Unbounded and lossless compression of multiparameter quantum information

Joe H. Jenne[1] and David R. M. Arvidsson-Shukur[2,3]
[1]*Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[2]*Hitachi Cambridge Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*
[3]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

(Received 12 September 2022; accepted 16 September 2022; published 4 October 2022)

Several tasks in quantum-information processing involve quantum learning. For example, quantum sensing, quantum machine learning, and quantum-computer calibration involve learning and estimating unknown parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$ from measurements of many copies of a quantum state $\hat{\rho}_{\boldsymbol{\theta}}$. This type of metrological information is described by the quantum Fisher information matrix, which bounds the average amount of information learned about $\boldsymbol{\theta}$ per measurement of $\hat{\rho}_{\boldsymbol{\theta}}$. In several scenarios, it is advantageous to compress the multiparameter information encoded in $\hat{\rho}_{\boldsymbol{\theta}}^{\otimes n}$ into $\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps}\,\otimes m}$, where $m \ll n$. Here, we present a "go-go" theorem proving that m/n can be made *arbitrarily* small, and that the information compression can happen *without loss* of information. We also demonstrate how to construct filters that perform this unbounded and lossless information compression. These filters can, for example, reduce arbitrarily the quantum-state intensity on experimental detectors, while retaining all initial information. Finally, we prove that the ability to compress quantum Fisher information is a nonclassical advantage that stems from the *negativity* of a particular quasiprobability distribution, a quantum extension of a probability distribution.

DOI: 10.1103/PhysRevA.106.042404

I. INTRODUCTION

The diverse fields of metrology and machine learning concern estimating, or learning, multiple unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_M)$ from experiments [1–3]. A common measure of an experiment's usefulness in learning is the Fisher information matrix $I(\boldsymbol{\theta})$ [4]. $I(\boldsymbol{\theta})$ quantifies the average information learned about $\boldsymbol{\theta}$ per experimental trial. The covariance matrix of a locally unbiased estimator $\boldsymbol{\theta}^{\text{e}}$ is lower bounded by $I(\boldsymbol{\theta})$ via the Cramér-Rao inequality: $\Sigma(\boldsymbol{\theta}^{\text{e}}) \geqslant [NI(\boldsymbol{\theta})]^{-1}$, where N is the number of independent experimental trials [5,6]. Naïvely, the theoretical task is to adjust the experimental input state and final measurement to optimize the Fisher information matrix and to minimize the estimator's risk with respect to some risk function [7–12]. However, theoretical strategies are not necessarily suitable for real technologies—especially not for quantum technologies, which generate data from measurements of quantum states $\hat{\rho}_{\boldsymbol{\theta}}$.

While a theorist aims to optimize the Fisher information, an experimentalist must beware experimental limitations and *costs* [13,14]. Recent works, theoretical and practical, have focused on mitigating experimental problems associated with the measurement and postprocessing of output quantum states. *Weak-value* amplification [15–17] and postselected metrology [18–21] allow the rate of output states per unit time to be lowered while a significant fraction of the information about a single parameter θ_1 is retained. This enables detectors to operate at lower intensities and can, if the postselection is experimentally *cheap*, reduce temporal overheads associated with measurements and postprocessing. The protocols cannot increase the information content, but can reduce the experimental costs of accessing it. A shortcoming of previous information-compression protocols is that they only work when information has been encoded by unitaries of specific forms, and require perfect knowledge of all but one of the experimental parameters—unrealistic settings [22].

Given the important role of multiparameter learning in quantum metrology and quantum machine learning, a generalization of these cost-reducing results is warranted for both practical and foundational reasons. A generalization will help facilitate postselected metrology in diverse experiments, where several parameters are unknown, as well as in quantum machine learning, where the overhead associated with the postprocessing of output data can be monumental. From a foundational perspective, a generalization could provide useful knowledge about postselection as a tool to amplify quantum resources. A previously unanswered fundamental question is, How much metrological information can be encoded in a quantum state?

In this paper, we answer this question. First, we review theoretical results, establishing that scalar risk functions based on the *quantum* Fisher information matrix are suitable objects to minimize, when optimizing quantum learning and quantifying quantum metrological information. Second, we derive a formula for the distilled (postselected) quantum Fisher information matrix. Third, we design a quantum-learning methodology to distill the useful information from an arbitrarily large number of states $\hat{\rho}_{\boldsymbol{\theta}}$ into an arbitrarily small number of states $\hat{\rho}_{\boldsymbol{\theta}}^{\text{ps}}$ (Theorem 1). Our methodology is general: It applies to any reasonable encoding unitary. It is a well-known fact that filtering a number of particles can distill the average information content of a subset of the particles, but not

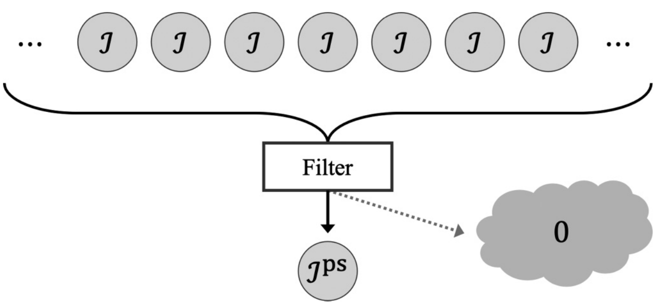

FIG. 1. Compression of metrological information. Our postselected metrology protocol compresses the multiparameter information encoded in $\hat{\rho}_{\boldsymbol{\theta}}{}^{\otimes n}$ into $\hat{\rho}^{\mathrm{ps}}_{\boldsymbol{\theta}}{}^{\otimes m}$. The filter imposes postselection. There is no bound on how small m/n can be. The information lost to discarded particles or to the environment can be vanishingly small.

increase the total average information of all particles. Nevertheless, our protocol is lossless: No information is wasted in the distillation (postselection) procedure, and the total information of all the initial particles is compressed into an arbitrarily small number of final particles. The main results of our work are summarized in Fig. 1. Fourth, we discuss how our results can be applied to improve quantum learning in the presence of imperfect detectors or postprocessing costs. Finally, we show that our results leverage *negativity* [26], a narrower nonclassicality concept than noncommutation.

II. PRELIMINARIES

Consider a learning experiment with finite and discrete outcomes k with corresponding probabilities $p(k|\boldsymbol{\theta})$. The Fisher information matrix is defined as

$$I_{i,j}(\boldsymbol{\theta}) = \sum_k p(k|\boldsymbol{\theta})\{\partial_i \ln[p(k|\boldsymbol{\theta})]\}\{\partial_j \ln[p(k|\boldsymbol{\theta})]\}, \quad (1)$$

where $\partial_i \equiv \frac{\partial}{\partial\theta_i}$. The Fisher information matrix lower-bounds the covariance matrix $\Sigma(\boldsymbol{\theta}^{\mathrm{e}})$ via the Cramér-Rao inequality: $\Sigma(\boldsymbol{\theta}^{\mathrm{e}}) \geqslant [NI(\boldsymbol{\theta})]^{-1}$. Choosing a positive, real, $M\times M$ weight matrix W introduces a scalar Cramér-Rao bound:

$$s(\Sigma(\boldsymbol{\theta}^{\mathrm{e}}), W) \equiv \mathrm{Tr}[W\Sigma(\boldsymbol{\theta}^{\mathrm{e}})] \geqslant \frac{1}{N}\mathrm{Tr}[WI^{-1}(\boldsymbol{\theta})]. \quad (2)$$

If, e.g., $W = \mathbb{1}$ and $\boldsymbol{\theta}^{\mathrm{e}}$ is an unbiased estimator, the scalar risk function $s(\Sigma(\boldsymbol{\theta}^{\mathrm{e}}), W)$ equals the sum of the individual mean-square errors of the parameters in $\boldsymbol{\theta}^{\mathrm{e}}$. See Ref. [4] for a review. For unbiased, or "reasonable," estimators $\boldsymbol{\theta}^{\mathrm{e}}$ and $N \to \infty$, inequality (2) is saturated [27]. In what follows, we shall assume these conditions, such that $s(\Sigma(\boldsymbol{\theta}^{\mathrm{e}}), W) \equiv s(I(\boldsymbol{\theta}), W) = \mathrm{Tr}[WI^{-1}(\boldsymbol{\theta})]/N$.

From a learnability perspective, it is often useful to consider the most informative experiment that extracts (Fisher) information from quantum states $\hat{\rho}_{\boldsymbol{\theta}}$:

$$s^{(\mathrm{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W) \equiv s\Big(\max_{\mathcal{M}} I(\boldsymbol{\theta}), W\Big) = \frac{1}{N}\min_{\mathcal{M}} \mathrm{Tr}[WI^{-1}(\boldsymbol{\theta})]. \quad (3)$$

Here, $\mathcal{M}$ is the set of all possible measurements.

In the Supplemental Material, we review previous results, showing that $s^{(\mathrm{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W)$ is bounded by

$$\frac{1}{N}\mathrm{Tr}[W\mathcal{I}^{-1}(\boldsymbol{\theta})] \leqslant s^{(\mathrm{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W) \leqslant 2\frac{1}{N}\mathrm{Tr}[W\mathcal{I}^{-1}(\boldsymbol{\theta})]. \quad (4)$$

$\mathcal{I}(\boldsymbol{\theta})$ is the symmetric-logarithmic-derivative quantum Fisher information. In this theoretical proof-of-principle study, we focus on pure states. An investigation of distilled quantum learning in the presence of noise is left for an upcoming paper. For pure states,

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) = 4\,\mathrm{Re}[\langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle - \langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle]. \quad (5)$$

Within a factor of 2, $\mathcal{I}(\boldsymbol{\theta})$ sets $s^{(\mathrm{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W)$. This constitutes our main motivation for focusing on $\mathcal{I}(\boldsymbol{\theta})$ as a measure of quantum learnability. Furthermore, empirical machine-learning motivation can be found in Refs. [28–30]. Below, we show that the quantum Fisher information encoded in an arbitrarily large number of identical states can be compressed into an arbitrarily small number of identical states, without any information loss.

III. POSTSELECTED QUANTUM FISHER INFORMATION MATRIX

Here, we consider an experiment where an initial state, ρ_0, is evolved by an M-parameter unitary, $\hat{U}(\boldsymbol{\theta})$: $\hat{\rho}_0 \to \hat{\rho}_{\boldsymbol{\theta}} \equiv \hat{U}(\boldsymbol{\theta})\hat{\rho}_0\hat{U}^{\dagger}(\boldsymbol{\theta})$. Then it is subject to a postselective measurement $\{\hat{F}_1 = \hat{F}, \hat{F}_2 = \hat{1} - \hat{F}\}$. $\hat{F}_i$ need not be projective. The experiment is depicted in Fig. 2. Throughout this paper, we assume a discrete Hilbert space of dimension d.

The output states in Fig. 2 are given by $|\psi^{\mathrm{ps}}_{\boldsymbol{\theta}}\rangle \equiv \hat{K}|\psi_{\boldsymbol{\theta}}\rangle / \sqrt{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}$, where $p^{\mathrm{ps}}_{\boldsymbol{\theta}} = \mathrm{Tr}[\hat{F}\hat{\rho}_{\boldsymbol{\theta}}]$ is the probability of successful postselection and $\hat{K}$ is the Kraus operator that sets the postselection filter: $\hat{F} = \hat{K}^{\dagger}\hat{K}$.

We now derive a formula for the postselected quantum Fisher information matrix. By evaluating Eq. (5) for $|\psi_{\boldsymbol{\theta}}\rangle \to |\psi^{\mathrm{ps}}_{\boldsymbol{\theta}}\rangle$, we extend and generalize the single-parameter results of Ref. [19]. The first inner product of Eq. (5) is then given by

$$\begin{aligned}
&\left(\partial_i \frac{\langle\psi_{\boldsymbol{\theta}}|\hat{K}^{\dagger}}{\sqrt{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}}\right)\cdot\left(\partial_j \frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}}\right)\\
&= \left(\frac{\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{K}^{\dagger}}{\sqrt{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}} - \frac{1}{2}\frac{\langle\psi_{\boldsymbol{\theta}}|\hat{K}^{\dagger}}{(p^{\mathrm{ps}}_{\boldsymbol{\theta}})^{\frac{3}{2}}}\partial_i p^{\mathrm{ps}}_{\boldsymbol{\theta}}\right)\\
&\quad\cdot\left(\frac{\hat{K}|\partial_j\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}} - \frac{1}{2}\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{(p^{\mathrm{ps}}_{\boldsymbol{\theta}})^{\frac{3}{2}}}\partial_j p^{\mathrm{ps}}_{\boldsymbol{\theta}}\right)\\
&= \frac{\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\partial_j\psi_{\boldsymbol{\theta}}\rangle}{p^{\mathrm{ps}}_{\boldsymbol{\theta}}} - \frac{1}{2}\frac{\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\partial_j\psi_{\boldsymbol{\theta}}\rangle}{(p^{\mathrm{ps}}_{\boldsymbol{\theta}})^2}\partial_i p^{\mathrm{ps}}_{\boldsymbol{\theta}}\\
&\quad - \frac{1}{2}\frac{\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle}{(p^{\mathrm{ps}}_{\boldsymbol{\theta}})^2}\partial_j p^{\mathrm{ps}}_{\boldsymbol{\theta}} + \frac{1}{4}\frac{(\partial_i p^{\mathrm{ps}}_{\boldsymbol{\theta}})(\partial_j p^{\mathrm{ps}}_{\boldsymbol{\theta}})}{(p^{\mathrm{ps}}_{\boldsymbol{\theta}})^2},
\end{aligned} \quad (6)$$

where the last equality follows from $p^{\mathrm{ps}}_{\boldsymbol{\theta}} = \langle\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle$. The second inner product of Eq. (5) is

$$\left(\partial_i \frac{\langle\psi_{\boldsymbol{\theta}}|\hat{K}^{\dagger}}{\sqrt{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}}\right)\cdot\left(\frac{\hat{K}|\psi_{\boldsymbol{\theta}}\rangle}{\sqrt{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}}\right) = \frac{\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle}{p^{\mathrm{ps}}_{\boldsymbol{\theta}}} - \frac{1}{2}\frac{(\partial_i p^{\mathrm{ps}}_{\boldsymbol{\theta}})}{p^{\mathrm{ps}}_{\boldsymbol{\theta}}}. \quad (7)$$

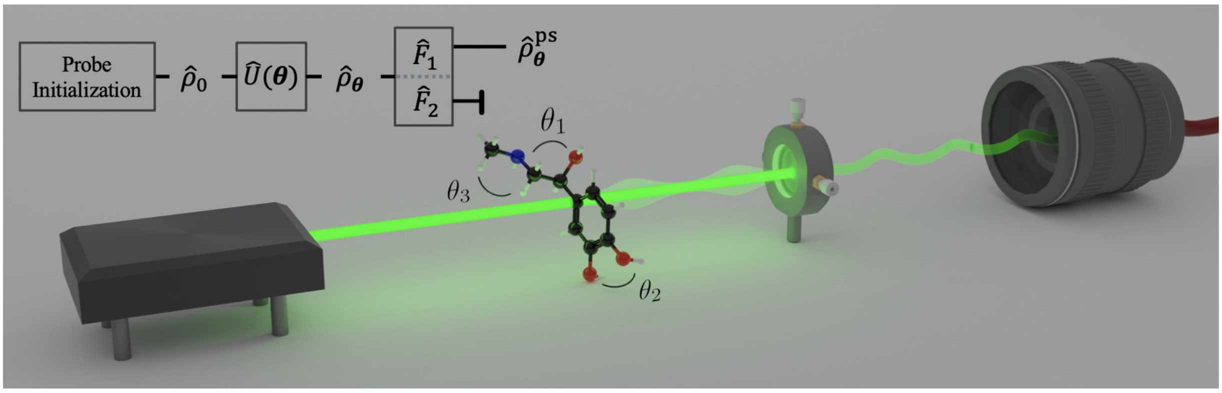

FIG. 2. Preparation and distillation of quantum states. Potential optical realization (main), and schematic figure (inset). First, M unknown parameters $\boldsymbol{\theta}$ are encoded in the initial state $\hat{\rho}_0$ by the unitary $\hat{U}(\boldsymbol{\theta})$: $\hat{\rho}_0 \to \hat{\rho}_{\boldsymbol{\theta}}$. Second, the encoded state $\hat{\rho}_{\boldsymbol{\theta}}$ is past through a postselective measurement $\{\hat{F}_1 = \hat{F}, \hat{F}_2 = \hat{1} - \hat{F}\}$. The postselection filters out the quantum states unless outcome $\hat{F} = \hat{K}^\dagger \hat{K}$ happens. Third, the experiment outputs the distilled states $\hat{\rho}^{\text{ps}}_{\boldsymbol{\theta}} = \hat{K}\hat{\rho}_{\boldsymbol{\theta}}\hat{K}^\dagger / p^{\text{ps}}_{\boldsymbol{\theta}}$ with success probability $p^{\text{ps}}_{\boldsymbol{\theta}} = \text{Tr}(\hat{F}\hat{\rho}_{\boldsymbol{\theta}})$.

The third inner product, $\langle \psi_{\boldsymbol{\theta}} | \partial_j \psi_{\boldsymbol{\theta}} \rangle$, is evaluated similarly. Combining these expressions, we find that the quantum Fisher information matrix of the distilled fraction of the output states (the states that have passed the postselection filter) is given by

$$\mathcal{I}_{i,j}\big(\boldsymbol{\theta}\big|\psi^{\text{ps}}_{\boldsymbol{\theta}}\big) = 4\,\text{Re}\Bigg[\frac{1}{p^{\text{ps}}_{\boldsymbol{\theta}}} \langle \partial_i \psi_{\boldsymbol{\theta}} | \hat{F} | \partial_j \psi_{\boldsymbol{\theta}} \rangle - \frac{1}{\big(p^{\text{ps}}_{\boldsymbol{\theta}}\big)^2} \langle \partial_i \psi_{\boldsymbol{\theta}} | \hat{F} | \psi_{\boldsymbol{\theta}} \rangle \langle \psi_{\boldsymbol{\theta}} | \hat{F} | \partial_j \psi_{\boldsymbol{\theta}} \rangle \Bigg]. \quad (8)$$

IV. DISTILLING QUANTUM LEARNABILITY

It is possible to use filtering (postselection) to distill quantum Fisher information such that we get anomalous matrix entries: $|\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}})| > \max_{\hat{\rho}_0} |\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})|$. In other words, postselection can outperform optimal state preparation in condensing the information per output state. (Below, we shall connect such anomalous information values to *negativity*, a narrower nonclassicality concept than noncommutation.) One would expect that anomalously large values of the postselected quantum Fisher information matrix can lead to better learning-per-experimental-cost rates (by reducing the number of measurements needed to obtain a certain amount of information). However, $\mathcal{I}^{-1}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}})$ bounds $s^{(\text{MI})}(\psi^{\text{ps}}_{\boldsymbol{\theta}}, W)$ via matrix inequalities [inequalities (4)], and it is generally hard to know which entry, $\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}})$, it would be beneficial to amplify. Furthermore, setting a postselection operator $\hat{F}$ to optimize one entry in $\mathcal{I}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}})$ could have a detrimental effect on another entry. Next, we show that it is possible to choose $\hat{F}$ such that the entire matrix $\mathcal{I}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}})$ is optimized: $\det \mathcal{I}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}}) \to \infty$, and $s^{(\text{MI})}(\psi^{\text{ps}}_{\boldsymbol{\theta}}, W) \to 0$. The price to pay for $\mathcal{I}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}})$ is smaller success chances $p^{\text{ps}}_{\boldsymbol{\theta}}$. That is, the metrological multiparameter information can be compressed into fewer quantum states, but the compression (filtering) cannot increase the overall information. First, we provide a guiding example of two-parameter estimation of a postselected qubit. Then, we present the general theory.

A. Example

Consider a qubit in an initial state $|\psi_0\rangle = |0\rangle$. The quantum circuit of interest is parametrized by two parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and represented by the unitary $\hat{U}(\boldsymbol{\theta}) = e^{i(\hat{\sigma}_x + \hat{\sigma}_z)\theta_2/\sqrt{2}} e^{i\hat{\sigma}_x\theta_1}$, where $\hat{\sigma}_k$ is the kth Pauli operator. The quantum Fisher information matrix of the output state $|\psi_{\boldsymbol{\theta}}\rangle = \hat{U}(\boldsymbol{\theta})\,|0\rangle$ is

$$\mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) = \begin{pmatrix} 4 & 2\sqrt{2} \\ 2\sqrt{2} & 3 - \cos(4\theta_1) \end{pmatrix}. \quad (9)$$

We assume that our initial guess of $\boldsymbol{\theta}$, $\hat{\rho}_{\boldsymbol{\theta}^0}$, is off by $1/10$ for both θ_1 and θ_2: $\boldsymbol{\theta}^0 = (\theta_1 + \frac{1}{10}, \theta_2 + \frac{1}{10})$. We set the Kraus operator to $\hat{K} = (\frac{1}{\sqrt{5}} - 1)\hat{\rho}_{\boldsymbol{\theta}^0} + \hat{1}$. The probability of a successful postselection is given by $p^{\text{ps}}_{\boldsymbol{\theta}} = \text{Tr}[\hat{K}^\dagger \hat{K} \hat{\rho}_{\boldsymbol{\theta}}] \approx 1/5$. Moreover, the postselected (distilled) quantum Fisher information matrix is given by

$$\mathcal{I}\big(\boldsymbol{\theta}\big|\psi^{\text{ps}}_{\boldsymbol{\theta}}\big) \approx 5 \begin{pmatrix} 4 & 2\sqrt{2} \\ 2\sqrt{2} & 3 - \cos(4\theta_1) \end{pmatrix}. \quad (10)$$

All entries of $\mathcal{I}(\boldsymbol{\theta}|\psi^{\text{ps}}_{\boldsymbol{\theta}})$ are anomalous and exceed their classical maximum of 4 (see below). By reducing (via postselective filtering) the number of quantum states that will reach the final detector by a factor of 5, we have also achieved a fivefold increase of the information content of the remaining states.

B. General theory

Here, we outline how to achieve a diverging quantum Fisher information matrix in the general scenario. Our results assume that we possess an *initial* estimate of $\boldsymbol{\theta}$, $\boldsymbol{\theta}^0$, that is sufficiently close to the true value: $\boldsymbol{\theta}^0 \approx \boldsymbol{\theta}$ [31]. In the limit of many trials ($N \to \infty$), we can always “sacrifice” a vanishingly small fraction of the trials to achieve such an initial estimate. $\boldsymbol{\theta}^0$ can also be improved iteratively, suitably using a Kalman filter [36].

Theorem 1: Arbitrary distillation ofquantum learnability. For a sufficiently accurate initial estimate, the theoretically attainable average distilled multiparameter information per output state about the unknown parameter vector $\boldsymbol{\theta}$ has no upper limit: It is possible to distill quantum states such that $\mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) \to \infty$ and $s^{(\mathrm{MI})}(\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}, W) \to 0$ in a lossless fashion.

Proof of Theorem 1. Our proof is constructive. We present a specific protocol that achieves the objective; other protocols might exist. We consider the setup depicted in Fig. 2, with postselected quantum Fisher information given by Eq. (8).

First, we express our initial quantum-state estimate, $\hat{\rho}_{\boldsymbol{\theta}^0}$, in terms of the true state $\hat{\rho}_{\boldsymbol{\theta}}$. We start by expanding $\hat{U}(\boldsymbol{\theta}^0)$ around $\boldsymbol{\theta}$. To simplify the notation, we define $\boldsymbol{\delta} \equiv (\delta_1, \dots, \delta_M)$ such that $\boldsymbol{\delta} = \boldsymbol{\theta} - \boldsymbol{\theta}^0$. The following calculations assume that $|\boldsymbol{\delta}|^2 \ll 1$.

$$\hat{U}(\boldsymbol{\theta}^0) = \hat{U}(\boldsymbol{\theta}) + [\nabla_{\boldsymbol{\theta}}\hat{U}(\boldsymbol{\theta})]^\top(\boldsymbol{\theta}^0 - \boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2) \tag{11}$$

$$= \hat{U}(\boldsymbol{\theta}) - i\hat{U}(\boldsymbol{\theta})\hat{d} + O(|\boldsymbol{\delta}|^2). \tag{12}$$

In the last step we have defined the $O(|\boldsymbol{\delta}|)$ Hermitian operator $\hat{d} \equiv \hat{U}^\dagger(\boldsymbol{\theta})[-i\nabla_{\boldsymbol{\theta}}\hat{U}(\boldsymbol{\theta})]^\top\boldsymbol{\delta}$. We can now evaluate, to $O(|\boldsymbol{\delta}|^2)$, $\hat{\rho}_{\boldsymbol{\theta}^0}$:

$$\hat{\rho}_{\boldsymbol{\theta}^0} = \hat{U}(\boldsymbol{\theta}^0)\hat{\rho}_0\hat{U}^\dagger(\boldsymbol{\theta}^0) \tag{13}$$

$$= \hat{U}(\boldsymbol{\theta})(\hat{1} - i\hat{d})\hat{\rho}_0(\hat{1} + i\hat{d})\hat{U}^\dagger(\boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2) \tag{14}$$

$$= \hat{U}(\boldsymbol{\theta})(\hat{\rho}_0 + i[\hat{\rho}_0, \hat{d}])\hat{U}^\dagger(\boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2) \tag{15}$$

$$= \hat{\rho}_{\boldsymbol{\theta}} + i[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}] + O(|\boldsymbol{\delta}|^2). \tag{16}$$

Again, we have defined a $O(|\boldsymbol{\delta}|)$ Hermitian operator $\hat{D} \equiv \hat{U}(\boldsymbol{\theta})\hat{d}\hat{U}^\dagger(\boldsymbol{\theta})$. Equation (16) is general (the Supplemental Material provides guiding examples). The only assumption is that $\hat{U}(\boldsymbol{\theta}^0)$ can be locally approximated by its first-order Taylor expansion.

Second, we set the Kraus operator $\hat{K}$ with respect to the initial estimate of the quantum state before postselection [37]:

$$\hat{K} = (t - 1)\hat{\rho}_{\boldsymbol{\theta}^0} + \hat{1}, \tag{17}$$

where $0 \leqslant t \leqslant 1$. Physically, this choice of $\hat{K}$ generates a postselection (distillation) procedure that transmits the expected state $\hat{\rho}_{\boldsymbol{\theta}^0}$ with probability t^2 and transmits fully any state orthogonal to $\hat{\rho}_{\boldsymbol{\theta}^0}$. In interferometric language, our filter allows a detector to operate in a dark fringe, but to access all information.

We proceed by evaluating the postselected Fisher information matrix $\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})$ [Eq. (8)] for the above-outlined choice of Kraus operator, $\hat{K} = (t-1)\hat{\rho}_{\boldsymbol{\theta}^0} + \hat{1}$. This Kraus operator gives the postselection operator $\hat{F} = (t^2 - 1)\hat{\rho}_{\boldsymbol{\theta}^0} + \hat{1}$. The following calculations assume that $M|\boldsymbol{\delta}|^2 \ll t^2$. We define the $O(|\boldsymbol{\delta}|)$ Hermitian operator $\hat{C} \equiv i[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}]$ to simplify notation. We begin by evaluating the individual terms of Eq. (8). Then, we combine these terms.

First, we calculate the postselection probability $p_{\boldsymbol{\theta}}^{\mathrm{ps}}$ in Eq. (8):

$$p_{\boldsymbol{\theta}}^{\mathrm{ps}} = \mathrm{Tr}[\hat{F}\hat{\rho}_{\boldsymbol{\theta}}] \tag{18}$$

$$= \mathrm{Tr}[(\hat{\rho}_{\boldsymbol{\theta}^0}(t^2 - 1) + \hat{1})\hat{\rho}_{\boldsymbol{\theta}}] \tag{19}$$

$$= (t^2 - 1)\mathrm{Tr}[(\hat{\rho}_{\boldsymbol{\theta}} + \hat{C})\hat{\rho}_{\boldsymbol{\theta}}] + 1 + O(|\boldsymbol{\delta}|^2) \tag{20}$$

$$= t^2 + O(|\boldsymbol{\delta}|^2). \tag{21}$$

Here, we have used that $\mathrm{Tr}[\hat{\rho}_{\boldsymbol{\theta}}\hat{C}] = 0$.

Second, we calculate the first inner product in Eq. (8):

$$\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\partial_j\psi_{\boldsymbol{\theta}}\rangle = \langle\partial_i\psi_{\boldsymbol{\theta}}|(\hat{\rho}_{\boldsymbol{\theta}^0}(t^2 - 1) + \hat{1})|\partial_j\psi_{\boldsymbol{\theta}}\rangle \tag{22}$$

$$= (t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|(\hat{\rho}_{\boldsymbol{\theta}} + \hat{C})|\partial_j\psi_{\boldsymbol{\theta}}\rangle + \langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + O(|\boldsymbol{\delta}|^2) \tag{23}$$

$$= (t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + (t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{C}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + \langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + O(|\boldsymbol{\delta}|^2). \tag{24}$$

Third, we calculate the second inner product in Eq. (8):

$$\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle = \langle\partial_i\psi_{\boldsymbol{\theta}}|(\hat{\rho}_{\boldsymbol{\theta}^0}(t^2 - 1) + \hat{1})|\psi_{\boldsymbol{\theta}}\rangle \tag{25}$$

$$= (t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|(\hat{\rho}_{\boldsymbol{\theta}} + \hat{C})|\psi_{\boldsymbol{\theta}}\rangle + \langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle + O(|\boldsymbol{\delta}|^2) \tag{26}$$

$$= t^2\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle + (t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{C}|\psi_{\boldsymbol{\theta}}\rangle + O(|\boldsymbol{\delta}|^2). \tag{27}$$

Fourth, in a similar manner we calculate the third inner product in Eq. (8):

$$\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\partial_j\psi_{\boldsymbol{\theta}}\rangle = t^2\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + (t^2 - 1)\langle\psi_{\boldsymbol{\theta}}|\hat{C}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + O(|\boldsymbol{\delta}|^2). \tag{28}$$

Fifth, we calculate the product of the second and third inner products in Eq. (8):

$$\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{F}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\hat{F}|\partial_j\psi_{\boldsymbol{\theta}}\rangle = t^4\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + t^2(t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\hat{C}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + t^2(t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{C}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + O(|\boldsymbol{\delta}|^2) \tag{29}$$

$$= t^4\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + t^2(t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\,\hat{C}\,|\partial_j\psi_{\boldsymbol{\theta}}\rangle + O(|\boldsymbol{\delta}|^2). \tag{30}$$

Finally, we combine the calculated expressions:

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) = 4\,\mathrm{Re}\bigg\{\frac{1}{t^2}[(t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + (t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\hat{C}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + \langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle] - \frac{1}{t^4}[t^4\langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle + t^2(t^2 - 1)\langle\partial_i\psi_{\boldsymbol{\theta}}|\,\hat{C}\,|\partial_j\psi_{\boldsymbol{\theta}}\rangle]\bigg\} + O(|\boldsymbol{\delta}|^2) \tag{31}$$

$$= \frac{1}{t^2}4\,\mathrm{Re}\{\langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle - \langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle\} + O(|\boldsymbol{\delta}|^2) \tag{32}$$

$$= \frac{1}{t^2}\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) + O(|\boldsymbol{\delta}|^2). \tag{33}$$

To summarize, substituting $\hat{K}$ and $\hat{\rho}_{\boldsymbol{\theta}^0}$ into $\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})$ [Eq. (8)] yields

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) = \frac{1}{t^2}\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) + O(|\boldsymbol{\delta}|^2). \tag{34}$$

We now interpret Eq. (34). $\hat{K}$ is independent of i, j, such that our distillation technique amplifies all nonzero entries of $\mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})$ simultaneously: $\mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) = \mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})/t^2 + O(|\boldsymbol{\delta}|^2)$.

Combining this result with inequalities (4), $s^{(\mathrm{MI})}(\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}, W) \to 0$ when $|\boldsymbol{\delta}|^2 \ll t^2 \to 0$ [38]. Consequently, we have found a protocol that distills the information of all parameters in $\boldsymbol{\theta}$ in a way such that the scalar cost function tends to zero.

It is important to note that while information can be compressed from many states $\psi_{\boldsymbol{\theta}}$ to fewer states $\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}$, the procedure cannot increase the average information of the total system. That is, the larger information content of the states that survive the filtering (postselection) comes at the cost of an information decrease of the states that fail the filtering. Phrased in algebraic terms, $p_{\boldsymbol{\theta}}^{\mathrm{ps}} \times \mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) \leqslant \mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})$. As outlined above, the probability of successful postselection in our protocol is given by $p_{\boldsymbol{\theta}}^{\mathrm{ps}} = t^{-2} + O(|\boldsymbol{\delta}|^2)$. Thus we reach the remarkable conclusion that the distillation of information in our proposed protocol is *lossless*: $p_{\boldsymbol{\theta}}^{\mathrm{ps}} \times \mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) = \mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) + O(|\boldsymbol{\delta}|^2)$. This concludes our constructive proof. ■

V. APPLICATIONS

By distilling the multiparameter Fisher information, the intensity of output states is reduced. This can lead to learnability improvements by allowing metrologists and machine learners to use input-state intensities that would normally have caused the output detectors to saturate. The information content available in the compressed, low-intensity output is identical to what the initial, high-intensity output would have been.

As an application example, consider encoding an image in quantum states $\hat{\rho}_{\boldsymbol{\theta}}$. $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$ is a vector of the image's pixels' intensities. Perhaps our task is to learn about imperfections in the image-encoding procedure of a certain target image $\boldsymbol{\theta}^{\star}$. Then $\hat{\rho}_{\boldsymbol{\theta}^{\star}}$ is a good initial guess to learn the imperfectly encoded, true image $\boldsymbol{\theta} \approx \boldsymbol{\theta}^{\star}$. Alternatively, perhaps we want to learn an image that deviates slightly from a blank image. Then $\boldsymbol{\theta}^0 = \mathbf{0}$, and $\hat{\rho}_{\boldsymbol{\theta}^0} = \hat{\rho}_0$ is a good initial guess. Our distillation protocol allows us to improve the benchmarking of the image-state preparation as well as to avoid detector saturation, and to increase sensitivity, without losing information, when measuring $\hat{\rho}_{\boldsymbol{\theta}}$ to learn the image. More specifically, imagine that we have access to a source supplying coherent photonic states $\hat{\rho}_{\boldsymbol{\theta}}$. The current state-of-the-art experiment produces such states at a rate of 1 GHz [39]. To confirm the findings of the experiment conducted in Ref. [39], the beam of photons had to be attenuated by a factor of 9.9 to avoid detector saturation [39]. In a setup where this photonic beam and these detectors had been used for quantum-learning purposes, our distillation protocol could have been used to attenuate the beam. Compared with a scenario where the supply of photons had been reduced instead, our protocol could lead to an 890% increased rate of Fisher information per sampling time.

A particle-number detector will suffer from a *dead time*, the time needed to reset the detector after triggering it. In the jargon of experimental costs, the dead time associates a temporal cost with the measurement [20]. Also, measurements call for postprocessing, which costs further time and computation. Under the right conditions, our distillation protocol enables an experimentalist to incur the final measurement's cost only when the probe state carries a great deal of information. The "right conditions" are when the postselection is experimentally *cheaper* than the final measurement. For example, if $\hat{\rho}_{\boldsymbol{\theta}}$ can be supplied ten times faster than the dead time of the detector that measures $\hat{\rho}_{\boldsymbol{\theta}}$, then optimal filtering can increase the information-per-time rate tenfold. In Ref. [20], postselected single-parameter metrology improved the sensitivity of a polarization measurement by a factor of >200. Our results diversifies such methods to the multiparameter regime.

Many quantum schemes can be sped up by using several quantum processors in parallel [32,33]. By using our protocol to distill the output from parallel processors, it could be possible to reduce the number of final-measurement apparatuses in setups, decreasing the monetary cost of parallel-processor schemes.

One can also envision scenarios where the encoding and final measurements (which may include a premeasurement quantum computation using $\hat{\rho}_{\boldsymbol{\theta}}$ as an input [40]) are spatially separated and connected by quantum channels. Our distillation protocol allows the rate of quantum-state transmission to decrease, while keeping the average information flow constant. Similarly, if the measurement is to take place long after the interaction, our distillation protocol can reduce the number of quantum states that need to be stored in quantum memory predetection.

VI. QUASIPROBABILISTIC ANALYSIS

After having examined the practical aspects of our information-distillation protocol, we now turn to a foundational analysis. One might wonder, What is the fundamental resource that enables postselection to probabilistically boost the Fisher information? The answer, we show in this section, is *negativity*, a nonclassicality concept that stems from a certain type of noncommutation between observables [26,41,42].

So far, we have made few assumptions regarding the form of $\hat{U}(\boldsymbol{\theta})$. However, in many scenarios, $\hat{U}(\boldsymbol{\theta})$ will be composed of a series of M sequential unitary operators, $\hat{U}(\boldsymbol{\theta}) = \prod_{m=M}^{1} \hat{U}_m(\theta_m)$, where $\hat{U}_m(\theta_m)$ satisfies Stone's theorem on one-parameter unitary groups [43] such that $\hat{U}_m(\theta_m) = e^{i\theta_m \hat{A}_m}$ $\forall m \in 1, \dots, M$ [44]. The Hermitian generators $\hat{A}_m$ are in general noncommuting.

Here, we examine such sequential unitaries and use quasiprobabilistic techniques to bound $\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})$ with respective to classical and quantum statistics. Quasiprobability distributions are mathematical objects that behave similarly to probability distributions: They sum to unity, and marginalizing over all but one of the arguments yields a classical probability distribution. However, individual quasiprobabilities can be nonclassical by having values outside [0,1].

The complex-valued Kirkwood-Dirac (KD) quasiprobability distribution [45,46] is a relative of the Wigner function that can describe discrete systems—even qubits. The KD distribution has recently illuminated quantum effects in weak-value amplification [47–49], measurement disturbance [48,50–52], tomography [53–57], quantum chaos [49,58–62], metrology [19,20], thermodynamics [63,64], and the foundations of quantum mechanics [50,65–73]. By optimizing a formula with respect to a classical (real and non-negative) and a quantum (complex) Kirkwood-Dirac distribution, classical and quantum bounds can be found, respectively. Below we deploy this technique.

A KD distribution represents a quantum state $\hat{\rho}$ in terms of $k \geqslant 2$ sets of measurement operators. Equation (8) can be decomposed naturally in terms of a KD distribution defined by a discrete $\hat{\rho}$ and $k = 3$ sets of measurement operators. Two sets are composed of the projectors onto the subspaces of distinct eigenvalues of $\hat{\tilde{A}}_i$ and $\hat{\tilde{A}}_j$, and one set contains the postselection measurement operators:

$$\{\hat{\Pi}_k^{(i)} : \hat{\Pi}_k^{(i)}\hat{\tilde{A}}_i = a_k^{(i)}\hat{\Pi}_k^{(i)}\},$$
$$\{\hat{\Pi}_l^{(j)} : \hat{\Pi}_l^{(j)}\hat{\tilde{A}}_j = a_l^{(j)}\hat{\Pi}_l^{(j)}\},$$
$$\{\hat{F}_1 = \hat{F}, \hat{F}_2 = \hat{1} - \hat{F}\}.$$

Here, $\hat{\tilde{A}}_m \equiv [\prod_{i=M}^{m+1}\hat{U}_i(\theta_i)]\hat{A}_m[\prod_{j=m+1}^{M}\hat{U}_j^\dagger(\theta_j)]$ for $m < M$. For $m = M$, $\hat{\tilde{A}}_M \equiv \hat{A}_M$. We order the eigenvalues of $\hat{\tilde{A}}_i$ and $\hat{\tilde{A}}_j$ ascendingly, $a_1^{(i)} \leqslant \cdots \leqslant a_d^{(i)}$, and define the spectral eigengap $\Delta a^{(i)} \equiv a_d^{(i)} - a_1^{(i)}$, etc. We can now define our operational KD distribution with respect to the operators above:

$$\{q_{k,l,m}^{\hat{\rho}}\} \equiv \{\mathrm{Tr}[\hat{\Pi}_k^{(i)}\hat{F}_m\hat{\Pi}_l^{(j)}\hat{\rho}]\}. \quad (35)$$

The KD distribution obeys an analog of Bayes's theorem [49,53]. Consequently, we can define a distribution that corresponds to $\{q_{k,l,m}^{\hat{\rho}}\}$ conditioned on the postselection yielding outcome $\hat{F}$:

$$\{Q_{k,l}^{\hat{\rho}}\} \equiv \left\{\frac{q_{k,l,m=1}^{\hat{\rho}}}{\sum_{k,l,m=1} q_{k,l,m}^{\hat{\rho}}}\right\} = \{\mathrm{Tr}[\hat{\Pi}_k^{(i)}\hat{F}\hat{\Pi}_l^{(j)}\hat{\rho}]/p_{\boldsymbol{\theta}}^{\mathrm{ps}}\}. \quad (36)$$

Transforming the inner products in Eq. (8) to traces, we obtain

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) = 4\,\mathrm{Re}\left\{\frac{1}{p_{\boldsymbol{\theta}}^{\mathrm{ps}}}\mathrm{Tr}[\hat{F}\hat{\tilde{A}}_j\hat{\rho}_{\boldsymbol{\theta}}\hat{\tilde{A}}_i] - \frac{1}{(p_{\boldsymbol{\theta}}^{\mathrm{ps}})^2}\mathrm{Tr}[\hat{F}\hat{\rho}_{\boldsymbol{\theta}}\hat{\tilde{A}}_i]\mathrm{Tr}[\hat{F}\hat{\tilde{A}}_j\hat{\rho}_{\boldsymbol{\theta}}]\right\}. \quad (37)$$

Here, we have used that

$$|\partial_j\psi_{\boldsymbol{\theta}}\rangle = \partial_j\hat{U}(\boldsymbol{\theta})\,|\psi_0\rangle = \hat{\tilde{A}}_j\,|\psi_{\boldsymbol{\theta}}\rangle\,. \quad (38)$$

We can now use distribution (36) to recast Eq. (37):

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}) = 4\,\mathrm{Re}\left\{\sum_{k,l} a_k^{(i)}a_l^{(j)}Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}} - \left(\sum_{k',l'} a_{k'}^{(i)}Q_{k',l'}^{\hat{\rho}_{\boldsymbol{\theta}}}\right) \cdot \left(\sum_{k'',l''} a_{l''}^{(j)}Q_{k'',l''}^{\hat{\rho}_{\boldsymbol{\theta}}}\right)\right\}. \quad (39)$$

When $\{Q_{k,l}^{\hat{\rho}}\}$ is classical, all $|Q_{k,l}^{\hat{\rho}}| \leqslant 1$. Negative quasiprobabilities allow the denominators of Eq. (36) to approach 0 even for finite numerators. Then, $|Q_{k,l}^{\hat{\rho}}|$ can be arbitrarily large. Such negativity, an example above shows, enables $|\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})|$ to be anomalously large, compared with experiments described by classical distributions. This can increase distilled states' multiparameter information to nonclassically large values.

Theorem 2: Necessary condition for anomalous postselected quantumFisher information matrix. Suppose that a postselected quantum Fisher information matrix has some entry $|\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})| > \Delta a^{(i)}\Delta a^{(j)}$. Then, an underlying KD distribution $\{Q_{k,l}^{\hat{\rho}}\}$ necessarily contains at least one negative value.

Proof of Theorem 2. We prove this theorem by contradiction. First, we note that Eq. (39) is a quantum extension of a covariance, where $Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}}$ replaces classical joint probabilities. Second, we assume that $\{Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}}\}$ is classical. Third, ignoring the specific form of $\{Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}}\}$, we maximize and minimize Eq. (39) over all classical distributions. When $Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}} \in [0, 1]$ and $i \neq j$, Eq. (39) has the form of (four times) a classical covariance with maximum and minimum values $\Delta a^{(i)}\Delta a^{(j)}$ and $-\Delta a^{(i)}\Delta a^{(j)}$, respectively [74]. When $Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}} \in [0, 1]$ and $i = j$, Eq. (39) is upper bounded by $(\Delta a^{(i)})^2$ and lower bounded by 0 [19]. Per definition, an anomalous quantum Fisher information matrix (QFIM) entry breaks these bounds, such that the assumption of a classical distribution $\{Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}}\}$ cannot be satisfied. Consequently, if $|\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})| > \Delta a^{(i)}\Delta a^{(j)}$, then $\{Q_{k,l}^{\hat{\rho}_{\boldsymbol{\theta}}}\}$ is nonclassical. The form of Eq. (39) implies that any nonreal values cancel. Thus the nonclassicality must be in the form of negativity. ■

An immediate corollary follows.

Corollary 1. In a classically commuting theory, a theory in which operators commute, the quantum Fisher information matrix satisfies $|\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})| \leqslant \Delta a^{(i)}\Delta a^{(j)}$.

Proof. Reference [26] proves that noncommutation is necessary for nonclassical KD distributions [75]. The corollary thus follows from Theorem 2. ■

One might wonder if a filter applied before the information-encoding unitary could also generate arbitrarily large information encoded in the final quantum states. The answer is no. If the postselection filter is moved from after to before the unitary, the KD distribution is classical. Thus, by Theorem 2, the postselected quantum Fisher information is not anomalous.

VII. CONCLUSION

We have shown that postselection enables distillation of quantum learnability in a lossless fashion. The quantum Fisher information matrix enables scalar quantification of quantum learnability in multiparameter metrology and machine learning. We proved (Theorem 1) that there is no upper bound on how much multiparameter Fisher information can be compressed into a small number of states. From a practical perspective, our result generalizes, to the multiparameter-quantum-learnability regime, previous techniques from single-parameter postselected metrology and weak-value amplification. The implementation of our distillation protocol could mitigate the impact of detector imperfections and enable simplified setups in parallelized quantum schemes. Finally, we used quasiprobabilistic techniques to study experiments with sequential unitaries $\hat{U}(\boldsymbol{\theta}) = \prod_{m=M}^{1} e^{i\theta_m\hat{A}_m}$. Classically, we find that the entries in the postselected quantum Fisher information matrix are bounded by $|\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}})| > \Delta a^{(i)}\Delta a^{(j)}$, where $\Delta a^{(i)}$ is the spectral gap of $\hat{\tilde{A}}_i$. Via a quantum analog of Bayes's theorem, negative quasiprobabilities allow the entries to break these bounds

(Theorem 2). Thus nonclassical negativity underlies the anomalously large values of $\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}^{\text{ps}})$ found in the example above.

ACKNOWLEDGMENTS

The authors would like to thank Crispin Barnes, Rafal Demkowicz-Dobrzanski, Bobak Kiani, Helena Knowles, Aleks Lasek, Zi-Wen Liu, Seth Lloyd, Noah Lupu-Gladstein, Milad Marvian, Yordan Yordanov, and Nicole Yunger Halpern for useful discussions. This work was supported by the EPSRC, Lars Hierta's Memorial Foundation, and Girton College.

APPENDIX A: BOUNDS ON THE MOST INFORMATIVE SCALAR RISK FUNCTION

In the main text, we considered a scalar risk function with respect to the most informative experiment that extracts (Fisher) information from quantum states $\hat{\rho}_{\boldsymbol{\theta}}$:

$$s^{(\text{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W) \equiv s\Big(\max_{\mathcal{M}} I(\boldsymbol{\theta}), W\Big) = \frac{1}{N}\min_{\mathcal{M}} \text{Tr}[W I^{-1}(\boldsymbol{\theta})]. \quad \text{(A1)}$$

$\mathcal{M}$ is the set of all possible measurements. We argued that the quantum Fisher information matrix is the object of interest when finding bounds on $s^{(\text{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W)$. In this Appendix, we review results that support this claim.

The Fisher information matrix is upper bounded by the quantum Fisher information matrix [4,76–78]: $I(\boldsymbol{\theta}) \leqslant \mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})$. The quantum Fisher information matrix is defined by

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}}) = \text{Tr}(\hat{L}_j \partial_i \hat{\rho}_{\boldsymbol{\theta}}). \quad \text{(A2)}$$

Here, $\hat{L}_j$ is the logarithmic derivative operator, which is not uniquely defined [78]. It can be defined using a symmetric logarithmic derivative (SLD), $2\partial_i\hat{\rho}_{\boldsymbol{\theta}} = \hat{L}_i^{(\text{SLD})}\hat{\rho}_{\boldsymbol{\theta}} + \hat{\rho}_{\boldsymbol{\theta}}\hat{L}_i^{(\text{SLD})}$, or with a right logarithmic derivative (RLD), $\partial_i\hat{\rho}_{\boldsymbol{\theta}} = \hat{\rho}_{\boldsymbol{\theta}}\hat{L}_i^{(\text{RLD})}$. In the multiparameter scenario ($M > 1$), noncommutation often forbids measurements such that $I_{i,j}(\boldsymbol{\theta}) = \mathcal{I}_{i,j}(\boldsymbol{\theta})$ for all i, j. Thus $I(\boldsymbol{\theta}) \leqslant \mathcal{I}(\boldsymbol{\theta}|\hat{\rho}_{\boldsymbol{\theta}})$ cannot commonly be saturated. Either the symmetric-logarithmic-derivative or the right-logarithmic-derivative quantum Fisher information matrix can give a bound that lies closer to the achievable bound. For pure states, $\hat{L}_i^{(\text{SLD})} = 2\hat{L}_i^{(\text{RLD})} = 2\partial_i\hat{\rho}_{\boldsymbol{\theta}}$, and the symmetric-logarithmic-derivative quantum Fisher information matrix [Eq. (A2)] is

$$\mathcal{I}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) = 4\,\text{Re}[\langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle - \langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\,\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle], \quad \text{(A3)}$$

where $\hat{\rho}_{\boldsymbol{\theta}} \equiv |\psi_{\boldsymbol{\theta}}\rangle\,\langle\psi_{\boldsymbol{\theta}}|$ [78].

The quantum Fisher information matrix yields a scalar Cramér-Rao bound [4]:

$$s^{(\text{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W) \geqslant \frac{1}{N}\text{Tr}[W\mathcal{I}^{-1}(\boldsymbol{\theta})]. \quad \text{(A4)}$$

It is this bound that (directly or indirectly) leads quantum machine-learning algorithms to optimize the quantum Fisher information matrix of their subroutines [28–30]. However, Eq. (A4) "only" provides a lower bound on $s^{(\text{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W)$. Consequently, it is reasonable to ask, How good a measure of learnability is the quantum Fisher information matrix? From an information theoretic perspective, the answer [79,80] is given by

$$\frac{1}{N}\text{Tr}[W\mathcal{I}^{-1}(\boldsymbol{\theta})] \leqslant h(\boldsymbol{\theta}, W) \leqslant (1+\mathcal{Q})\frac{1}{N}\text{Tr}[W\mathcal{I}^{-1}(\boldsymbol{\theta})], \quad \text{(A5)}$$

where $h(\boldsymbol{\theta}, W)$ is Holevo's lower bound of the Cramér-Rao inequality [81]. The "geometric quantumness" measure $\mathcal{Q}$ satisfies $0 \leqslant \mathcal{Q} \leqslant 1$ [82]. Generally, it is hard to calculate $h(\boldsymbol{\theta}, W)$ (see Ref. [4] for the exact form). Nevertheless, for pure states, $s^{(\text{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W) = h(\boldsymbol{\theta}, W)$ [83].

For the purpose of the theoretical pure-state investigation in this paper, the formulas above can be summarized as

$$\frac{1}{N}\text{Tr}[W\mathcal{I}^{-1}(\boldsymbol{\theta})] \leqslant s^{(\text{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W) \leqslant 2\frac{1}{N}\text{Tr}[W\mathcal{I}^{-1}(\boldsymbol{\theta})]. \quad \text{(A6)}$$

Within a factor of 2, $\mathcal{I}(\boldsymbol{\theta})$ sets $s^{(\text{MI})}(\hat{\rho}_{\boldsymbol{\theta}}, W)$.

APPENDIX B: VALIDITY OF EQUATION (16)

In this Appendix we confirm that the estimated state $\hat{\rho}_{\boldsymbol{\theta}^0}$ satisfies $\hat{\rho}_{\boldsymbol{\theta}^0} \equiv \hat{U}(\boldsymbol{\theta}^0)\hat{\rho}_0\hat{U}^\dagger(\boldsymbol{\theta}^0) = \hat{\rho}_{\boldsymbol{\theta}} + i[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}] + O(|\boldsymbol{\delta}|^2)$ [Eq. (16)], by examining two very common forms of encoding unitary evolutions.

First, we assume that the unitary takes the form of $\hat{U}(\boldsymbol{\theta}) = \prod_{m=M}^{1}\hat{U}_m(\theta_m)$, where $\hat{U}_m(\theta_m) \equiv e^{i\theta_m\hat{A}_m}$ $\forall m \in 1, \ldots, M$ [44]. The Hermitian generators $\hat{A}_m$ are in general noncommuting. Then,

$$\hat{U}(\boldsymbol{\theta}^0) = \prod_{j=M}^{1}\hat{U}_j(\theta_j^0) \quad \text{(B1)}$$

$$= \prod_{j=M}^{1}\hat{U}_j(\theta_j)e^{-i\delta_j\hat{A}_j} \quad \text{(B2)}$$

$$= \prod_{j=M}^{1}\hat{U}_j(\theta_j)(\hat{1} - i\delta_j\hat{A}_j) + O(|\boldsymbol{\delta}|^2) \quad \text{(B3)}$$

$$= \hat{U}(\boldsymbol{\theta})\left(\hat{1} - i\sum_{j=1}^{M}\delta_j\hat{A}_j\right) + O(|\boldsymbol{\delta}|^2) \quad \text{(B4)}$$

$$= \hat{U}(\boldsymbol{\theta})(\hat{1} - i\hat{\gamma}) + O(|\boldsymbol{\delta}|^2). \quad \text{(B5)}$$

Here, we have defined the $O(\delta)$ Hermitian operator $\hat{\gamma} \equiv \sum_{j=1}^{M}\delta_j\hat{A}_j$. Similarly,

$$\hat{U}^\dagger(\boldsymbol{\theta}^0) = (\hat{1} + i\hat{\gamma})\hat{U}^\dagger(\boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2). \quad \text{(B6)}$$

We can now evaluate, to $O(|\boldsymbol{\delta}|^2)$, $\hat{\rho}_{\boldsymbol{\theta}^0}$:

$$\hat{\rho}_{\boldsymbol{\theta}^0} = \hat{U}(\boldsymbol{\theta}^0)\hat{\rho}_0\hat{U}^\dagger(\boldsymbol{\theta}^0) \quad \text{(B7)}$$

$$= \hat{U}(\boldsymbol{\theta})(\hat{1} - i\hat{\gamma})\hat{\rho}_0(\hat{1} + i\hat{\gamma})\hat{U}^\dagger(\boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2) \quad \text{(B8)}$$

$$= \hat{U}(\boldsymbol{\theta})(\hat{\rho}_0 + [\hat{\rho}_0, i\hat{\gamma}])\hat{U}^\dagger(\boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2). \quad \text{(B9)}$$

Setting $\hat{U}(\boldsymbol{\theta})\hat{\gamma}\hat{U}^\dagger(\boldsymbol{\theta}) \to \hat{D}$, we obtain the required form: $\hat{\rho}_{\boldsymbol{\theta}^0} = \hat{\rho}_{\boldsymbol{\theta}} + i[\hat{\rho}_{\boldsymbol{\theta}}, \hat{D}] + O(|\boldsymbol{\delta}|^2)$.

Second, we assume that the unitary takes the form of $\hat{U}(\boldsymbol{\theta}) = e^{i\sum_{j=1}^{M}\theta_j\hat{A}_j}$. Then,

$$\hat{U}(\boldsymbol{\theta}^0) = e^{i\sum_{j=1}^{M}\theta_j^0\hat{A}_j} \quad \text{(B10)}$$

$$= e^{i\sum_{j=1}^{M}\theta_j\hat{A}_j - i\sum_{k=1}^{M}\delta_k\hat{A}_k} \quad \text{(B11)}$$

$$\equiv e^{i\hat{H}_{\boldsymbol{\theta}} - i\hat{G}_\delta} \quad \text{(B12)}$$

$$= e^{i\hat{H}_{\boldsymbol{\theta}}} e^{-i\hat{G}_\delta} e^{\frac{1}{2}[\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]} e^{\frac{i}{6}[\hat{H}_{\boldsymbol{\theta}},\, [\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]]} + O(|\boldsymbol{\delta}|^2) \quad \text{(B13)}$$

$$= \hat{U}(\boldsymbol{\theta}) e^{-i\hat{G}_\delta} e^{\frac{1}{2}[\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]} e^{\frac{i}{6}[\hat{H}_{\boldsymbol{\theta}},\, [\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]]} + O(|\boldsymbol{\delta}|^2) \quad \text{(B14)}$$

$$= \hat{U}(\boldsymbol{\theta})(\hat{1} - i\hat{G}_\delta)\left(\hat{1} + \frac{1}{2}[\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]\right) \times \left(\hat{1} + \frac{i}{6}[\hat{H}_{\boldsymbol{\theta}},\, [\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]]\right) + O(|\boldsymbol{\delta}|^2) \quad \text{(B15)}$$

$$= \hat{U}(\boldsymbol{\theta})\left(\hat{1} - i\hat{G}_\delta + \frac{1}{2}[\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta] + \frac{i}{6}[\hat{H}_{\boldsymbol{\theta}},\, [\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]]\right) + O(|\boldsymbol{\delta}|^2) \quad \text{(B16)}$$

$$= \hat{U}(\boldsymbol{\theta})(\hat{1} - i\hat{\kappa}) + O(|\boldsymbol{\delta}|^2). \quad \text{(B17)}$$

Here, we have used Zassenhaus' formula and defined Hermitian operators $\hat{H}_{\boldsymbol{\theta}} \equiv \sum_{j=1}^{M}\theta_j\hat{A}_j$, $\hat{G}_\delta \equiv \sum_{k=1}^{M}\delta_k\hat{A}_kj$, and $\hat{\kappa} \equiv \hat{G}_\delta - i[\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]/2 - [\hat{H}_{\boldsymbol{\theta}},\, [\hat{H}_{\boldsymbol{\theta}},\, \hat{G}_\delta]]/6$. $\hat{\kappa}$ is $O(\delta)$. Similarly,

$$\hat{U}^\dagger(\boldsymbol{\theta}^0) = (\hat{1} + i\hat{\kappa})\hat{U}^\dagger(\boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2). \quad \text{(B18)}$$

As before, we can now evaluate, to $O(|\boldsymbol{\delta}|^2)$, $\hat{\rho}_{\boldsymbol{\theta}^0}$:

$$\hat{\rho}_{\boldsymbol{\theta}^0} = \hat{U}(\boldsymbol{\theta})(\hat{\rho}_0 + [\hat{\rho}_0, i\hat{\kappa}])\hat{U}^\dagger(\boldsymbol{\theta}) + O(|\boldsymbol{\delta}|^2). \quad \text{(B19)}$$

Setting $\hat{U}(\boldsymbol{\theta})\hat{\kappa}\hat{U}^\dagger(\boldsymbol{\theta}) \to \hat{D}$, we obtain the required form: $\hat{\rho}_{\boldsymbol{\theta}^0} = \hat{\rho}_{\boldsymbol{\theta}} + i[\hat{\rho}_{\boldsymbol{\theta}},\, \hat{D}] + O(|\boldsymbol{\delta}|^2)$.

APPENDIX C: POSTSELECTED GEOMETRIC QUANTUMNESS

The geometric quantumness measure $\mathcal{Q}$ in inequalities (A5) is given by

$$\mathcal{Q} = ||i\mathcal{I}^{-1}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})\mathcal{J}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})||_\infty, \quad \text{(C1)}$$

where $||X||_\infty$ denotes the largest eigenvalue of X. $\mathcal{J}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})$ is the Uhlmann curvature [84,85] given by

$$\mathcal{J}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) = 4\,\mathrm{Im}[\langle\partial_i\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle - \langle\partial_i\psi_{\boldsymbol{\theta}}|\psi_{\boldsymbol{\theta}}\rangle\,\langle\psi_{\boldsymbol{\theta}}|\partial_j\psi_{\boldsymbol{\theta}}\rangle]. \quad \text{(C2)}$$

The same tricks used in the main text can be used to show that

$$\mathcal{J}_{i,j}\big(\boldsymbol{\theta}\big|\psi_{\boldsymbol{\theta}}^{\mathrm{ps}}\big) = \frac{1}{t^2}\mathcal{J}_{i,j}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}}) + O(|\boldsymbol{\delta}|^2). \quad \text{(C3)}$$

Thus, at least to $O(|\boldsymbol{\delta}|^2)$, the geometric quantumness $\mathcal{Q}$ is constant with respect to the postselection.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photon. **5**, 222 (2011).
[2] M. Szczykulska, T. Baumgratz, and A. Datta, Adv. Phys.: X **1**, 621 (2016).
[3] B. T. Kiani, G. D. Palma, M. Marvian, Z.-W. Liu, and S. Lloyd, Quantum Sci. Technol. **7**, 045002 (2022).
[4] F. Albarelli, M. Barbieri, M. Genoni, and I. Gianani, Phys. Lett. A **384**, 126311 (2020).
[5] C. R. Rao, in *Breakthroughs in Statistics* (Springer, New York, 1992), pp. 235–247.
[6] H. Cramér, *Mathematical Methods of Statistics (PMS-9)*, Princeton Mathematical Series Vol. 9 (Princeton University Press, Princeton, 2016).
[7] M. A. Ballester, Phys. Rev. A **69**, 022303 (2004).
[8] H. Imai and A. Fujiwara, J. Phys. A: Math. Theor. **40**, 4391 (2007).
[9] M. G. Genoni, M. G. A. Paris, G. Adesso, H. Nha, P. L. Knight, and M. S. Kim, Phys. Rev. A **87**, 012107 (2013).
[10] P. C. Humphreys, M. Barbieri, A. Datta, and I. A. Walmsley, Phys. Rev. Lett. **111**, 070403 (2013).
[11] L. Pezzè, M. A. Ciampini, N. Spagnolo, P. C. Humphreys, A. Datta, I. A. Walmsley, M. Barbieri, F. Sciarrino, and A. Smerzi, Phys. Rev. Lett. **119**, 130504 (2017).
[12] Y. Chen and H. Yuan, New J. Phys. **19**, 063023 (2017).
[13] P. Liuzzo-Scorpo, L. A. Correa, F. A. Pollock, A. Górecka, K. Modi, and G. Adesso, New J. Phys. **20**, 063009 (2018).
[14] P. Lipka-Bartosik and R. Demkowicz-Dobrzański, J. Phys. A: Math. Theor. **51**, 474001 (2018).
[15] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Rev. Mod. Phys. **86**, 307 (2014).
[16] J. Harris, R. W. Boyd, and J. S. Lundeen, Phys. Rev. Lett. **118**, 070802 (2017).
[17] L. Xu, Z. Liu, A. Datta, G. C. Knee, J. S. Lundeen, Y.-Q. Lu, and L. Zhang, Phys. Rev. Lett. **125**, 080501 (2020).
[18] D. R. M. Arvidsson-Shukur and C. H. W. Barnes, Phys. Rev. A **99**, 060102(R) (2019).
[19] D. R. M. Arvidsson-Shukur, N. Yunger Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, Nat. Commun. **11**, 3775 (2020).
[20] N. Lupu-Gladstein, Y. B. Yilmaz, D. R. M. Arvidsson-Shukur, A. Brodutch, A. O. T. Pang, A. M. Steinberg, and N. Y. Halpern, Phys. Rev. Lett. **128**, 220504 (2022).
[21] D. R. M. Arvidsson-Shukur, A. G. McConnell, and N. Y. Halpern, arXiv:2207.07666.
[22] Initial studies of weak-value amplification with specific forms of multiparameter unitaries are given in Refs. [23–25].
[23] A. Vella, S. T. Head, T. G. Brown, and M. A. Alonso, Phys. Rev. Lett. **122**, 123603 (2019).
[24] B. Xia, J. Huang, C. Fang, H. Li, and G. Zeng, Phys. Rev. Appl. **13**, 034023 (2020).
[25] L. B. Ho and Y. Kondo, J. Math. Phys. **62**, 012102 (2021).
[26] D. R. M. Arvidsson-Shukur, J. C. Drori, and N. Y. Halpern, J. Phys. A: Math. Theor. **54**, 284001 (2021).
[27] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer, New York, 2006).
[28] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, Nat. Comput. Sci. **1**, 403 (2021).
[29] T. Haug, K. Bharti, and M. S. Kim, PRX Quantum **2**, 040309 (2021).
[30] J. J. Meyer, Quantum **5**, 539 (2021).
[31] Also in single-parameter metrology and weak-value amplification, conducting the optimal measurement generally requires a good initial estimate of the unknown parameter of interest. Moreover, many variational quantum algorithms, e.g., for

quantum computational chemistry, require good initial estimates: $\boldsymbol{\theta}^0 \approx \boldsymbol{\theta}$ [32–35].
[32] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Rev. Mod. Phys. **92**, 015003 (2020).
[33] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Commun. Phys. **4**, 228 (2021).
[34] W. Lavrijsen, A. Tudor, J. Müller, C. Iancu, and W. de Jong, in *Proceedings of the 2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Piscataway, NJ, 2020), pp. 267–277.
[35] L. Bittel and M. Kliesch, Phys. Rev. Lett. **127**, 120502 (2021).
[36] *Fundamentals of Kalman Filtering: A Practical Approach*, edited by P. Zarchan and H. Musoff, Progress in Astronautics and Aeronautics (American Institute of Aeronautics and Astronautics, Reston, VA, 2000).
[37] This choice of $\hat{K}$ generalizes the technique used by Lupu-Gladstein *et al.* in single-parameter metrology of optical qubits [20].
[38] We have assumed that the nonpostselected quantum Fisher information $\mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})$ is nonsingular. If it is singular, one has to remove the singularity-producing parameters from the analysis.
[39] N. Tomm, A. Javadi, N. O. Antoniadis, D. Najer, M. C. Löbl, A. R. Korsch, R. Schott, S. R. Valentin, A. D. Wieck, A. Ludwig, and R. J. Warburton, Nat. Nanotechnol. **16**, 399 (2021).
[40] B. T. Kiani, G. De Palma, D. Englund, W. Kaminsky, M. Marvian, and S. Lloyd, Phys. Rev. A **105**, 022415 (2022).
[41] S. De Bièvre, Phys. Rev. Lett. **127**, 190404 (2021).
[42] S. De Bievre, arXiv:2207.07451.
[43] M. H. Stone, Ann. Math. **33**, 643 (1932).
[44] Experiments where $\hat{U}_m(\theta_m) \neq e^{i\theta_m \hat{A}_m}$ can often be transformed into the required form via an artificial reparametrization.
[45] J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[46] P. A. M. Dirac, Rev. Mod. Phys. **17**, 195 (1945).
[47] A. M. Steinberg, Phys. Rev. A **52**, 32 (1995).
[48] J. Dressel, Phys. Rev. A **91**, 032116 (2015).
[49] N. Yunger Halpern, B. Swingle, and J. Dressel, Phys. Rev. A **97**, 042105 (2018).
[50] H. F. Hofmann, New J. Phys. **13**, 103009 (2011).
[51] J. Dressel and A. N. Jordan, Phys. Rev. A **85**, 012107 (2012).
[52] J. T. Monroe, N. Yunger Halpern, T. Lee, and K. W. Murch, Phys. Rev. Lett. **126**, 100403 (2021).
[53] L. M. Johansen, Phys. Rev. A **76**, 012119 (2007).
[54] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber, Nature (London) **474**, 188 (2011).
[55] J. S. Lundeen and C. Bamber, Phys. Rev. Lett. **108**, 070402 (2012).
[56] C. Bamber and J. S. Lundeen, Phys. Rev. Lett. **112**, 070405 (2014).
[57] G. S. Thekkadath, L. Giner, Y. Chalich, M. J. Horton, J. Banker, and J. S. Lundeen, Phys. Rev. Lett. **117**, 120401 (2016).
[58] B. Swingle, G. Bentsen, M. Schleier-Smith, and P. Hayden, Phys. Rev. A **94**, 040302(R) (2016).
[59] N. Y. Halpern, A. Bartolotta, and J. Pollack, Commun. Phys. **2**, 92 (2019).
[60] J. R. González Alonso, N. Yunger Halpern, and J. Dressel, Phys. Rev. Lett. **122**, 040404 (2019).
[61] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe, Nature (London) **567**, 61 (2019).
[62] R. Mohseninia, J. R. G. Alonso, and J. Dressel, Phys. Rev. A **100**, 062336 (2019).
[63] A. Levy and M. Lostaglio, PRX Quantum **1**, 010309 (2020).
[64] M. Lostaglio, Phys. Rev. Lett. **125**, 230603 (2020).
[65] R. B. Griffiths, J. Stat. Phys. **36**, 219 (1984).
[66] S. Goldstein and D. N. Page, Phys. Rev. Lett. **74**, 3715 (1995).
[67] J. B. Hartle, Phys. Rev. A **70**, 022104 (2004).
[68] H. F. Hofmann, New J. Phys. **14**, 043031 (2012).
[69] H. F. Hofmann, Phys. Rev. A **89**, 042115 (2014).
[70] H. F. Hofmann, Phys. Rev. A **91**, 062123 (2015).
[71] H. F. Hofmann, Eur. Phys. J. D **70**, 118 (2016).
[72] J. J. Halliwell, Phys. Rev. A **93**, 022123 (2016).
[73] B. C. Stacey, arXiv:1907.02432.
[74] Applying the Cauchy-Schwarz inequality to a covariance of random variables X and Y yields $|\mathrm{Cov}(X,Y)| \leqslant \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$.
[75] In fact, noncommutation is necessary, but not sufficient, for KD nonclassicality [26].
[76] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[77] C. Helstrom, Phys. Lett. A **25**, 101 (1967).
[78] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, J. Phys. A: Math. Theor. **53**, 023001 (2020)..
[79] F. Albarelli, J. F. Friel, and A. Datta, Phys. Rev. Lett. **123**, 200503 (2019).
[80] A. Carollo, B. Spagnolo, A. A. Dubkov, and D. Valenti, J. Stat. Mech. (2019) 094010.
[81] A. Holevo, Rep. Math. Phys. **12**, 251 (1977).
[82] Appendix C shows that the geometric quantumness $\mathcal{Q}$ in inequalities (A5) is constant [to $O(|\boldsymbol{\delta}|^2)$] with respect to the postselection.
[83] K. Matsumoto, J. Phys. A: Math. Gen. **35**, 3111 (2002).
[84] For pure states, $\mathcal{J}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})$ is (four times) the imaginary part of the quantum geometric tensor. $\mathcal{I}(\boldsymbol{\theta}|\psi_{\boldsymbol{\theta}})$ is (four times) the real part.
[85] A. Carollo, B. Spagnolo, and D. Valenti, Sci. Rep. **8**, 9852 (2018).