)_{jk} w(\theta)_{jk} \ln\left(\frac{\sum_l v(\theta)_{lk}^2}{v(\theta)_{jk}^2} \right) \right|\\ &\leqslant 2 \sqrt{ \sum_{k,j} v(\theta)_{jk}^2 \ln^2 \left(\frac{\sum_l v(\theta)_{lk}^2}{v(\theta)_{jk}^2} \right) }, \qquad (13)\end{aligned}$$

where in the last line we used the Cauchy–Schwarz inequality with $|w(\theta)| = 1$.

Now, recall that the $v(\theta)_{jk}^2$ form a valid probability distribution (i.e. they are non-negative values summing to 1), since $|v| = 1$. Also, note that the argument of $\ln^2$ in the final line above, i.e., $\sum_l v(\theta)_{lk}^2 / v(\theta)_{jk}^2$, lies in the interval $[1, \infty)$. Thus, we now construct an increasing concave upper bound $f(x)$ for $\ln^2 x$ on $x \in [1, \infty)$, as this would allow us to "move the summation" over k, j [weighted by the probabilities $v(\theta)_{jk}^2$] into the argument of the function. To begin, note that

$$\begin{aligned}\frac{d}{dx} \ln^2 x &= 2 \frac{\ln x}{x},\\ \frac{d^2}{dx^2} \ln^2 x &= 2 \frac{1 - \ln x}{x^2},\end{aligned}$$

so $\ln^2 x$ is convex for all $x \in [1, e]$ and concave for all $x \in [e, \infty)$. Then to produce $f(x)$, we seek a line $y(x) = m(x - a)$ such that $y(1) = \ln^2(1) = 0$, and such that there exists $x_0 \in [1, \infty)$ with $x_0 \geqslant e$, $y(x_0) = \ln^2(x_0)$ and $y'(x_0) = (\frac{d}{dx} \ln^2 x)_{x_0} = 2 \ln x_0 / x_0$. Then we must solve the system

$$\ln^2 x_0 = y(x_0) = 2 \frac{\ln x_0}{x_0} (x_0 - 1),$$

for x_0, which has solutions $x_0 = 1$ and $x_0 = \exp[W_0(-\frac{2}{e})] \approx 4.922$. We discard the solution $x_0 = 1 < e$ and keep the other solution $x_0 \approx 4.922 \geqslant e$. Thus, our increasing concave upper bound for $\ln^2 x$ is

$$f(x) := \begin{cases} 2 \frac{\ln x_0}{x_0} (x - 1), & 1 \leqslant x \leqslant x_0 \\ \ln^2 x, & x \geqslant x_0. \end{cases}$$

With $f(x)$, we can write

$$\begin{aligned}
|H_c'(\theta)| &\leqslant 2\sqrt{\sum_{k,j} v(\theta)_{jk}^2 \ln^2\left(\frac{\sum_l v(\theta)_{lk}^2}{v(\theta)_{jk}^2}\right)} \\
&\leqslant 2\sqrt{\sum_{k,j} v(\theta)_{jk}^2 f\left(\frac{\sum_l v(\theta)_{lk}^2}{v(\theta)_{jk}^2}\right)} \\
&\leqslant 2\sqrt{f\left(\sum_{k,j} v(\theta)_{jk}^2 \frac{\sum_l v(\theta)_{lk}^2}{v(\theta)_{jk}^2}\right)} \\
&\leqslant 2\sqrt{f\left(d_A \sum_{k,l} v(\theta)_{lk}^2\right)} \\
&= 2\sqrt{f(d_A)} \\
&= \begin{cases} \sqrt{8\frac{\ln x_0}{x_0}}\sqrt{d_A - 1}, & 1 \leqslant d_A \leqslant 4 \\ 2\ln d_A, & d_A \geqslant 5, \end{cases} \\
&= u(d_A).
\end{aligned} \tag{14}$$

In the above, the first line is Eq. (13). The second line follows since f is an upper bound on $\ln^2 x$ for $x \geqslant 1$. The third line follows since f is concave and $|v(\theta)| = 1$. The fourth line follows since f is increasing and $|S_A^k| \leqslant d_A$ for all k. The fifth line follows since $|v(\theta)| = 1$. The sixth and seventh lines follow from the definitions and the fact that $d_A \in \mathbb{N}$. ■

In comparison to the proof in Ref. [9] for unconditioned entropies, the main difference in our proof here is essentially that there are additional contributions to the derivative $\mathrm{H}_c'(\theta)$ arising from the $\mathrm{H}(\rho_B)$ term in the conditional entropy. Informally, these contributions act in the "opposite direction" from those of the $\mathrm{H}(\rho_{AB})$ term, reducing the magnitude of the derivative and yielding a final bound that is independent of d_B, in contrast to what we would have obtained had we only considered the derivative of the $\mathrm{H}(\rho_{AB})$ term alone—see Eq. (8). Another small difference is that we have constructed the concave upper bound in a slightly different and arguably simpler way.

B. Potential improvements

How tight is the bound in Eq. (9)? To address this question empirically, we began by randomly sampling 100 000 pairs of quantum-classical states according to the procedure in Appendix D 1, and for each pair (ρ, σ), we computed their angular distance $\mathrm{A}(\rho, \sigma)$ and conditional entropy difference $|\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma|$. We then plotted the conditional entropy differences against the angular distances in Fig. 1. This provides an empirical estimate for the tightness of Eq. (9) at all feasible angular distances $\mathrm{A}(\rho, \sigma) \in [0, \frac{\pi}{2}]$.

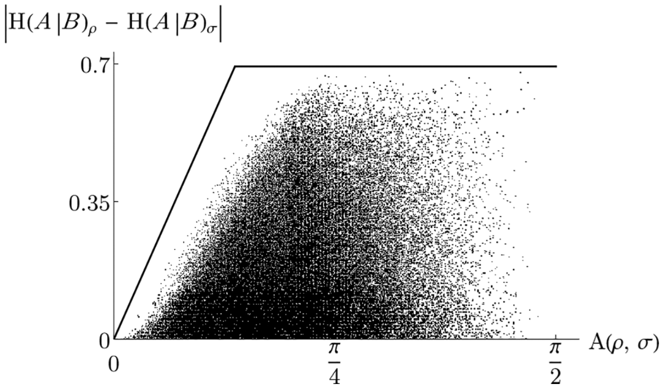

FIG. 1. Scatter plot of the conditional entropy difference $|\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma|$ against the angular distance $\mathrm{A}(\rho, \sigma)$ for 100 000 pairs of randomly sampled quantum-classical states (ρ, σ) with $d_A = 2 = d_B$ (see Appendix D 1). For comparison, we have also plotted the curve $\min\{u(d_A)\,\mathrm{A}(\rho, \sigma), \ln d_A\}$, where $u(d_A)\,\mathrm{A}(\rho, \sigma)$ is the bound in Eq. (9), and $\ln d_A$ is a hard upper bound on the conditional entropy difference between quantum-classical states.

Next, we explore the tightness of Eq. (9) at small angular distances (since in most applications we are mainly interested in this case). For each angular distance $\mathrm{A}(\rho, \sigma) \in \{0.1\times10^{-5}, 0.2\times10^{-5}, \ldots, 1.0\times10^{-5}\}$, we randomly sampled 10 000 pairs of classical states (ρ, σ) with that angular distance, as described in Appendix D 2. Then, we computed the conditional entropy differences $|\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma|$ and plotted them in Fig. 2. The results suggest that at small angular distances, our continuity bound is close to the "true" tight expression when d_A is small, but there may be room for improvement when d_A is larger (note that random sampling typically yields less representative results in high dimensions, so the latter claim should not be taken as conclusive).

Note that, in order to saturate Eq. (9), a pair of states (ρ, σ) must saturate both inequalities (13) (Cauchy–Schwarz) and (14) (which roughly speaking is due to the concavity of f). However, it seems that these inequalities cannot be simultaneously saturated, which is consistent with the above empirical evidence that there is room for sharpening the bound.

It is also worth briefly comparing "conversions" between our result and the continuity bounds based on trace distance. Specifically, note that the Fuchs–van de Graaf inequalities [11] upper bound the trace distance in terms of angular distance and vice versa. Thus, a continuity bound in terms of either distance measure in principle yields a continuity bound in terms of the other. However, such a conversion is potentially quite suboptimal—we provide a brief scaling comparison in Appendix A, where we find that, starting from a bound on angular distance and then applying the previous continuity bound (5), yields highly suboptimal results, while the other direction [starting from a bound on trace distance and then applying our bound (9)] is somewhat better, although still not tight.

Finally, we discuss avenues for generalizing Eq. (9), since many applications in quantum key distribution require Eq. (9) (or a similar bound satisfying both conditions 1 and 2) to hold for classical-quantum states as well. However, our proof technique does not appear to generalize readily to classical-quantum (or fully quantum) states, since our proof relies on the simple eigenvalue relationship between a quantum-classical state ρ and its partial trace $\mathrm{tr}_A(\rho)$. This eigenvalue relationship becomes much more complicated in the classical-quantum (or fully quantum) case, since the eigenvalues of $\mathrm{tr}_A(\rho)$ now depend on the eigenvectors of ρ as

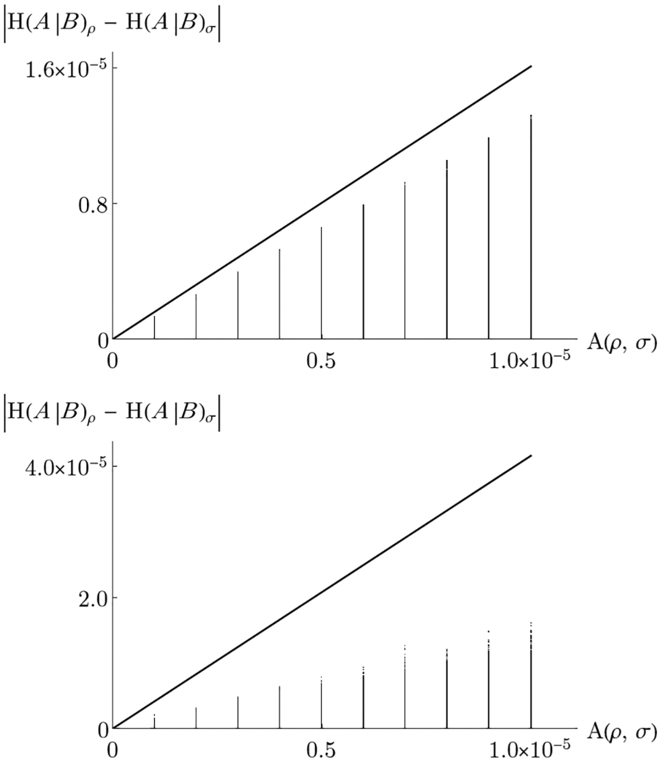

FIG. 2. Scatter plot of the conditional entropy difference $|\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma|$ against the angular distance $\mathrm{A}(\rho,\sigma)$ for 100 000 pairs of classical states (ρ,σ) randomly sampled at fixed angular distances $\mathrm{A}(\rho,\sigma) \in \{0.1\times10^{-5}, 0.2\times10^{-5}, \ldots, 1.0\times10^{-5}\}$ (see Appendix D 2). The slanted line shows the bound $u(d_A)\,\mathrm{A}(\rho,\sigma)$ in Eq. (9), while the vertical "lines" are formed by the data points from the 100 000 sampled pairs (ρ,σ). (Top panel) $d_A = 2 = d_B$; (Bottom panel) $d_A = 8$, $d_B = 2$.

well. We also highlight that as observed in Ref. [5], it seems difficult to use purification-based arguments to obtain such a result, because purifications usually do not "preserve" the conditional entropies in a useful way—we give some further details in Appendix B.

In attempting to generalize Eq. (9), it is important to note that the bound does not hold for arbitrary (fully quantum) states $\rho, \sigma \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$. To see this, let $\{|e_j\rangle\}_j$ and $\{|f_k\rangle\}_k$ be orthonormal bases for $\mathcal{H}^A$ and $\mathcal{H}^B$, respectively. Let $d_M := \min\{d_A, d_B\}$, and consider the maximally entangled state

$$\rho = \frac{1}{d_M}\sum_{j,k=1}^{d_M} |e_j f_j\rangle\langle e_k f_k|, \qquad (15)$$

which has the most negative conditional entropy $\mathrm{H}(A|B)_\rho = -\ln d_M$. Next, consider the maximally mixed state

$$\tau = \frac{1}{d_A d_B}\mathbb{1},$$

which has the most positive conditional entropy $\mathrm{H}(A|B)_\tau = \ln d_A$. Now, consider the segment connecting ρ and τ,

$$\sigma = \lambda\tau + (1-\lambda)\rho, \qquad (16)$$

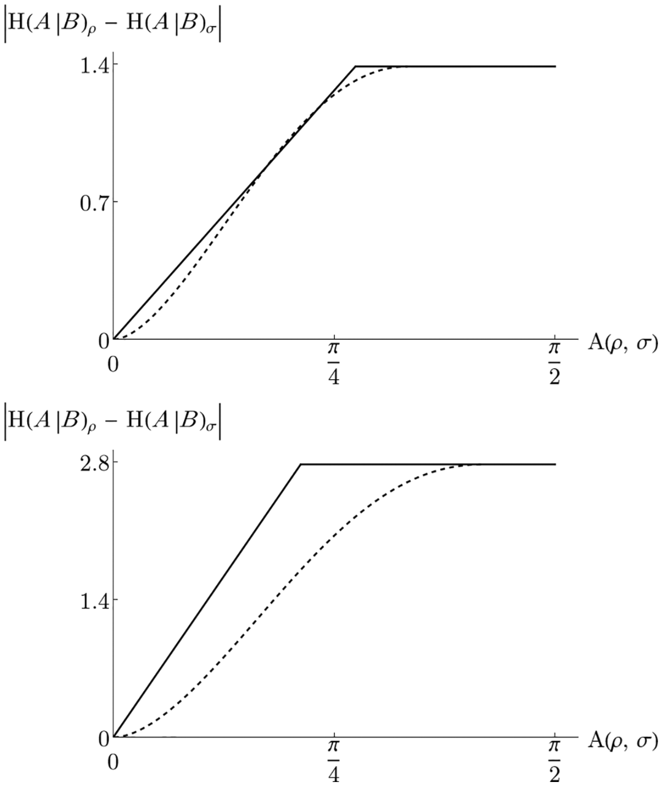

FIG. 3. Parametric plot (dashed line) of $|\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma|$ against $\mathrm{A}(\rho,\sigma)$, where ρ and σ are as in Eqs. (15) and (16), respectively, for $\lambda \in [0,1]$. For comparison, the solid line shows $\min\{u(d_A)\,\mathrm{A}(\rho,\sigma), \ln(d_A d_M)\}$ with $d_M := \min\{d_A, d_B\}$, where $u(d_A)\,\mathrm{A}(\rho,\sigma)$ is the bound in Eq. (9), and $\ln(d_A d_M)$ is a hard upper bound on the conditional entropy difference between fully quantum states. (Top panel) $d_A = 2 = d_B$; (Bottom panel) $d_A = 8$, $d_B = 2$.

where $\lambda \in [0,1]$. By direct computation, one can show that

$$\mathrm{A}(\rho,\sigma) = \arccos\left(\sqrt{1 - \frac{d_A d_B - 1}{d_A d_B}\lambda}\right),$$

and that

$$\begin{aligned}|H(A|B)_\rho - H(A|B)_\sigma| &= -\ln d_M + \left(1 - \frac{d-1}{d}\lambda\right)\ln\left(1 - \frac{d-1}{d}\lambda\right)\\ &\quad + (d-1)\frac{\lambda}{d}\ln\frac{\lambda}{d} - (d_B - d_M)\frac{\lambda}{d_B}\ln\frac{\lambda}{d_B}\\ &\quad - d_M\left(\frac{\lambda}{d_B} + \frac{1-\lambda}{d_M}\right)\ln\left(\frac{\lambda}{d_B} + \frac{1-\lambda}{d_M}\right),\end{aligned}$$

where $d := d_A d_B$. For the counterexample to Eq. (9), let $d_A = 2 = d_B$ and $\lambda = \frac{1}{2}$. Then

$$|H(A|B)_\rho - H(A|B)_\sigma| \approx 1.074,$$

but

$$u(d_A)\,\mathrm{A}(\rho,\sigma) \approx 1.061,$$

in violation of Eq. (9). This situation is depicted in the top panel of Fig. 3.

However, it appears that violations of Eq. (9) are uncommon and relatively small in magnitude—the above is the only and the most egregious counterexample to Eq. (9) known to the authors. Also, numerical computation shows that for every choice of $d_A \in \{2, \dots, 10\}$, every choice of $d_B \in \{1, \dots, 10\}$, and for $\lambda \in [0, 1]$, the above construction produces a counterexample only when $d_A = 2 = d_B$ and $\lambda \approx 0.5$ (see the bottom panel of Fig. 3 for a representative example). Thus, a slight modification of the bound in Eq. (9), perhaps by including an extra factor of two [i.e., $u(d) \to 2\,u(d)$ in Eq. (7)], will resolve the only counterexample known to the authors and may generalize the bound in Eq. (9) to all states $\rho, \sigma \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$.

III. TRACE DISTANCE AND ANGULAR DISTANCE

We now shift our attention to the relationship between the trace distance and the angular distance. Specifically, we study the set of states which saturate the Fuchs–van de Graaf inequalities [11]

$$1 - \mathrm{F}(\rho, \sigma) \leqslant \mathrm{T}(\rho, \sigma) \leqslant \sqrt{1 - \mathrm{F}(\rho, \sigma)^2}. \tag{17}$$

To motivate this direction of investigation, note that Audenaert's continuity bound for the von Neumann entropy in Eq. (4) is tight [5]. That is, for any $T \in [0, 1]$, there exist $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ such that $\mathrm{T}(\rho, \sigma) = T$ and $|\mathrm{H}(\rho) - \mathrm{H}(\sigma)| = T \ln(d-1) + \mathrm{h}(T)$. But suppose that some such ρ, σ which saturate Eq. (4) also saturate the right-hand side of Eq. (17) (which we refer to as the "upper Fuchs–van de Graaf inequality"), i.e.,

$$\mathrm{T}(\rho, \sigma) = \sqrt{1 - \mathrm{F}(\rho, \sigma)^2} = \sin \mathrm{A}(\rho, \sigma).$$

Then for $\mathrm{T}(\rho, \sigma) \ll 1$, we have $\mathrm{T}(\rho, \sigma) = \sin \mathrm{A}(\rho, \sigma) \approx \mathrm{A}(\rho, \sigma) =: A$. But then it is impossible to obtain a continuity bound for the conditional entropy which satisfies both condition 1 and condition 2, since whenever $d_B = 1$, such a pair ρ, σ would satisfy

$$\begin{aligned} |\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma| &= |\mathrm{H}(\rho) - \mathrm{H}(\sigma)| \\ &= T \ln(d-1) + \mathrm{h}(T) \\ &\approx A \ln(d-1) + \mathrm{h}(A), \end{aligned}$$

so the conditional entropy scales badly with the angular distance at $A \approx 0$, and condition 2 would not be satisfiable.

A. Invertible states

With this motivation, our goal is now to characterize the set of all states which saturate the Fuchs–van de Graaf inequalities. We begin by studying the subset of such states that are also invertible (i.e. positive definite), since this case is easier to handle. That is, we characterize the sets

$$\mathcal{S}_1 := \{(\rho, \sigma) \in \mathcal{D}_{\mathrm{inv}}(\mathcal{H})^2 \mid 1 - \mathrm{F}(\rho, \sigma) = \mathrm{T}(\rho, \sigma)\}, \tag{18}$$

and

$$\mathcal{S}_2 := \{(\rho, \sigma) \in \mathcal{D}_{\mathrm{inv}}(\mathcal{H})^2 | \, \mathrm{T}(\rho, \sigma) = \sqrt{1 - \mathrm{F}(\rho, \sigma)^2}\}, \tag{19}$$

where $\mathcal{D}_{\mathrm{inv}}(\mathcal{H})$ denotes the set of all invertible density operators on $\mathcal{H}$. To do this, we apply the following general strategy: A generic inequality $\mathrm{G}_1(\rho, \sigma) \leqslant \mathrm{G}_n(\rho, \sigma)$ is usually proven via a chain of inequalities

$$\mathrm{G}_1(\rho, \sigma) \leqslant \mathrm{G}_2(\rho, \sigma) \leqslant \cdots \leqslant \mathrm{G}_{n-1}(\rho, \sigma) \leqslant \mathrm{G}_n(\rho, \sigma).$$

Thus, to determine which states ρ, σ satisfy $\mathrm{G}_1(\rho, \sigma) = \mathrm{G}_n(\rho, \sigma)$, we derive the equality conditions for each inequality $\mathrm{G}_j(\rho, \sigma) \leqslant \mathrm{G}_{j+1}(\rho, \sigma)$ in the above chain of inequalities, and we combine the equality conditions (with logical and) for all $j \in \{1, \dots, n-1\}$ to obtain the equality condition for $\mathrm{G}_1(\rho, \sigma) \leqslant \mathrm{G}_n(\rho, \sigma)$. However, there usually exist multiple different proofs of the inequality $\mathrm{G}_1(\rho, \sigma) \leqslant \mathrm{G}_n(\rho, \sigma)$, which each proceed through alternative chains of intermediate inequalities

$$\mathrm{G}_1(\rho, \sigma) \leqslant \tilde{\mathrm{G}}_2(\rho, \sigma) \leqslant \cdots \leqslant \tilde{\mathrm{G}}_{n-1}(\rho, \sigma) \leqslant \mathrm{G}_n(\rho, \sigma).$$

Thus, a careful consideration of the various proof techniques for a given inequality is needed for analyzing equality conditions, since different proof techniques may yield equality conditions vastly different in appearance. Of course, all equality conditions for the same inequality $\mathrm{G}_1(\rho, \sigma) \leqslant \mathrm{G}_n(\rho, \sigma)$ should be logically equivalent regardless of the underlying proof technique, but some equality conditions may not be "compatible" with others in our overall proof, making it more difficult to condense all the intermediate equality conditions into a final concise characterization.

To proceed, we need to introduce a few definitions. First, for any probability distributions p and q on a finite alphabet $\mathcal{X}$, we denote their classical trace distance by

$$\mathrm{T}_c(p, q) := \frac{1}{2} \sum_{x \in \mathcal{X}} |p(x) - q(x)|$$

and their classical fidelity by

$$\mathrm{F}_c(p, q) := \sum_{x \in \mathcal{X}} \sqrt{p(x) q(x)},$$

which are special cases of Eqs. (1) and (2) for commuting density operators. Next, we denote the set of all rank-1 projective measurements on a Hilbert space $\mathcal{H}$ by

$$\begin{aligned} \mathcal{M} := \{\{|e_x\rangle\langle e_x|\}_{x \in \mathcal{X}} \subseteq \mathcal{L}(\mathcal{H}) | \\ \{|e_x\rangle\}_{x \in \mathcal{X}} \text{ is an orthonormal basis for } \mathcal{H}\}, \end{aligned}$$

where $\mathcal{L}(\mathcal{H})$ denotes the space of linear operators on $\mathcal{H}$. Now, each $\Lambda := \{|e_x\rangle\langle e_x|\}_{x \in \mathcal{X}} \in \mathcal{M}$ and $\tau \in \mathcal{D}(\mathcal{H})$ induce a natural probability distribution

$$\mathrm{tr}(\Lambda \tau) := [\mathrm{tr}(|e_x\rangle\langle e_x| \tau)]_{x \in \mathcal{X}} = (\langle e_x | \tau\, e_x\rangle)_{x \in \mathcal{X}} \in \mathbb{R}^d,$$

where $d := \dim \mathcal{H}$. That is, this probability distribution is given by

$$\mathrm{tr}(\Lambda \tau)(x) := \mathrm{tr}(|e_x\rangle\langle e_x| \tau) = \langle e_x | \tau\, e_x\rangle$$

for all $x \in \mathcal{X}$. Finally, for any positive-definite operators A and B on a Hilbert space $\mathcal{H}$, we denote their geometric mean [12] by

$$A \# B := A^{\frac{1}{2}} \sqrt{A^{-\frac{1}{2}} B A^{-\frac{1}{2}}} A^{\frac{1}{2}} = B \# A. \tag{20}$$

Now, to obtain equality conditions for Eq. (17), we found that Fuchs and van de Graaf's original proof technique in

Ref. [11] seemed amenable for analysis. Their proof of Eq. (17) invoked a variational characterization of the trace distance, a variational characterization of the fidelity, and a classical version of the bound in Eq. (17). We first introduce these previously known results in turn. Then, we derive new equality conditions for these results as lemmas. Finally, we combine the lemmas in Theorem III A to obtain characterizations of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ in Eqs. (18) and (19). Proofs of the lemmas have been deferred to Appendix C for readability.

We first consider the variational characterization of the trace distance, which is stated in Ref. [11] and can be traced back to the work of Helstrom and Toussaint [13,14]. For any $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, their trace distance can be expressed as

$$\mathrm{T}\,(\rho,\sigma) = \max_{\Lambda\in\mathcal{M}} \mathrm{T}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)). \tag{21}$$

We define

$$\mathcal{T}(\rho,\sigma) := \{\Lambda \in \mathcal{M} \mid \mathrm{T}\,(\rho,\sigma) = \mathrm{T}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma))\}$$

to be the (nonempty) set of all rank-1 projective measurements which achieve the optimum in Eq. (21). We characterize this set in the following lemma:

Lemma 1. For any $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, we have

$$\mathcal{T}(\rho,\sigma) = \{\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M} | \\ \forall\, x \in \mathcal{X},\, |e_x\rangle \in \ker P \vee |e_x\rangle \in \ker Q\} \neq \emptyset,$$

where P and Q are the positive and negative parts of $\rho - \sigma$, respectively.

Proof. See Appendix C 1. ■

Next, we consider the variational characterization of the fidelity, which was first introduced by Fuchs and Caves in Ref. [15]. For any $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, their fidelity can be expressed as

$$\mathrm{F}\,(\rho,\sigma) = \min_{\Lambda\in\mathcal{M}} \mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)). \tag{22}$$

We define

$$\mathcal{F}(\rho,\sigma) := \{\Lambda \in \mathcal{M} \mid \mathrm{F}\,(\rho,\sigma) = \mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma))\}$$

to be the (nonempty) set of all rank-1 projective measurements which achieve the optimum in Eq. (22). We characterize this set for invertible ρ and σ in the following lemma:

Lemma 2. For any invertible $\rho, \sigma \in \mathcal{D}_{\mathrm{inv}}(\mathcal{H})$, we have

$$\mathcal{F}(\rho,\sigma) = \{\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M} | \\ \{|e_x\rangle\}_{x\in\mathcal{X}} \text{ is an eigenbasis for } M\} \neq \emptyset,$$

where

$$M := \rho^{-1}\#\sigma = \rho^{-\frac{1}{2}}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}\,\rho^{-\frac{1}{2}}$$

is the operator geometric mean between ρ^{-1} and σ.

Proof. See Appendix C 2. ■

Finally, we consider the classical analog of Eq. (17). As shown in Ref. [11], for any probability distributions p and q on a finite alphabet $\mathcal{X}$, the classical trace distance and classical fidelity are related by

$$1 - \mathrm{F}_c\,(p,q) \leqslant \mathrm{T}_c\,(p,q) \leqslant \sqrt{1 - \mathrm{F}_c\,(p,q)^2}. \tag{23}$$

We define

$$\mathcal{C}_1(\rho,\sigma) := \{\Lambda \in \mathcal{M} | 1 - \mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)) \\ = \mathrm{T}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma))\},$$

and

$$\mathcal{C}_2(\rho,\sigma) := \{\Lambda \in \mathcal{M} |\, \mathrm{T}_c(\mathrm{tr}(\Lambda\rho), \mathrm{tr}(\Lambda\sigma)) \\ = \sqrt{1 - \mathrm{F}_c(\mathrm{tr}(\Lambda\rho), \mathrm{tr}(\Lambda\sigma))^2}\}$$

to be the sets of rank-1 projective measurements which induce classical probability distributions $p := \mathrm{tr}(\Lambda\rho)$ and $q := \mathrm{tr}(\Lambda\sigma)$ which saturate the left-hand and right-hand inequalities in Eq. (23), respectively. We characterize these sets in the following lemma:

Lemma 3. For any $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, we have

$$\mathcal{C}_1(\rho,\sigma) = \{\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M} | \\ \forall\, x \in \mathcal{X},\ \langle e_x|\rho\, e_x\rangle = \langle e_x|\sigma\, e_x\rangle \\ \vee \langle e_x|\rho\, e_x\rangle = 0 \vee \langle e_x|\sigma\, e_x\rangle = 0\},$$

and

$$\mathcal{C}_2(\rho,\sigma) \\ = \Bigg\{\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M} | \\ \mathrm{tr}\,(\Lambda\rho) = \mathrm{tr}\,(\Lambda\sigma) \vee \mathrm{tr}\,(\Lambda\rho) \times \mathrm{tr}\,(\Lambda\sigma) = 0 \vee \\ \begin{pmatrix} \exists\, b \in (0,1), \forall\, x \in \mathcal{X},\, \langle e_x|\sigma\, e_x\rangle = b\langle e_x|\rho\, e_x\rangle \\ \vee \langle e_x|\sigma\, e_x\rangle = \frac{1}{b}\langle e_x|\rho\, e_x\rangle \end{pmatrix}\Bigg\}.$$

Proof. See Appendix C 3. ■

We are now ready to state the main result of this section: equality conditions for the left-hand and right-hand inequalities in Eq. (17). Intuitively, Theorem III A says that, for invertible states, the lower Fuchs–van de Graaf inequality is saturated only for trivial cases, while the upper Fuchs–van de Graaf inequality is saturated exactly when the geometric mean operator $M := \rho^{-1}\#\sigma$ and the difference operator $\rho - \sigma$ are simultaneously diagonalizable, *and* there exists some constant $c \in (0, 1]$ such that the eigenvalues of M are all either c or $\frac{1}{c}$.

Theorem 2. With $\mathcal{S}_1$ and $\mathcal{S}_2$ as defined in Eqs. (18) and (19) respectively, we have

$$\mathcal{S}_1 = \{(\rho,\sigma) \in \mathcal{D}_{\mathrm{inv}}(\mathcal{H})^2 | \rho = \sigma\}, \tag{24}$$

and

$$\mathcal{S}_2 = \Bigg\{(\rho,\sigma) \in \mathcal{D}_{\mathrm{inv}}(\mathcal{H})^2 | \rho = \sigma \vee \Bigg(\exists\, c \in (0,1), \\ \mathrm{spec}(M) = \left\{c, \frac{1}{c}\right\} \wedge [M, \rho - \sigma] = 0\Bigg)\Bigg\}, \tag{25}$$

where

$$M := \rho^{-1}\#\sigma = \rho^{-\frac{1}{2}}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}\,\rho^{-\frac{1}{2}}$$

is the operator geometric mean between ρ^{-1} and σ, and $\mathrm{spec}(M)$ denotes the spectrum of M.

Proof. We first show one way to prove the left-hand inequality in Eq. (17) for arbitrary $\rho, \sigma \in \mathcal{D}(\mathcal{H})$. Observe that, for any $\Lambda_T \in \mathcal{T}(\rho, \sigma)$ and $\Lambda_F \in \mathcal{F}(\rho, \sigma)$, we have

$$\begin{aligned} 1 - \mathrm{F}(\rho, \sigma) &= 1 - \mathrm{F}_c(\mathrm{tr}\,(\Lambda_F \rho), \mathrm{tr}\,(\Lambda_F \sigma)) \\ &\leqslant \mathrm{T}_c(\mathrm{tr}\,(\Lambda_F \rho), \mathrm{tr}(\Lambda_F \sigma)) \\ &\leqslant \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T \rho), \mathrm{tr}(\Lambda_T \sigma)) \\ &= \mathrm{T}(\rho, \sigma), \end{aligned} \tag{26}$$

where the second line follows from Eq. (23) and the third line follows from Eq. (21).

With this, we can now prove Eq. (24). Restrict attention to invertible $\rho, \sigma \in \mathcal{D}_{\mathrm{inv}}(\mathcal{H})$. First note that, if $\rho = \sigma$, then clearly $1 - \mathrm{F}(\rho, \sigma) = \mathrm{T}(\rho, \sigma)$. Conversely, suppose that $1 - \mathrm{F}(\rho, \sigma) = \mathrm{T}(\rho, \sigma)$. Let $\{|e_x\rangle\}_{x\in\mathcal{X}}$ be an orthonormal eigenbasis for the operator geometric mean M and denote the corresponding projective measurement by Λ_F, in which case by Lemma 2 we have

$$\Lambda_F := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{F}(\rho, \sigma). \tag{27}$$

Since $\Lambda_F \in \mathcal{F}(\rho, \sigma)$, the inequalities in the bound (26) hold [given some arbitrary choice of $\Lambda_T \in \mathcal{T}(\rho, \sigma)$]. But since we assumed that $1 - \mathrm{F}(\rho, \sigma) = \mathrm{T}(\rho, \sigma)$, both inequalities in the bound must in fact be equalities. Focusing on the first inequality, we see that $\Lambda_F \in \mathcal{C}_1(\rho, \sigma)$. Then by Lemma 3, we have

$$\begin{aligned} \forall\, x \in \mathcal{X},\ &\langle e_x|\rho\, e_x\rangle = \langle e_x|\sigma\, e_x\rangle \vee \\ &\langle e_x|\rho\, e_x\rangle = 0 \vee \langle e_x|\sigma\, e_x\rangle = 0. \end{aligned}$$

Since ρ and σ are assumed to be invertible and all $|e_x\rangle$ are nonzero, this implies

$$\forall\, x \in \mathcal{X},\ \ \langle e_x|\rho\, e_x\rangle = \langle e_x|\sigma\, e_x\rangle. \tag{28}$$

Now, for each $x \in \mathcal{X}$, let $c_x > 0$ be the eigenvalue of M corresponding to $|e_x\rangle$. Since $\sigma = M\rho M$, Eq. (28) implies

$$\forall\, x \in \mathcal{X},\ \ \langle e_x|\rho\, e_x\rangle = c_x^2\langle e_x|\rho\, e_x\rangle.$$

Since $\rho > 0$, this implies

$$\forall\, x \in \mathcal{X}, \quad c_x = 1.$$

Thus, $M = \mathbb{1}$, which implies that $\rho = \sigma$, as needed.

Next, we show one way to prove the right-hand inequality in Eq. (17) for arbitrary $\rho, \sigma \in \mathcal{D}(\mathcal{H})$. Observe that, for any $\Lambda_T \in \mathcal{T}(\rho, \sigma)$ and $\Lambda_F \in \mathcal{F}(\rho, \sigma)$, we have

$$\begin{aligned} \mathrm{T}\,(\rho, \sigma) &= \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T \rho), \mathrm{tr}\,(\Lambda_T \sigma)) \\ &\leqslant \sqrt{1 - \mathrm{F}_c(\mathrm{tr}\,(\Lambda_T \rho), \mathrm{tr}\,(\Lambda_T \sigma))^2} \\ &\leqslant \sqrt{1 - \mathrm{F}_c(\mathrm{tr}\,(\Lambda_F \rho), \mathrm{tr}\,(\Lambda_F \sigma))^2} \\ &= \sqrt{1 - \mathrm{F}\,(\rho, \sigma)^2}, \end{aligned} \tag{29}$$

where the second line follows from Eq. (23) and the third line follows from Eq. (22).

With this, we now prove Eq. (25). Restrict attention to invertible $\rho, \sigma \in \mathcal{D}_{\mathrm{inv}}(\mathcal{H})$. First note that, if $\rho = \sigma$, then clearly $\mathrm{T}(\rho, \sigma) = [1 - \mathrm{F}(\rho, \sigma)^2]^{1/2}$, so this case is trivial. Otherwise, suppose that

$$\exists\, c \in (0, 1),\ \mathrm{spec}\,(M) = \left\{c, \frac{1}{c}\right\} \wedge [M, \rho - \sigma] = 0.$$

Then there exists a basis $\{|e_x\rangle\}_{x\in\mathcal{X}}$ for $\mathcal{H}$ that simultaneously diagonalizes M and $\rho - \sigma$. Moreover, there exists $c \in (0, 1)$ such that for all $x \in \mathcal{X}$, $M|e_x\rangle = c\,|e_x\rangle$ or $M|e_x\rangle = \frac{1}{c}\,|e_x\rangle$. Define $\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M}$. By Lemma 1, $\Lambda \in \mathcal{T}(\rho, \sigma)$, and by Lemma 2, $\Lambda \in \mathcal{F}(\rho, \sigma)$. Thus, the inequalities in the bound in (29) hold with $\Lambda_T := \Lambda =: \Lambda_F$. Now, since $\Lambda_T = \Lambda_F$, the third line in Eq. (29) becomes an equality. Moreover, for each $x \in \mathcal{X}$, we have

$$\begin{aligned} \langle e_x|\sigma\, e_x\rangle &= \langle e_x|M\rho M\, e_x\rangle \\ &= \begin{cases} c^2\langle e_x|\rho\, e_x\rangle, & M|e_x\rangle = c|e_x\rangle \\ \frac{1}{c^2}\langle e_x|\rho\, e_x\rangle, & M|e_x\rangle = \frac{1}{c}|e_x\rangle. \end{cases} \end{aligned}$$

With $b := c^2 \in (0, 1)$, we see that, by Lemma 3, $\Lambda \in \mathcal{C}_2(\rho, \sigma)$. Thus, the second line in Eq. (29) becomes an equality. Then we have

$$\mathrm{T}\,(\rho, \sigma) = \sqrt{1 - \mathrm{F}\,(\rho, \sigma)^2},$$

as needed.

Conversely, suppose that $\mathrm{T}(\rho, \sigma) = [1 - \mathrm{F}(\rho, \sigma)^2]^{1/2}$. Let $\{|f_y\rangle\}_{y\in\mathcal{X}}$ be an orthonormal eigenbasis for $\rho - \sigma$ and denote the corresponding projective measurement by Λ_T, in which case by Lemma 1 we have

$$\Lambda_T := \{|f_y\rangle\langle f_y|\}_{y\in\mathcal{X}} \in \mathcal{T}(\rho, \sigma). \tag{30}$$

Since $\Lambda_T \in \mathcal{T}(\rho, \sigma)$, the inequalities in the bound (29) hold [given some arbitrary choice of $\Lambda_F \in \mathcal{F}(\rho, \sigma)$]. But since we assumed that $\mathrm{T}(\rho, \sigma) = [1 - \mathrm{F}(\rho, \sigma)^2]^{1/2}$, both inequalities in the bound must in fact be equalities. Thus we have $\Lambda_T \in \mathcal{C}_2(\rho, \sigma)$ and $\Lambda_T \in \mathcal{F}(\rho, \sigma)$ (i.e., Λ_T must also be a fidelity-preserving measurement). Then by Lemma 2, $\{|f_y\rangle\}_{y\in\mathcal{X}}$ is an orthonormal basis for M, so $[M, \rho - \sigma] = 0$. Moreover, by Lemma 3, we have

$$\begin{aligned} &(\forall\, y \in \mathcal{X},\ \langle f_y|\rho\, f_y\rangle = \langle f_y|\sigma\, f_y\rangle) \vee \\ &\big(\forall\, y \in \mathcal{X},\ \langle f_y|\rho\, f_y\rangle = 0 \vee \langle f_y|\sigma\, f_y\rangle = 0\big) \vee \\ &\left(\begin{array}{c} \exists\, b \in (0, 1),\ \forall\, y \in \mathcal{X}, \\ \langle f_y|\sigma\, f_y\rangle = b\langle f_y|\rho\, f_y\rangle \vee \langle f_y|\sigma\, f_y\rangle = \frac{1}{b}\langle f_y|\rho\, f_y\rangle \end{array}\right). \end{aligned}$$

Now, if the first line above holds, then we appeal to a previous argument [see Eq. (28)] and conclude that $\rho = \sigma$. If the second line above holds, then we obtain a contradiction, since ρ and σ are assumed to be invertible and all $|f_y\rangle$ are nonzero. Now, suppose that the third line above holds. For each $y \in \mathcal{X}$, let $c_y > 0$ be the eigenvalue of M corresponding to $|f_y\rangle$. With $\sigma = M\rho M$, we then have

$$\begin{aligned} &\exists\, b \in (0, 1)\, \forall\, y \in \mathcal{X}, \\ &c_y^2\langle f_y|\rho\, f_y\rangle = b\langle f_y|\rho\, f_y\rangle \vee c_y^2\langle f_y|\rho\, f_y\rangle = \frac{1}{b}\langle f_y|\rho\, f_y\rangle. \end{aligned}$$

With $c := \sqrt{b} \in (0, 1)$ and $\rho > 0$, this implies

$$\exists\, c \in (0, 1)\, \forall\, y \in \mathcal{X}, \quad c_y = c \vee c_y = \frac{1}{c}.$$

Thus, every eigenvalue c_y of M is either c or $\frac{1}{c}$, so $\operatorname{spec}(M) = \{c, \frac{1}{c}\}$, as needed. ■

B. Noninvertible states

How can we generalize Theorem 2 to arbitrary density operators $\rho, \sigma \in \mathcal{D}(\mathcal{H})$? To begin addressing this question, note that Lemmas 1 and 3 already apply to arbitrary $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, but Lemma 2 applies only to invertible ρ, σ. Next, note that Theorem 2 appears to generalize readily to noninvertible ρ and σ given a characterization of $\mathcal{F}(\rho, \sigma)$ for noninvertible ρ and σ. Thus, it appears that the main difficulty in generalizing Theorem 2 is with generalizing Lemma 2 to arbitrary ρ and σ. In other words, if we can characterize the rank-1 projective measurements $\Lambda \in \mathcal{M}$ which achieve the optimum in Eq. (22) for arbitrary ρ and σ, i.e., if we can characterize $\mathcal{F}(\rho, \sigma)$ for arbitrary ρ and σ, then it seems relatively straightforward to extend our characterization of the sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{D}_{\text{inv}}(\mathcal{H})$ to a characterization of similarly defined sets $\tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2 \subseteq \mathcal{D}(\mathcal{H})$ for arbitrary ρ and σ.

Thus, we discuss possibilities for characterizing the set $\mathcal{F}(\rho, \sigma)$ for arbitrary ρ and σ. We highlight that the proofs of the variational characterization (22) in e.g., [1,15] only yield specific examples of measurements attaining the optimum, rather than characterizing the set of all such measurements, which (as we shall soon discuss) appears significantly larger for some noninvertible states. As for the proof in Ref. [4], it implicitly uses a compactness argument that also does not seem to yield a precise characterization of this set.

At first, it seems plausible that some appropriate generalization of the operator $M := \rho^{-1}\#\sigma$ may be involved in the characterization of $\mathcal{F}(\rho, \sigma)$. However, it is not immediately clear what this generalization of M might be. For example, one possibility for generalizing M is to consider pseudoinverses. If $\rho \in \mathcal{D}(\mathcal{H})$ is noninvertible, the (Moore–Penrose) pseudoinverse is informally the "inverse on the support" of ρ. That is (using ρ^{-1} to denote the pseudoinverse of ρ), $\rho^{-1} \in \mathcal{L}(\mathcal{H})$ is an operator with the property that $\rho\,\rho^{-1} = \rho^{-1}\rho = \Pi_\rho$, where Π_ρ is the projector onto the support of ρ. We could then try to define $M := \rho^{-1}\#\sigma$ using pseudoinverses. However, with this approach, it is not possible to say that the set $\mathcal{F}(\rho, \sigma)$ for arbitrary ρ, σ is just the same as in Lemma 2 except with M defined via pseudoinverses. To see this, consider the case of pure states $\rho = |\rho\rangle\langle\rho|$ and $\sigma = |\sigma\rangle\langle\sigma|$. Then with this definition of M, we would have

$$\rho^{-1}\#\sigma = |\langle\rho|\sigma\rangle||\rho\rangle\langle\rho|,$$

but

$$\sigma^{-1}\#\rho = |\langle\rho|\sigma\rangle||\sigma\rangle\langle\sigma|.$$

Now, observe that $\mathcal{F}$ is symmetric in its arguments, i.e., $\mathcal{F}(\rho, \sigma) = \mathcal{F}(\sigma, \rho)$, since the quantum and classical fidelities are both symmetric in their arguments. Thus, if Lemma 2 held for this choice of definition for M, then every $\Lambda \in \mathcal{F}(\rho, \sigma) = \mathcal{F}(\sigma, \rho)$ must be an eigenbasis of both $\rho^{-1}\#\sigma$ and $\sigma^{-1}\#\rho$. But when ρ and σ are pure states as above, this implies that $|\rho\rangle\langle\rho| \in \Lambda$ and $|\sigma\rangle\langle\sigma| \in \Lambda$, which is impossible whenever ρ and σ are distinct and nonorthogonal. This would imply that $\mathcal{F}(\rho, \sigma)$ is empty whenever ρ and σ are distinct, nonorthogonal pure states, which completely contradicts Eq. (22). [Essentially, the fundamental issue here is that defining M using pseudoinverses causes it to lose a symmetry property $\rho^{-1}\#\sigma = (\sigma^{-1}\#\rho)^{-1}$ that held for invertible operators.]

Another potential approach for generalizing M is to note that for noninvertible operators, one can choose to define the operator geometric mean as (see, e.g., Ref. [16], p. 211)

$$A\#B := \lim_{\delta\to 0^+} (A + \delta\mathbb{1})\#(B + \delta\mathbb{1}), \tag{31}$$

where the right-hand-side can be computed using the definition (20) since $A + \delta\mathbb{1}$ and $B + \delta\mathbb{1}$ are both invertible for $\delta > 0$. It can be shown that the limit in (31) indeed exists for all positive semidefinite A and B, although this definition of $A\#B$ has the drawback that it is not continuous with respect to A and B [16]. However, this approach in our context faces the difficulty that the term in our result is not $\rho\#\sigma$, but rather $\rho^{-1}\#\sigma$. Therefore, even if we were to choose some generalized definition of the operator geometric mean for noninvertible operators, it would still not be enough by itself to resolve the issue of generalizing $\rho^{-1}\#\sigma$, since ρ^{-1} is already ill defined if ρ is noninvertible.

Drawing on the above idea, however, we could still consider "δ-perturbed" versions of ρ and σ (such that the perturbed versions are invertible), and analyze the $\delta \to 0^+$ limit in the broader context of our desired result rather than the operator geometric mean specifically. We sketch the starting points of this approach here, deferring further analysis to Appendix E. To begin, consider the following states, where $d := \dim\mathcal{H}$ (here we shall define the perturbations slightly differently from (31), in order to ensure that $\rho_\delta, \sigma_\delta$ are normalized states):

$$\rho_\delta := (1-\delta)\rho + \delta\frac{\mathbb{1}}{d}, \quad \sigma_\delta := (1-\delta)\sigma + \delta\frac{\mathbb{1}}{d}. \tag{32}$$

Using the (reverse) triangle inequality for angular distance, these states can be seen to satisfy

$$|\mathrm{A}(\rho_\delta, \sigma_\delta) - \mathrm{A}(\rho, \sigma)| \leqslant \operatorname{negl}(\delta),$$

where for brevity we use the notation $\operatorname{negl}(\delta)$ to indicate any expression such that $\lim_{\delta\to 0^+}\operatorname{negl}(\delta) = 0$. In other words, the perturbations as defined in (32) only change the angular distance (and thus also the fidelity) by an amount that vanishes in the $\delta \to 0^+$ limit.

From this, we see that if, for instance, ρ and σ saturate the upper Fuchs–van de Graaf inequality, then ρ_δ and σ_δ "approximately saturate" it as well, in the sense that

$$\mathrm{T}(\rho, \sigma) - \sqrt{1 - \mathrm{F}(\rho, \sigma)^2} = 0$$
$$\Rightarrow |\mathrm{T}(\rho, \sigma) - \sqrt{1 - \mathrm{F}(\rho_\delta, \sigma_\delta)^2}| \leqslant \operatorname{negl}(\delta). \tag{33}$$

(In the above, we have only considered perturbing the fidelity term rather than the trace-distance term; the reason for this will become apparent in our more detailed analysis in Appendix E.) Following this form of reasoning, we could continue onwards and attempt to repeat the proof of Theorem 2, except with "approximate equalities" instead of equalities. We were able to make some progress with this approach, which we describe in Appendix E. However, it still does not seem sufficient to resolve the question of extending Theorem 2 to arbitrary noninvertible states, and perhaps raises the question of whether some notion of the geometric mean operator is even the "right" object to consider in this characterization. We also remark that the above considerations seem to suggest the main challenges for noninvertible states mostly arise when ρ is noninvertible—at a high level, it seems that the proofs in the preceding sections should basically carry through for noninvertible σ as long as ρ is still invertible.

To end off, we highlight what seems to be a significant broad obstacle in generalizing our characterization of $\mathcal{F}(\rho,\sigma)$ to arbitrary ρ and σ, by presenting the characterization of this set for the case of pure states $\rho = |\rho\rangle\langle\rho|$ and $\sigma = |\sigma\rangle\langle\sigma|$. While it is possible to work through the proof of Lemma 2 to study this special case, an easier approach is via direct computation. For any $\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M}$, we have

$$\begin{aligned}
\mathrm{F}(\rho,\sigma) &= |\langle\rho|\sigma\rangle| \\
&= \left|\sum_{x\in\mathcal{X}} \langle\rho|e_x\rangle\langle e_x|\sigma\rangle\right| \\
&\leqslant \sum_{x\in\mathcal{X}} |\langle\rho|e_x\rangle\langle e_x|\sigma\rangle| \\
&= \mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)).
\end{aligned}$$

Now, equality holds in the above if and only if the triangle inequality in the third line is a strict equality. This occurs if and only if there exists $\theta \in \mathbb{R}$ such that for every $x \in \mathcal{X}$, $\arg\langle\rho|e_x\rangle\langle e_x|\sigma\rangle \equiv \theta \mod 2\pi \vee \langle\rho|e_x\rangle\langle e_x|\sigma\rangle = 0$. Thus, for pure ρ and σ, we have

$$\begin{aligned}
&\mathcal{F}(|\rho\rangle\langle\rho|, |\sigma\rangle\langle\sigma|) \\
&\quad = \{\Lambda \in \mathcal{M} \mid \exists\, \theta \in \mathbb{R}, \forall\, x \in \mathcal{X}, \\
&\qquad \arg\langle e_x|\sigma\rangle - \arg\langle e_x|\rho\rangle \equiv \theta \mod 2\pi \vee \\
&\qquad \langle e_x|\sigma\rangle = 0 \vee \langle e_x|\rho\rangle = 0\}.
\end{aligned}$$

Note that the global phases of the representative state vectors $|\rho\rangle, |\sigma\rangle$ can be chosen arbitrarily, as the description of the above set is invariant under such changes of phase.

With this, we see that $\mathcal{F}(\rho,\sigma)$ is a much larger set in the pure-state case than in the invertible case: in the latter case, all $\Lambda \in \mathcal{F}(\rho,\sigma)$ are essentially equivalent up to degeneracy in the spectral decomposition of M, whereas in the former case, we have for instance that *any* orthonormal basis in which all the components $\langle e_x|\rho\rangle$, $\langle e_x|\sigma\rangle$ are real and non-negative yields a measurement in $\mathcal{F}(\rho,\sigma)$ (note that it is easy to construct examples of pure states $|\rho\rangle$ and $|\sigma\rangle$ such that many orthonormal bases do have this property). Thus, it seems unclear how to generalize our characterization of $\mathcal{F}(\rho,\sigma)$ to noninvertible states, since any such generalization must capture the special case above for pure ρ and σ.

IV. CONCLUSION

In this work, we derived a continuity bound for the conditional entropy of quantum-classical states with respect to angular distance. This bound satisfies both of conditions 1 and 2, which are desirable in many applications to quantum key distribution [7,9]. However, in those applications, a continuity bound for classical-quantum states is usually required. Further work is thus needed to extend our result to classical-quantum states. Numerical evidence suggests that our bound in Eq. (9) may indeed hold for such states, and with minor modifications, our bound may also be valid for fully quantum states.

To find such a generalization, one approach could be to consider the large body of work on entropic continuity bounds in terms of trace distance, for instance, Refs. [6,17–20], and study whether any of the proof approaches in those works could be modified to use angular distance instead. To begin, the works [17,18] used techniques from majorization theory to prove entropic continuity bounds, including for families of Rényi entropies. However, those techniques do not seem straightforward to apply when the conditioning system is quantum. For such scenarios, continuity bounds were derived in Ref. [19] for Rényi entropies, and a recent work [20] proved an "almost locally affine" property of the relative entropy that (among other results) reproduces the bound [6]. However, qualitatively speaking, the approaches in those works seem to rely on studying "additive perturbations" to the states ρ, σ, which are naturally related to trace distance but seem more difficult to express in terms of angular distance.

We also note the observation in Ref. [21] that continuity bounds for fully classical conditional entropies can often be quite "generically" extended to the quantum-classical case (essentially, whenever the distance measure satisfies a data-processing inequality), but such a generic extension for the classical-quantum or fully quantum cases seems to require new ideas or techniques. If this is indeed so, extending our result to cover those latter cases would require exploiting some specific property of angular distance that is not shared by trace distance, for instance the characterization via Uhlmann's theorem.

In the second part of our work, to relate previous continuity bounds based on trace distance to our result based on angular distance, we explored the relationship between trace distance and angular distance via the Fuchs–van de Graaf inequalities. In particular, we derived necessary and sufficient conditions for invertible states to saturate either side of the Fuchs–van de Graaf inequalities relating the trace distance and fidelity. We remark that this may have independent applications in other topics, such as computing keyrates for QKD; we briefly outline this in Appendix F.

In addition, we showed that generalizing our result to noninvertible states appears nontrivial, and that this generalization is closely related to characterizing the set of rank-1 projective measurements which preserve the fidelity. Future work could continue by generalizing Lemma 2 and Theorem 2 to noninvertible states; it appears that generalizing Lemma 2 is the difficult part, whereas generalizing Theorem 2 after that seems to be relatively straightforward.

We note that an alternative approach for such an analysis could be to utilize a proof of the upper Fuchs–van de Graaf inequality via Uhlmann's theorem, as presented in, e.g., Ref. [4]. To give a high-level overview, that approach would yield (via the discussion at the beginning of Sec. III A) that the set of states saturating the upper Fuchs–van de Graaf inequality is precisely the set of states such that the Uhlmann purifications have the same trace distance as the original states. However, the construction of the Uhlmann purification involves a polar decomposition that appears similar to the one which arises in the proof of Lemma 2, and it seems unclear whether it can be analyzed more fruitfully.

ACKNOWLEDGMENTS

We thank Norbert Lütkenhaus, Joseph M. Renes, and Jinzhao Wang for helpful feedback and discussions. The work has been performed at the Institute for Quantum Computing, at the University of Waterloo, which is supported by Innovation, Science, and Economic Development Canada. The research has been supported by NSERC under the Discovery Grants Program, Grant No. 341495. The NSERC CRD No. CRDPJ 522308-17. The Huawei Technologies Canada Grant No. YBN2017010034A1. Numerical data were generated using *Mathematica*® 12.1.

APPENDIX A: CONVERSIONS BETWEEN CONTINUITY BOUNDS

Using the Fuchs–van de Graaf inequalities in Eq. (17), we can convert continuity bounds in trace distance to continuity bounds in angular distance and vice versa. For states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ with trace distance $T := \mathrm{T}(\rho, \sigma)$ and angular distance $A := \mathrm{A}(\rho, \sigma)$, the upper Fuchs–van de Graaf inequality gives $T \leqslant \sin A$. However, as illustrated in the calculations at the beginning of Sec. III, combining this with the previously known bounds (4) and (5) in trace distance yields results which scale badly with A as compared with our result.

In the reverse direction, the lower Fuchs–van de Graaf inequality gives $A \leqslant \arccos(1 - T)$. Plugging this into our continuity bound (9) yields an upper bound of $u(d_A) \arccos(1 - T)$ on the difference in conditional entropies. At small T, we can approximate this as

$$u(d_A) \arccos{(1 - T)} \approx u(d_A)\sqrt{2T}.$$

For comparison, in the bound (4) for the unconditioned entropies (which corresponds to the $d_B = 1$ case in our bound), we can use an approximation $\mathrm{h}(x) \lesssim 2\sqrt{x}$ at small x for the binary entropy function to obtain

$$\begin{aligned} T \ln{(d_A - 1)} + \mathrm{h}\,(T) &\lesssim T \ln{(d_A - 1)} + 2\sqrt{T} \\ &< [\ln{(d_A - 1)} + 2]\sqrt{T}. \end{aligned}$$

From the first line, we see that at small T, both the bound obtained via our result and the bound from (4) basically scale on the order of $\sqrt{T}$; however, the coefficient in front of that term is quite different. The second line above reveals that our bound cannot outperform the bound (4) (as should be expected, since the latter is tight): for $d_A \leqslant 3$ one can verify numerically that $\ln(d_A - 1) + 2 \leqslant u(d_A)$, while for $d_A \geqslant 4$ we have $\ln(d_A - 1) + 2 \leqslant \ln(d_A - 1) + \ln(d_A) \leqslant 2\ln(d_A) \leqslant u(d_A)$.

As for the bound (5) for conditional entropies (which is nearly tight but not exactly so [6]), we first highlight that for quantum-classical states, the prefactor of two on the first term in that bound can be omitted. Since our bound only holds for quantum-classical states, we can compare it to that version. However, with that change, the modified version of (5) is quite close to the unconditioned-entropy bound (4) at small T, and hence (for small T, at least) our result also cannot provide an improvement over (5) via a conversion of this form.

APPENDIX B: CHALLENGES IN PURIFICATION-BASED ARGUMENTS

At first sight, it might appear that, since we are considering angular distance as our metric, it would be useful to consider purifications, since we could, for instance, apply results such as Uhlmann's theorem. Unfortunately, this appears to encounter difficulties regarding the conditional entropies, as we shall now describe.

We first observe that for fully quantum states, the following issue arises. Take any $\rho, \sigma \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$ such that $\rho_A = \sigma_A$. If we purify ρ and σ to some pure states ρ_{ABR} and σ_{ABR}, then we have $\mathrm{H}(A|BR)_\rho = -\mathrm{H}(A)_\rho = -\mathrm{H}(A)_\sigma = \mathrm{H}(A|BR)_\sigma$, i.e., the difference in conditional entropies $\mathrm{H}(A|BR)$ is zero. Hence these conditional entropies cannot give us any information about the difference in the original conditional entropies $\mathrm{H}(A|B)$. If we instead consider the entropies $\mathrm{H}(AR|B)$, an analogous problem arises whenever $\rho_B = \sigma_B$.

For classical-quantum states, we could consider a modified version of this approach, by taking "individual purifications" of the conditional quantum states on the $\mathcal{H}^B$ systems. Specifically, for a state ρ_{AB} of the form $\rho_{AB} = \sum_a p_a |a\rangle\langle a| \otimes \rho_B^{(a)}$, we can consider an extension $\rho_{ABR} = \sum_a p_a |a\rangle\langle a| \otimes \rho_{BR}^{(a)}$ such that each $\rho_{BR}^{(a)}$ is a purification of $\rho_B^{(a)}$. For simplicity, let us suppose here that the other state σ_{AB} satisfies $\sigma_A = \rho_A$ (and is also classical-quantum). In that case, by an appropriate application of Uhlmann's theorem, it can be shown that, for any extension of ρ_{AB} as described above, we can construct an analogous extension of σ_{AB} with the property $\mathrm{F}(\rho_{AB}, \sigma_{AB}) = \mathrm{F}(\rho_{ABR}, \sigma_{ABR})$. This seems promising because it preserves the angular distance—if we could furthermore show that the difference in conditional entropies is nondecreasing under some such extension, this would imply that it suffices to consider pure conditional states in this scenario, simplifying the analysis. (In fact, if that were true, it would already be sufficient to yield a continuity bound with the desired properties by noting that the pure conditional states would all be supported on a subspace of dimension at most $2d_A$, although the resulting Lipschitz constant might be suboptimal.)

However, this encounters the following obstacle: one can show[2] that there exist states ρ_{AB} such that for some $\delta > 0$, *any* extension ρ_{ABR} in the above sense satisfies $\mathrm{H}(A|BR)_\rho < \mathrm{H}(A|B)_\rho - \delta$; i.e., the conditional entropy of the extension is bounded away from the original value by a constant. In that case, if we take any classical-quantum σ_{AB} with $\mathrm{H}(A|B)_\sigma = 0$ [in which case any extension in the above sense must have $\mathrm{H}(A|BR)_\sigma = 0$ as well, due to strong subadditivity], we have

$$\begin{aligned} &|\mathrm{H}(A|BR)_\rho - \mathrm{H}(A|BR)_\sigma| \\ &\quad = \mathrm{H}\,(A|BR)_\rho < \mathrm{H}(A|B)_\rho - \delta \\ &\quad = |\mathrm{H}(A|B)_\rho - \mathrm{H}(A|B)_\sigma| - \delta, \end{aligned}$$

[2] To outline the key ideas, in Ref. [22] the following bound was derived [see Eq. (21) of that work]: $\mathrm{H}(A|B)_\rho \geqslant 1 - \mathrm{h}([1 - \mathrm{F}[\rho_B^{(0)}, \rho_B^{(1)}]]/2)$ for ρ_{AB} of the form $\rho_{AB} = \sum_{a\in\{0,1\}} (1/2)|a\rangle\langle a| \otimes \rho_B^{(a)}$, with equality holding when the two conditional states $\rho_B^{(a)}$ are pure. Take any ρ_{AB} such that this bound is a strict inequality, then observe that any purifications of the conditional states must satisfy $\mathrm{F}(\rho_B^{(0)}, \rho_B^{(1)}) \geqslant \mathrm{F}(\rho_{BR}^{(0)}, \rho_{BR}^{(1)})$, and use the fact that the bound from [22] becomes an equality for pure $\rho_{BR}^{(a)}$.

or in other words, the difference in entropies $\mathrm{H}(A|BR)$ of the extensions is strictly smaller than the original difference (more precisely, bounded away from it by a constant amount). This is problematic for the goal of using these extensions to bound the original entropy difference. While this obstacle may not be impossible to overcome, it does suggest that there may not be a straightforward proof using this approach.

APPENDIX C: PROOFS OF LEMMAS

1. Proof of Lemma 1

Proof. Let $\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M}$ be arbitrary. We first prove that

$$\mathrm{T}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)) \leqslant \mathrm{T}\,(\rho,\sigma).$$

To see this, consider a spectral decomposition

$$\rho-\sigma = \sum_{y\in\mathcal{X}} \lambda_y |f_y\rangle\langle f_y|$$

of $\rho-\sigma$, where the $\lambda_y \in \mathbb{R}$ and the $|f_y\rangle$ form an orthonormal basis for $\mathcal{H}$. Define the positive and negative parts of $\rho-\sigma$ as

$$P := \sum_{y\in\mathcal{X},\lambda_y>0} \lambda_y |f_y\rangle\langle f_y|,$$

$$Q := \sum_{y\in\mathcal{X},\lambda_y<0} |\lambda_y| |f_y\rangle\langle f_y|,$$

respectively. Note that $\rho-\sigma = P-Q$ and $|\rho-\sigma| := [(\rho-\sigma)^\dagger(\rho-\sigma)]^{1/2} = P+Q$. Then for any $x\in\mathcal{X}$, we have

$$\begin{aligned}|\langle e_x|(\rho-\sigma)e_x\rangle| &= |\langle e_x|(P-Q)e_x\rangle|\\ &\leqslant \langle e_x|(P+Q)e_x\rangle\\ &= \langle e_x||\rho-\sigma|e_x\rangle,\end{aligned}\tag{C1}$$

where in the second line we used that $P,\ Q \geqslant 0$. Thus, we have

$$\begin{aligned}\mathrm{T}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)) &= \frac{1}{2}\sum_{x\in\mathcal{X}} |\langle e_x|\rho\, e_x\rangle - \langle e_x|\sigma\, e_x\rangle|\\ &\leqslant \frac{1}{2}\sum_{x\in\mathcal{X}} \langle e_x||\rho-\sigma|e_x\rangle\\ &= \frac{1}{2}\,\mathrm{tr}\,(|\rho-\sigma|)\\ &= \mathrm{T}\,(\rho,\sigma),\end{aligned}$$

as needed. Now, in the second line above, we applied Eq. (C1) for each $x\in\mathcal{X}$. Thus, equality holds in the above if and only if

$$\forall\, x\in\mathcal{X},\quad |\langle e_x|(P-Q)e_x\rangle| = \langle e_x|(P+Q)e_x\rangle.$$

Since $P, Q \geqslant 0$, this occurs if and only if

$$\forall\, x\in\mathcal{X},\quad \langle e_x|Pe_x\rangle = 0 \vee \langle e_x|Qe_x\rangle = 0.$$

Since any positive semidefinite $A\in\mathcal{L}(\mathcal{H})$ satisfies $\langle e_x|Ae_x\rangle = 0 \iff |e_x\rangle \in \ker A$, the above is equivalent to

$$\forall\, x\in\mathcal{X},\quad |e_x\rangle \in \ker P \vee |e_x\rangle \in \ker Q,$$

as needed. To see that $\mathcal{T}(\rho,\sigma)$ is nonempty, consider $\tilde{\Lambda} := \{|f_y\rangle\langle f_y|\}_{y\in\mathcal{X}} \in \mathcal{M}$. Clearly, this satisfies

$$\forall\, y\in\mathcal{X},\quad |f_y\rangle \in \ker P \vee |f_y\rangle \in \ker Q,$$

so $\tilde{\Lambda}$ is an optimizing projective measurement, as needed. ■

2. Proof of Lemma 2

Proof. Let $\Lambda := \{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}} \in \mathcal{M}$ be arbitrary. We first prove that

$$\mathrm{F}\,(\rho,\sigma) \leqslant \mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)).$$

To see this, note that by the polar decomposition, there exists a unitary U on $\mathcal{H}$ such that

$$\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}} = \sqrt{\rho}\,\sqrt{\sigma}\,U.\tag{C2}$$

Then we have

$$\begin{aligned}\mathrm{F}\,(\rho,\sigma) &= \|\sqrt{\rho}\,\sqrt{\sigma}\|_1\\ &= \mathrm{tr}(\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}})\\ &= \mathrm{tr}(\sqrt{\rho}\,\sqrt{\sigma}\,U)\\ &= \sum_{x\in\mathcal{X}} \mathrm{tr}(\sqrt{\rho}\,|e_x\rangle\langle e_x|\sqrt{\sigma}\,U)\\ &= \left|\sum_{x\in\mathcal{X}} \mathrm{tr}\left(\sqrt{\rho}\,|e_x\rangle\langle e_x|\sqrt{\sigma}\,U\right)\right|,\end{aligned}$$

where the last line follows since $\mathrm{F}(\rho,\sigma) \geqslant 0$. Applying the triangle inequality followed by the Cauchy–Schwarz inequality, we continue with

$$\begin{aligned}\mathrm{F}\,(\rho,\sigma) &= \left|\sum_{x\in\mathcal{X}} \mathrm{tr}(\sqrt{\rho}\,|e_x\rangle\langle e_x|\sqrt{\sigma}\,U)\right|\\ &\leqslant \sum_{x\in\mathcal{X}} |\,\mathrm{tr}(\sqrt{\rho}\,|e_x\rangle\langle e_x||e_x\rangle\langle e_x|\sqrt{\sigma}\,U)|\\ &\leqslant \sum_{x\in\mathcal{X}} \sqrt{\mathrm{tr}(|e_x\rangle\langle e_x|\,\rho)}\sqrt{\mathrm{tr}(|e_x\rangle\langle e_x|\,\sigma)}\\ &= \mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho), \mathrm{tr}\,(\Lambda\sigma)),\end{aligned}\tag{C3}$$

as needed. Now, equality occurs in the above if and only if the triangle inequality and the Cauchy–Schwarz inequality are both saturated. This occurs if and only if each term $\mathrm{tr}(\sqrt{\rho}\,|e_x\rangle\langle e_x|\sqrt{\sigma}\,U)$ either is of a fixed complex phase or is equal to zero, and the set $\{|e_x\rangle\langle e_x|\sqrt{\rho},\ |e_x\rangle\langle e_x|\sqrt{\sigma}\,U\} \subseteq \mathcal{L}(\mathcal{H})$ is linearly dependent for all $x\in\mathcal{X}$. To continue, observe that $\sum_{x\in\mathcal{X}} \mathrm{tr}(\sqrt{\rho}\,|e_x\rangle\langle e_x|\sqrt{\sigma}\,U) = \mathrm{F}(\rho,\sigma) \geqslant 0$. In addition, observe that $|e_x\rangle\langle e_x|\sqrt{\rho} \neq 0$ and $|e_x\rangle\langle e_x|\sqrt{\sigma}\,U \neq 0$, since ρ and σ are assumed invertible and $|e_x\rangle \neq 0$. This implies that equality in Eq. (C3) occurs if and only if the terms $\mathrm{tr}(\sqrt{\rho}\,|e_x\rangle\langle e_x|\sqrt{\sigma}\,U)$ are all non-negative and $|e_x\rangle\langle e_x|\sqrt{\rho} = c_x|e_x\rangle\langle e_x|\sqrt{\sigma}\,U$ for all $x\in\mathcal{X}$, where $c_x\in\mathbb{C}\setminus\{0\}$. Equivalently, equality holds if and only if

$$\begin{aligned}&\forall\, x\in\mathcal{X},\quad \exists\, c_x\in\mathbb{C}\setminus\{0\},\\ &\langle e_x|\sqrt{\sigma}\,U\sqrt{\rho}\,e_x\rangle \geqslant 0 \wedge\\ &|e_x\rangle\langle e_x|\sqrt{\rho} = c_x|e_x\rangle\langle e_x|\sqrt{\sigma}\,U.\end{aligned}$$

Since ρ is assumed invertible, this is equivalent to

$$\begin{aligned}&\forall\, x\in\mathcal{X},\quad \exists\, c_x\in\mathbb{C}\setminus\{0\},\\ &\quad\langle e_x|\rho^{-\frac{1}{2}}\sqrt{\rho}\,\sqrt{\sigma}\,U\,\rho^{-\frac{1}{2}}\rho\, e_x\rangle\geqslant 0\,\wedge\\ &\quad\langle e_x| = c_x\langle e_x|\rho^{-\frac{1}{2}}\sqrt{\rho}\,\sqrt{\sigma}\,U\,\rho^{-\frac{1}{2}}.\end{aligned}$$

Recalling Eq. (C2) and the definition of the operator geometric mean $M := \rho^{-1}\#\sigma = \rho^{-\frac{1}{2}}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}\,\rho^{-\frac{1}{2}}$, this is equivalent to

$$\begin{aligned}&\forall\, x\in\mathcal{X},\quad \exists\, c_x\in\mathbb{C}\setminus\{0\},\\ &\quad\langle e_x|M\rho\, e_x\rangle\geqslant 0\wedge\langle e_x|M=\frac{1}{c_x}\langle e_x|.\end{aligned}$$

Since $M=M^{\dagger}$, this is equivalent to

$$\begin{aligned}&\forall\, x\in\mathcal{X},\quad \exists\, c_x\in\mathbb{C}\setminus\{0\},\\ &\quad\overline{c_x}\langle e_x|M\rho M e_x\rangle\geqslant 0\wedge M|e_x\rangle=\frac{1}{\overline{c_x}}|e_x\rangle.\end{aligned}$$

But since $M\rho M=\sigma$ and $\sigma>0$ (since σ is assumed invertible), this is equivalent to

$$\forall\, x\in\mathcal{X},\quad \exists\, c_x>0,\quad M|e_x\rangle=\frac{1}{c_x}|e_x\rangle.$$

Thus, since $M>0$, the eigenvalues of M are all positive, we see that equality in Eq. (C3) holds if and only if $\{|e_x\rangle\}_{x\in\mathcal{X}}$ is an eigenbasis for M, as needed. Clearly, an eigenbasis for M exists, so $\mathcal{F}(\rho,\sigma)$ is nonempty, as needed. ■

3. Proof of Lemma 3

Proof. For notational simplicity, we work with arbitrary probability distributions p and q on a finite alphabet $\mathcal{X}$. We first prove the left-hand inequality in Eq. (23). Observe that

$$\begin{aligned}1-\mathrm{F}_c(p,q)&=\frac{1}{2}\sum_{x\in\mathcal{X}}(p(x)+q(x)-2\sqrt{p(x)q(x)})\\ &=\frac{1}{2}\sum_{x\in\mathcal{X}}|\sqrt{p(x)}-\sqrt{q(x)}|^2\\ &\leqslant\frac{1}{2}\sum_{x\in\mathcal{X}}|\sqrt{p(x)}-\sqrt{q(x)}||\sqrt{p(x)}+\sqrt{q(x)}|\\ &=\frac{1}{2}\sum_{x\in\mathcal{X}}|p(x)-q(x)|\\ &=\mathrm{T}_c(p,q),\end{aligned}$$

as needed. Now, equality holds in the above if and only if

$$\begin{aligned}\forall\, x\in\mathcal{X},\quad &|\sqrt{p(x)}-\sqrt{q(x)}||\sqrt{p(x)}-\sqrt{q(x)}|\\ &=|\sqrt{p(x)}-\sqrt{q(x)}||\sqrt{p(x)}+\sqrt{q(x)}|,\end{aligned}$$

which occurs if and only if

$$\begin{aligned}\forall\, x\in\mathcal{X},\quad &|\sqrt{p(x)}-\sqrt{q(x)}|=0\\ &\vee|\sqrt{p(x)}-\sqrt{q(x)}|=|\sqrt{p(x)}+\sqrt{q(x)}|,\end{aligned}$$

which occurs if and only if

$$\forall\, x\in\mathcal{X},\quad p(x)=q(x)\vee p(x)=0\vee q(x)=0,$$

as needed. Next, we prove the right-hand inequality in Eq. (23). Observe that

$$\begin{aligned}&\mathrm{T}_c(p,q)^2\\ &\quad=\left(\frac{1}{2}\sum_{x\in\mathcal{X}}|p(x)-q(x)|\right)^2\\ &\quad=\frac{1}{4}\left(\sum_{x\in\mathcal{X}}|\sqrt{p(x)}-\sqrt{q(x)}||\sqrt{p(x)}+\sqrt{q(x)}|\right)^2\\ &\quad\leqslant\frac{1}{4}\sum_{x\in\mathcal{X}}(\sqrt{p(x)}-\sqrt{q(x)})^2\sum_{x\in\mathcal{X}}(\sqrt{p(x)}+\sqrt{q(x)})^2\\ &\quad=\frac{1}{4}[2-2\,\mathrm{F}_c\,(p,q)][2+2\,\mathrm{F}_c\,(p,q)]\\ &\quad=1-\mathrm{F}_c\,(p,q)^2,\end{aligned}$$

as needed. Note that, in the third line above, we applied the Cauchy–Schwarz inequality with

$$\begin{aligned}u&:=(|\sqrt{p(x)}-\sqrt{q(x)}|)_{x\in\mathcal{X}}\in\mathbb{R}^d,\\ v&:=(|\sqrt{p(x)}+\sqrt{q(x)}|)_{x\in\mathcal{X}}\in\mathbb{R}^d.\end{aligned}$$

Thus, equality holds in the above if and only if u and v saturate the Cauchy–Schwarz inequality, which occurs if and only if $\{u,v\}\subseteq\mathbb{R}^d$ is linearly dependent. Since $v\neq 0$, this is equivalent to $u\in\operatorname{Span}\{v\}$, i.e., $\exists\, a\in\mathbb{R}$, $u=a\,v$. Since $u\geqslant 0$ and $v>0$, this is equivalent to $\exists\, a\geqslant 0$, $u=a\,v$, i.e.,

$$\begin{aligned}&\exists\, a\geqslant 0,\quad\forall\, x\in\mathcal{X},\\ &\qquad|\sqrt{p(x)}-\sqrt{q(x)}|=a|\sqrt{p(x)}+\sqrt{q(x)}|.\end{aligned}$$

Breaking into cases, this is equivalent to

$$\begin{aligned}&\exists\, a\geqslant 0,\quad\forall\, x\in\mathcal{X},\\ &\qquad\sqrt{p(x)}-\sqrt{q(x)}=a(\sqrt{p(x)}+\sqrt{q(x)})\,\vee\\ &\qquad\sqrt{q(x)}-\sqrt{p(x)}=a(\sqrt{p(x)}+\sqrt{q(x)}).\end{aligned}$$

Rearranging, this is equivalent to

$$\begin{aligned}&\exists\, a\geqslant 0,\quad\forall\, x\in\mathcal{X},\\ &\qquad\frac{1-a}{1+a}\sqrt{p(x)}=\sqrt{q(x)}\vee\frac{1+a}{1-a}\sqrt{p(x)}=\sqrt{q(x)}.\end{aligned}$$

Now, since $\sqrt{p(x)},\sqrt{q(x)}\geqslant 0$ for all $x\in\mathcal{X}$, we can restrict $a\in[0,1]$, so the above is equivalent to

$$\begin{aligned}&\exists\, a\in[0,1],\quad\forall\, x\in\mathcal{X},\\ &\qquad\frac{1-a}{1+a}\sqrt{p(x)}=\sqrt{q(x)}\vee\frac{1+a}{1-a}\sqrt{p(x)}=\sqrt{q(x)}.\end{aligned}$$

Treating $a=0$ and $a=1$ separately, this is equivalent to

$$\begin{aligned}&(\forall\, x\in\mathcal{X},\ p(x)=q(x))\vee(\forall\, x\in\mathcal{X},\ p(x)=0\vee q(x)=0)\,\vee\\ &\left(\begin{array}{l}\exists\, a\in(0,1),\ \forall\, x\in\mathcal{X},\\ q(x)=\left(\frac{1-a}{1+a}\right)^2p(x)\vee q(x)=\left(\frac{1+a}{1-a}\right)^2p(x)\end{array}\right).\end{aligned}$$

Since $p, q \geqslant 0$, this is equivalent to

$$p = q \vee p \cdot q = 0 \vee \left(\exists\, a \in (0,1),\ \forall\, x \in \mathcal{X}, \right.$$
$$\left. q(x) = b(a)\, p(x) \vee q(x) = \frac{1}{b(a)}\, p(x) \right),$$

where we defined $b(a) := (\frac{1-a}{1+a})^2 \in (0,1)$ for $a \in (0,1)$. Note that $b(a) : (0,1) \to (0,1)$ is a bijection. Thus, the above is equivalent to

$$p = q \vee p \cdot q = 0 \vee \left(\exists\, b \in (0,1),\ \forall\, x \in \mathcal{X}, \right.$$
$$\left. q(x) = b\, p(x) \vee q(x) = \frac{1}{b}\, p(x) \right),$$

as needed. Thus, with $p := (\langle e_x | \rho\, e_x\rangle)_{x\in\mathcal{X}}$ and $q := (\langle e_x | \sigma\, e_x\rangle)_{x\in\mathcal{X}}$, the lemma holds. ■

APPENDIX D: RANDOM DENSITY OPERATORS

1. Random quantum-classical states

Our goal in this section is to randomly sample quantum-classical states $\rho \in \mathcal{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$ of the form

$$\rho = \sum_{k=1}^{d_B} \alpha_k \rho_k \otimes |f_k\rangle\langle f_k|, \tag{D1}$$

where $\rho_k \in \mathcal{D}(\mathcal{H}^A)$, $\alpha_k \geqslant 0$ satisfy $\sum_{k=1}^{d_B} \alpha_k = 1$, and $|f_k\rangle$ form an orthonormal basis for $\mathcal{H}^B$. To begin, we randomly sample the α_k from the $(d_B - 1)$-dimensional simplex spanned by the standard basis vectors $(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1) \in \mathbb{R}^{d_B}$. Then, we let $\{|e_j\rangle\}_j$ be a standard basis for $\mathcal{H}^A$ and $\{|f_k\rangle\}_k$ be a standard basis for $\mathcal{H}^B$. Next, for each k, we randomly choose the eigenvalues of ρ_k from the $(d_A - 1)$-dimensional simplex spanned by the standard basis vectors $(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1) \in \mathbb{R}^{d_A}$, and we place those random eigenvalues on the main diagonal of a matrix $D_k \in \mathcal{L}(\mathcal{H}^A)$ (written with respect to the standard basis $\{|e_j\rangle\}_j$). To randomize the eigenvectors of the ρ_k, we pick random unitaries U_k according to a Haar-uniform distribution, and we conjugate the D_k by the U_k to obtain the ρ_k. That is, we set $\rho_k := U_k D_k U_k^\dagger$ for each k. Applying Eq. (D1) then yields the desired random quantum-classical states as used in Fig. 1. The *Mathematica* code to implement the above procedure can be found on Github.

2. Random classical states with fixed angular distance

Our goal in this section is to randomly sample classical (i.e., commuting) states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ with a prescribed angular distance $\mathrm{A}(\rho, \sigma) = A$ for any $A \gtrsim 0$. To do this, consider any points $r, s \in \mathbb{R}^d$ on the unit $(d-1)$-sphere such that $r, s \geqslant 0$, where $d := \dim \mathcal{H}$. Now, let ρ and σ be commuting states with eigenvalues $\rho_j := r_j^2$ and $\sigma_j := s_j^2$, respectively, where $j = 1, \ldots, d$. Then,

$$\mathrm{A}(\rho, \sigma) = \arccos \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \arccos \sum_{j=1}^{d} \sqrt{\rho_j \sigma_j}$$
$$= \arccos \sum_{j=1}^{d} r_j s_j = \arccos r \cdot s,$$

where the last line is just the angular distance between r and s. Thus, by the above construction, it remains to randomly sample such points r, s. To do this, we randomly pick $\tilde{r}, \tilde{s}$ on the unit $(d-1)$-sphere. We then let $r = \mathrm{abs}(\tilde{r})$, where $\mathrm{abs}(\cdot)$ denotes component-wise absolute value. Next, we rotate r towards $\tilde{s}$ by A radians to obtain s, where $A \gtrsim 0$ is fixed. For A close to zero, this procedure generates points r, s in the positive hyperoctant of the unit $(d-1)$-sphere with high probability; s is very unlikely to lie outside the positive hyperoctant for A close enough to zero, we simply reject the few cases where s happens to lie outside the positive hyperoctant. As explained above, this yields random classical states ρ, σ at a fixed angular distance A, as used in Fig. 2. The *Mathematica* code to implement the above procedure can be found on Github.

APPENDIX E: PERTURBATION ARGUMENT

Consider any arbitrary (i.e., possibly noninvertible) ρ, σ saturating the upper Fuchs–van de Graaf inequality. We continue our analysis onwards from Eq. (33), attempting to apply a similar argument as in the proof of Theorem 2. Take any $\Lambda_T \in \mathcal{T}(\rho, \sigma)$ (note that this step is essentially why we avoided perturbing the trace-distance term when writing Eq. (33)—if we were to instead take $\Lambda_T \in \mathcal{T}(\rho_\delta, \sigma_\delta)$, then Λ_T would implicitly depend on δ, which poses some challenges in our subsequent analysis). Now note that if we were to perform this measurement on the states (ρ, σ), the trace distance between the resulting distributions is "close" to the trace distance between the distributions that would be obtained by performing this measurement on $(\rho_\delta, \sigma_\delta)$ instead:

$$\begin{aligned}
&|\mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho), \mathrm{tr}\,(\Lambda_T\sigma)) - \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho_\delta), \mathrm{tr}\,(\Lambda_T\sigma_\delta))| \\
&\quad\leqslant |\mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho), \mathrm{tr}\,(\Lambda_T\sigma)) - \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho_\delta), \mathrm{tr}\,(\Lambda_T\sigma))| \\
&\qquad + |\mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho_\delta), \mathrm{tr}\,(\Lambda_T\sigma)) - \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho_\delta), \mathrm{tr}\,(\Lambda_T\sigma_\delta))| \\
&\quad\leqslant \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho), \mathrm{tr}\,(\Lambda_T\rho_\delta)) + \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\sigma), \mathrm{tr}\,(\Lambda_T\sigma_\delta)) \\
&\quad\leqslant \mathrm{T}\,(\rho, \rho_\delta) + \mathrm{T}\,(\sigma, \sigma_\delta) =: f(\delta),
\end{aligned}$$

where the second inequality holds due to the reverse triangle inequality for trace distance, and the function $f(\delta)$ in the last line satisfies $\lim_{\delta\to 0^+} f(\delta) = 0$.

With this, we can obtain the following chain of inequalities by following the same arguments as in the derivation of Eq. (29):

$$\begin{aligned}
\mathrm{T}\,(\rho, \sigma) - f(\delta) &= \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho), \mathrm{tr}\,(\Lambda_T\sigma)) - f(\delta) \\
&\leqslant \mathrm{T}_c(\mathrm{tr}\,(\Lambda_T\rho_\delta), \mathrm{tr}\,(\Lambda_T\sigma_\delta)) \\
&\leqslant \sqrt{1 - \mathrm{F}_c(\mathrm{tr}\,(\Lambda_T\rho_\delta), \mathrm{tr}\,(\Lambda_T\sigma_\delta))^2} \\
&\leqslant \sqrt{1 - \mathrm{F}\,(\rho_\delta, \sigma_\delta)^2} \\
&\leqslant \mathrm{T}\,(\rho, \sigma) + g(\delta),
\end{aligned}$$

where the first inequality follows since $\mathrm{T}(\rho_\delta, \sigma_\delta) \leqslant \mathrm{T}(\rho, \sigma)$, and in the last line the function $g(\delta)$ represents the $\mathrm{negl}(\delta)$ bound in Eq. (33). Observe that the first and last expressions in the above chain of inequalities differ by only $f(\delta) + g(\delta)$. This implies that for each individual inequality in the chain, the difference between the two sides of each inequality is also at most $f(\delta) + g(\delta)$, which is a negligible function $\mathrm{negl}(\delta)$. From this fact, and the continuity of the function $(1 - x^2)^{1/2}$, we conclude that the measurement Λ_T necessarily satisfies

$$|\mathrm{F}_c(\mathrm{tr}\,(\Lambda_T\rho_\delta), \mathrm{tr}\,(\Lambda_T\sigma_\delta)) - \mathrm{F}\,(\rho_\delta, \sigma_\delta)| \leqslant \mathrm{negl}\,(\delta),$$

i.e., it "approximately preserves" the fidelity between ρ_δ and σ_δ.

This suggests that it may be useful to characterize the set of measurements that "approximately preserve" fidelity in the above sense. To this end, we prove the following lemma, which looks roughly similar in some ways to Lemma 2(note, however, that we have only proven one direction of the implications in this lemma, i.e., this result might not be a bidirectional implication).

Lemma 4. Consider any states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ and define ρ_δ, σ_δ as in Eq. (32), in which case $\rho_\delta, \sigma_\delta$ are invertible for all $\delta > 0$, and we can define the operator

$$M_\delta := \rho_\delta^{-1} \# \sigma_\delta. \tag{E1}$$

Suppose that $\Lambda \in \mathcal{M}$ is a rank-1 projective measurement with projectors $\{|e_x\rangle\langle e_x|\}_{x\in\mathcal{X}}$ such that, for all $\delta > 0$,

$$|\mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho_\delta), \mathrm{tr}\,(\Lambda\sigma_\delta)) - \mathrm{F}\,(\rho_\delta, \sigma_\delta)| \leqslant \mathrm{negl}\,(\delta). \tag{E2}$$

Then for any $x \in \mathcal{X}$ such that $|e_x\rangle \notin \ker\rho$ and any $\delta > 0$, there exists $\mu_{x\delta} \in \mathbb{C}$, which together satisfy

$$\lim_{\delta\to 0^+} \sqrt{\rho_\delta}(M_\delta - |\mu_{x\delta}|\mathbb{1})|e_x\rangle = 0. \tag{E3}$$

Proof. To make some steps easier to follow, we shall start by writing the condition (E2) in the form

$$|\mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho_\delta), \mathrm{tr}\,(\Lambda\sigma_\delta)) - \mathrm{F}\,(\rho_\delta, \sigma_\delta)| \leqslant \varepsilon_\delta, \tag{E4}$$

and only set $\varepsilon_\delta = \mathrm{negl}(\delta)$ near the end of the argument. The idea of the proof is to follow essentially the same steps as in the proof in Appendix C 2, except with "approximate equalities" instead of inequalities. We begin by letting U_δ be the operator (induced by polar decomposition) such that

$$\sqrt{\sqrt{\rho_\delta}\,\sigma_\delta\sqrt{\rho_\delta}} = \sqrt{\rho_\delta}\sqrt{\sigma_\delta}U_\delta.$$

For brevity, we introduce the following operators (note that this part of the construction also essentially works for a general measurement with POVM operators $\{E_x\}_{x\in\mathcal{X}}$; one just needs to use $\sqrt{E_x}$ in place of $|e_x\rangle\langle e_x|$):

$$A_{x\delta} := |e_x\rangle\langle e_x|\sqrt{\sigma_\delta}U_\delta, \quad B_{x\delta} := |e_x\rangle\langle e_x|\sqrt{\rho_\delta}.$$

Note that, since we are only considering the $\delta > 0$ regime, ρ_δ, σ_δ are invertible and hence the operators $A_{x\delta}$, $B_{x\delta}$ are always nonzero.

Now, following the same calculations as in the proof in Appendix C 2 gives

$$\begin{aligned}\mathrm{F}\,(\rho_\delta, \sigma_\delta) &= \sum_{x\in\mathcal{X}} \mathrm{tr}\left(\sqrt{\rho_\delta}\,|e_x\rangle\langle e_x|\sqrt{\sigma_\delta}\,U_\delta\right) = \sum_{x\in\mathcal{X}} \mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})\\ &\leqslant \sum_{x\in\mathcal{X}} |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})| \leqslant \sum_{x\in\mathcal{X}} \|A_{x\delta}\|_2\|B_{x\delta}\|_2\\ &= \sum_{x\in\mathcal{X}} \sqrt{\mathrm{tr}\,(|e_x\rangle\langle e_x|\,\rho_\delta)}\sqrt{\mathrm{tr}\,(|e_x\rangle\langle e_x|\,\sigma_\delta)}\\ &= \mathrm{F}_c(\mathrm{tr}\,(\Lambda\rho_\delta), \mathrm{tr}\,(\Lambda\sigma_\delta)),\end{aligned}$$

where the second inequality is the step where Cauchy-Schwarz was applied. Combining this with the condition (E4), we see that both the inequalities in the chain must be "tight up to ε_δ", i.e., we have

$$\left|\sum_{x\in\mathcal{X}} |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})| - \sum_{x\in\mathcal{X}} \mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})\right| \leqslant \varepsilon_\delta, \tag{E5}$$

and

$$\left|\sum_{x\in\mathcal{X}} \|A_{x\delta}\|_2\|B_{x\delta}\|_2 - \sum_{x\in\mathcal{X}} \mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})\right| \leqslant \varepsilon_\delta. \tag{E6}$$

For the bound (E5), if we write $z_x := \mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})$, then we have some values $z_x \in \mathbb{C}$ such that $|z_x| \leqslant \|A_{x\delta}\|_2\|B_{x\delta}\|_2 \leqslant 1$, and their sum $\sum_{x\in\mathcal{X}} z_x$ is real-valued [because it is equal to $\mathrm{F}(\rho_\delta, \sigma_\delta)$] and within ε_δ of the sum of their absolute values. Viewing these complex numbers z_x as vectors in the complex plane, a geometric argument then shows that we must have $||z_x| - z_x| \leqslant \sqrt{2\varepsilon_\delta}$ for all $x \in \mathcal{X}$, i.e.,

$$\forall\, x \in \mathcal{X}, \quad ||\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})| - \mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})| \leqslant \sqrt{2\varepsilon_\delta}. \tag{E7}$$

As for the bound (E6), note that the terms in the summations satisfy $|\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})| \leqslant \|A_{x\delta}\|_2\|B_{x\delta}\|_2$ (this is just Cauchy–Schwarz) for all $x \in \mathcal{X}$, from which we can deduce that

$$\forall\, x \in \mathcal{X}, \quad |\|A_{x\delta}\|_2\|B_{x\delta}\|_2 - |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|| \leqslant \varepsilon_\delta.$$

The above bound suggests that $\|A_{x\delta}\|_2^2\|B_{x\delta}\|_2^2$ and $|\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|^2$ should be close as well (for all $x \in \mathcal{X}$). To formalize this, we note that $\|A_{x\delta}\|_2\|B_{x\delta}\|_2$ and $|\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|$ are both upper bounded by 1, and thus

$$\begin{aligned}&\|A_{x\delta}\|_2^2\|B_{x\delta}\|_2^2 - |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|^2\\ &\quad = (\|A_{x\delta}\|_2\|B_{x\delta}\|_2 + |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|)\\ &\qquad \times (\|A_{x\delta}\|_2\|B_{x\delta}\|_2 - |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|)\\ &\quad \leqslant 2(\|A_{x\delta}\|_2\|B_{x\delta}\|_2 - |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|).\end{aligned}$$

Combined with the preceding bound, this gives

$$\forall\, x \in \mathcal{X}, \ \sqrt{\|A_{x\delta}\|_2^2\|B_{x\delta}\|_2^2 - |\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|^2} \leqslant \sqrt{2\varepsilon_\delta}. \tag{E8}$$

The reason for expressing the bound in the above form is so we can now make use of the following equality that appears in the derivation of Cauchy–Schwarz (which can be verified by expanding the left-hand-side):

$$\left\|A_{x\delta} - \frac{\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})}{\|B_{x\delta}\|_2^2}B_{x\delta}\right\|_2^2 = \|A_{x\delta}\|_2^2 - \frac{|\,\mathrm{tr}(B_{x\delta}^\dagger A_{x\delta})|^2}{\|B_{x\delta}\|_2^2}.$$

Let us now define

$$\mu_{x\delta} := \frac{\operatorname{tr}(B_{x\delta}^\dagger A_{x\delta})}{\|B_{x\delta}\|_2^2}, \tag{E9}$$

so that the left-hand-side of the preceding expression is just $\|A_{x\delta} - \mu_{x\delta} B_{x\delta}\|_2^2$. With this, we can write

$$\begin{aligned} \forall\, x \in \mathcal{X}, \quad \|A_{x\delta} - \mu_{x\delta} B_{x\delta}\|_2 &= \sqrt{\|A_{x\delta}\|_2^2 - \frac{|\operatorname{tr}(B_{x\delta}^\dagger A_{x\delta})|^2}{\|B_{x\delta}\|_2^2}} \\ &\leqslant \frac{\sqrt{2\varepsilon_\delta}}{\|B_{x\delta}\|_2}, \end{aligned}$$

where in the last line we applied the bound (E6). With this, we have for all $x \in \mathcal{X}$:

$$\begin{aligned} &\|A_{x\delta} - |\mu_{x\delta}| B_{x\delta}\|_2 \\ &\quad \leqslant \|A_{x\delta} - \mu_{x\delta} B_{x\delta}\|_2 + |\mu_{x\delta} - |\mu_{x\delta}|\,|\|B_{x\delta}\|_2 \\ &\quad = \|A_{x\delta} - \mu_{x\delta} B_{x\delta}\|_2 + \frac{|\operatorname{tr}(B_{x\delta}^\dagger A_{x\delta}) - |\operatorname{tr}(B_{x\delta}^\dagger A_{x\delta})||}{\|B_{x\delta}\|_2} \\ &\quad \leqslant \frac{\sqrt{2\varepsilon_\delta}}{\|B_{x\delta}\|_2} + \frac{\sqrt{2\varepsilon_\delta}}{\|B_{x\delta}\|_2} = \frac{2\sqrt{2\varepsilon_\delta}}{\|B_{x\delta}\|_2}, \end{aligned}$$

where in the last line we applied the bound (E7).

The above result is essentially the main bound that yields the desired claim. (Note that in the case where ρ, σ are both invertible and we set both δ and ε_δ to zero, the above bound reduces to $A_{x\delta} = |\mu_{x\delta}| B_{x\delta}$, which is basically equivalent to the $|e_x\rangle\langle e_x|\sqrt{\rho} = c_x|e_x\rangle\langle e_x|\sqrt{\sigma}\,U$ condition in the Appendix C 2 proof.) To finish up, we note that by the definition of $B_{x\delta}$ we have $\|B_{x\delta}\|_2 = \langle e_x|\rho_\delta|e_x\rangle$, and hence (because $\lim_{\delta\to 0^+}\rho_\delta$ exists and equals ρ):

$$\lim_{\delta\to 0^+} \|B_{x\delta}\|_2 = \lim_{\delta\to 0^+} \langle e_x|\rho_\delta|e_x\rangle = \langle e_x|\rho|e_x\rangle.$$

This means that for any $x \in \mathcal{X}$ such that $|e_x\rangle \notin \ker\rho$, the value $L_x := \lim_{\delta\to 0^+} \|B_{x\delta}\|_2$ exists and is strictly positive. Hence for such x, we know that for all sufficiently small δ, we would have $\|B_{x\delta}\|_2 \geqslant L_x/2$ and thus also

$$\|A_{x\delta} - |\mu_{x\delta}| B_{x\delta}\|_2 \leqslant \frac{4\sqrt{2\varepsilon_\delta}}{L_x}. \tag{E10}$$

With this, we finally substitute $\varepsilon_\delta = \operatorname{negl}(\delta)$ to conclude that for such x, we have

$$\lim_{\delta\to 0^+} \|A_{x\delta} - |\mu_{x\delta}| B_{x\delta}\|_2 = 0,$$

and thus

$$\lim_{\delta\to 0^+} (A_{x\delta} - |\mu_{x\delta}| B_{x\delta}) = 0. \tag{E11}$$

(It does not matter which operator norm is considered in the above convergence statement, because all norms on a finite-dimensional vector space yield the same topology.) Note that for all $\delta > 0$, by substituting the definitions of $A_{x\delta}$, $B_{x\delta}$, and U_δ (and also using the fact that ρ_δ is invertible), we get

$$\begin{aligned} A_{x\delta} - |\mu_{x\delta}| B_{x\delta} &= |e_x\rangle\langle e_x|(\sqrt{\sigma_\delta}U_\delta - |\mu_{x\delta}|\sqrt{\rho_\delta}) \\ &= |e_x\rangle\langle e_x|(M_\delta\sqrt{\rho_\delta} - |\mu_{x\delta}|\sqrt{\rho_\delta}). \end{aligned}$$

Substituting this into (E11), then left-multiplying by $\langle e_x|$ and taking the adjoint, we get the desired result. ■

Note that if it can be shown that $\lim_{\delta\to 0^+} M_\delta$ and $\lim_{\delta\to 0^+} \mu_{x\delta}$ exist (let us denote their limiting values as $\lim_{\delta\to 0^+} M_\delta =: M_0$ and $\lim_{\delta\to 0^+} \mu_{x\delta} =: \mu_{x0}$), then the condition (E3) reduces to a form of "skewed" eigenvalue condition:

$$\sqrt{\rho} M_0 |e_x\rangle = \sqrt{\rho}|\mu_{x0}||e_x\rangle. \tag{E12}$$

(In fact, it may not be strictly necessary to show that both $\lim_{\delta\to 0^+} M_\delta$ and $\lim_{\delta\to 0^+} \mu_{x\delta}$ exist; from the condition (E3) we know that, e.g., if $\lim_{\delta\to 0^+} \sqrt{\rho_\delta} M_\delta |e_x\rangle$ exists then so does $\lim_{\delta\to 0^+} \sqrt{\rho_\delta}|\mu_{x\delta}||e_x\rangle$ and vice versa, although the factors of $\sqrt{\rho_\delta}$ and $|e_x\rangle$ make this not entirely straightforward to work with). Roughly speaking, the main challenge in trying to get the above line of reasoning to yield a result fully similar to Lemma 2 is that when ρ is noninvertible, we cannot multiply both sides of Eq. (E12) by $(\rho^{-1})^{1/2}$ to remove the $\sqrt{\rho}$ prefactors and get a genuine eigenvalue equation. If instead we try working with one of the intermediate bounds in the proof, such as (E10), and multiply by $(\rho_\delta^{-1})^{1/2}$ (which is well defined for $\delta > 0$), the issue is that the maximum eigenvalue of $(\rho_\delta^{-1})^{1/2}$ diverges as $\delta \to 0^+$, making it difficult to bound the norm of the resulting quantities.

Regarding the question of whether $\lim_{\delta\to 0^+} M_\delta$ exists, we note that in the special case where both ρ and σ are pure qubit states, we can without loss of generality write $\rho = |0\rangle$ and $\sigma = \alpha|0\rangle + \beta|1\rangle$ in some basis, in which case we can compute the following limit for $\alpha \neq 0$:

$$\lim_{\delta\to 0^+} M_\delta = \begin{pmatrix} |\alpha| & \frac{ab^*}{|\alpha|} \\ \frac{ba^*}{|\alpha|} & \frac{|\beta|^2}{|\alpha|} + \sqrt{1 + \frac{|\beta|^2}{|\alpha|^2}} \end{pmatrix}.$$

(The $\alpha = 0$ case corresponds to ρ, σ being orthogonal, in which case it appears that M_δ diverges as $\delta \to 0^+$, but this case is not very relevant in our context since the Fuchs–van de Graaf inequalities are trivially saturated in this case.) More generally, for $\dim\mathcal{H} = 3$, we were able to compute the operator M_δ analytically in *Mathematica* and found that, as long as ρ, σ are nonorthogonal pure states, the limit $\lim_{\delta\to 0^+} M_\delta$ indeed exists. However, we currently do not have a generalization of the argument to higher dimensions.

Finally, we remark that one possible direction for further investigation is that rather than focusing on proving an analog of Lemma 2, we could instead try to use some of the intermediate steps in the proof [such as the bounds (E7) and (E8)] to more directly analyze the set of states saturating the upper Fuchs–van de Graaf inequality. However, it currently does not seem clear whether this gives a useful result.

APPENDIX F: APPLICATIONS IN KEYRATE CALCULATIONS

Here, we briefly outline some potential applications of our results in computing QKD keyrates, focusing on a form of QKD referred to as device-independent (DI) QKD [23]. Basically, it was observed in, e.g., Ref. [24], that one potential approach to compute such keyrates is to lower bound the fidelity between a particular pair of states. If it could be shown that these states saturate the upper Fuchs–van de Graaf inequality, then the fidelity can be written in terms of the trace distance,[3] which has an operational interpretation in terms of guessing probability [4]. Various methods are known for bounding guessing probabilities in DIQKD [25–27], and hence this could serve as another potential approach for DIQKD keyrate computations. In fact, if the upper Fuchs–van de Graaf inequality were saturated in this context, then the guessing-probability bound derived in Ref. [25] for a particular DIQKD protocol would yield an expression exactly matching the (tight) keyrate formula derived in Ref. [23] for that protocol, suggesting some plausibility in this approach.

Another aspect of DIQKD in which such a result could be useful would be protocols based on advantage distillation [28], which refers to the use of two-way communication for the information-reconciliation step [29] of the protocol. Again, the analysis in Ref. [28] is based on the fidelity between a particular pair of states (different from the above), and it was observed that if that pair of states saturates the upper Fuchs–van de Graaf inequality, then significantly better results could be obtained. Hence a more detailed characterization of the set of such states could be of use.

[3] To be more precise, one could always use the lower Fuchs–van de Graaf inequality to bound the fidelity, but in this context the resulting values are highly suboptimal, hence we are interested in whether the states could in fact saturate the upper bound.

[1] M. Wilde, *Quantum Information Theory* (Cambridge University Press, New York, 2013).

[2] R. Cleve and D. P. DiVincenzo, Schumacher's quantum data compression as a quantum computation, Phys. Rev. A **54**, 2636 (1996).

[3] M. Berta, O. Fawzi, and S. Wehner, Quantum to classical randomness extractors, in *Advances in Cryptology – CRYPTO 2012*, edited by R. Safavi-Naini and R. Canetti (Springer, Berlin, Heidelberg, 2012), pp. 776–793.

[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2010).

[5] K. Audenaert, A sharp continuity estimate for the von Neumann entropy, J. Phys. A: Math. Theor. **40**, 8127 (2007).

[6] A. Winter, Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints, Commun. Math. Phys. **347**, 291 (2016).

[7] T. Upadhyaya, T. van Himbeeck, J. Lin, and N. Lütkenhaus, Dimension reduction in quantum key distribution for continuous- and discrete-variable protocols, PRX Quantum **2**, 020325 (2021).

[8] F. Kanitschar, I. George, J. Lin, T. Upadhyaya, and N. Lütkenhaus, Finite-size security for discrete-modulated continuous-variable quantum key distribution protocols, arXiv:2301.08686

[9] P. Sekatski, J.-D. Bancal, X. Valcarce, E. Y.-Z. Tan, R. Renner, and N. Sangouard, Device-independent quantum key distribution from generalized CHSH inequalities, Quantum **5**, 444 (2021).

[10] L. Mirsky, A trace inequality of John von Neumann, Monatsh. Math. **79**, 303 (1975).

[11] C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[12] T. Ando, C.-K. Li, and R. Mathias, Geometric means, Linear Algebra Appl. **385**, 305 (2004).

[13] G. Toussaint, Comments on "The divergence and Bhattacharyya distance measures in signal selection," IEEE Trans. Commun. **20**, 485 (1972).

[14] C. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[15] C. A. Fuchs and C. M. Caves, Mathematical techniques for quantum communication theory, Open Syst. Inf. Dyn. **3**, 345 (1995).

[16] R. Bhatia, *Positive Definite Matrices*, Princeton Series in Applied Mathematics (Princeton University Press, Princeton, 2006).

[17] E. P. Hanson and N. Datta, Entropies, majorization flow, and continuity bounds, in *The Physics and Mathematics of Elliott Lieb The 90th Anniversary* (EMS Press, 2022), Vol. 1, pp. 473–514.

[18] M. G. Jabbour and N. Datta, A tight uniform continuity bound for the Arimoto-Rényi conditional entropy and its extension to classical-quantum states, IEEE Trans. Inf. Th. **68**, 2169 (2022).

[19] A. Marwah and F. Dupuis, Uniform continuity bound for sandwiched Rényi conditional entropy, J. Math. Phys. **63**, 052201 (2022).

[20] A. Bluhm, Á. Capel, P. Gondolf, and A. Pérez-Hernández, Continuity of quantum entropic quantities via almost convexity, arXiv:2208.00922v1.

[21] M. M. Wilde, Optimal uniform continuity bound for conditional entropy of classical–quantum states, Quantum Inf. Process. **19**, 61(2020).

[22] W. Roga, M. Fannes, and K. Życzkowski, Universal Bounds for the Holevo Quantity, Coherent Information, and the Jensen-Shannon Divergence, Phys. Rev. Lett. **105**, 040505 (2010).

[23] S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, Device-independent quantum key distribution secure against collective attacks, New J. Phys. **11**, 045021 (2009).

[24] E. Woodhead, Tight asymptotic key rate for the Bennett-Brassard 1984 protocol with local randomization and device imprecisions, Phys. Rev. A **90**, 022306 (2014).

[25] L. Masanes, S. Pironio, and A. Acín, Secure device-independent quantum key distribution with causally independent measurement devices, Nat. Commun. **2**, 238(2011).

[26] S. Pironio and S. Massar, Security of practical private randomness generation, Phys. Rev. A **87**, 012336 (2013).

[27] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio, Device-independent randomness generation from several Bell estimators, New J. Phys. **20**, 023049 (2018).

[28] Ernest Y.-Z. Tan, C. C.-W. Lim, and R. Renner, Advantage Distillation for Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **124**, 020502 (2020).

[29] R. Renner, *Security of Quantum Key Distribution*, Ph.D. thesis (ETH Zurich, 2005).