# Estimation of quantum finite mixtures

J. I. de Vicente, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan

*Grup de Física Teòrica, Facultat de Ciències, Edifici Cn, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain*
(Received 8 October 2009; published 29 January 2010)

We consider the problem of determining the weights of a quantum ensemble. That is to say, given a quantum system that is in a set of possible *known* states according to an *unknown* probability law, we give strategies to estimate the individual probabilities, weights, or mixing proportions. Such strategies can be used to estimate the frequencies at which different independent signals are emitted by a source. They can also be used to estimate the weights of particular terms in a canonical decomposition of a quantum channel. The quality of these strategies is quantified by a covariance-type error matrix. According with this cost function, we give optimal strategies in both the single-shot and multiple-copy scenarios. The latter is also analyzed in the asymptotic limit of large number of copies. We give closed expressions of the error matrix for two-component quantum mixtures of qubit systems. The Fisher information plays an unusual role in the problem at hand, providing exact expressions of the minimum covariance matrix for any number of copies.

## I. INTRODUCTION

Suppose we are given a quantum system which is known to be in one of several states with some unknown probability, such as a photon that travels through a communication channel and codifies some message. These states can be nonorthogonal due to, e.g., errors occurring during the transmission, but can also be made to overlap intentionally, e.g., to avoid possible eavesdropper attacks in quantum key distribution. Given this set of possible fixed states, we wish to find an estimate of the probabilities that best describe the state we have been provided with. More succinctly, assuming that a state $\rho_\lambda$ is a convex combination of a given set of states $\{\rho_r\}$,

$$\rho_\lambda = \sum_{r=1}^{M} \lambda_r \rho_r, \tag{1}$$

we wish to best estimate the value of the weights $\{\lambda_r\}$, which we arrange in the column vector $\boldsymbol{\lambda}$ and characterize the quantum ensemble $\{(\lambda_r, \rho_r)\}$, by performing suitable measurements on the system.

The analogous classical problem appears in the field of statistical modeling under the name of estimation of *finite mixtures* [1]. The formal study of finite mixtures was initiated by Pearson in 1894 [2]. He was conducting a biometric investigation on data collected from crabs, and found that the distribution of the size of their forehead (relative to the size of the body) presented an unexpected skewness, which could not be modeled with a symmetric normal distribution. Pearson showed that the data were very well fitted by a mixture of two normal distributions. The presence of two components was taken by Pearson as evidence that there were two different species of crabs. In this way finite mixture models can be used to expose any grouping in underlying data (clustering of data). With the prior knowledge on the individual component densities, which can be inferred or estimated by other means, finite mixture estimation enables one to estimate the weights, or proportions, of the different populations from the gathered coarse-grained data. A (classical) finite mixture, $p_\lambda(i) = \sum \lambda_r p_r(i)$ (in the obvious notation), can thus always be interpreted as describing situations where the information

on the grouping is lost, or in other words, as marginals of a joint distribution $p(i, r)$, such that $p_\lambda(i) = \sum_r p(i, r)$, i.e., $p_r(i)$ can be viewed as the conditioned probability $p_r(i) = p(i|r)$.

In this paper we approach the problem of estimating *quantum finite mixtures*. More precisely, we give optimal strategies to estimate the vector of weights $\boldsymbol{\lambda}$ under the assumptions given above (known set $\{\rho_r\}$ of possible states). We address also the situation in which we are provided with $N$ identical and independent copies of the state $\rho_\lambda$, to which we will refer as *average state*. In this multiple-copy scenario, we further assume that generalized collective measurements can be performed on $\rho_\lambda^{\otimes N}$. For large $N$, we also give (local) strategies based on projective measurements on individual copies that have the same performance as the optimal collective strategies.

Quantum ensembles are necessary to describe situations in which complete prior information is lacking. In the context of quantum communication, for instance, one estimates the frequency of different (known) states coming out of a source, i.e., one gathers information from the average state in connection to its particular preparation procedure. It is well known that in general there is no unique quantum ensemble consistent with a given mixed state [3]. Therefore there will be instances in quantum finite mixture estimation, called unidentifiable, where the average state $\rho_\lambda$ does not fully determine the value of the weights $\boldsymbol{\lambda}$, which therefore cannot be estimated with unlimited precision even when an arbitrary number of copies of $\rho_\lambda$ is provided. This problem is related to that of discrimination of quantum ensembles [4], where it is necessary to consider as inequivalent the different ensembles that are consistent with a given mixed state. We also note that, as in the classical case, a quantum finite mixture can be interpreted as the marginal density matrix of an extended system-ancilla state when a particular measurement is done on the ancilla. A quantum ensemble also describes the output of a stochastic quantum channel (or generalized measurement) for a fixed input state. In particular, if the input state is taken to be one part of a bipartite maximally entangled state, the stochastic channel is fully characterized by the output state, and it can be interpreted as a quantum finite mixture. Therefore, the results that we present here can be applied to the estimation of the weights of

the individual (or of a subset of) Kraus operators in a particular operator sum representation of a channel. For example, we can easily give bounds on the precision of estimating the weight of bit flip, phase flip, and combined bit-phase flip errors, or also the total weight of two-qubit Pauli errors versus single qubit Pauli errors.

Quantum finite mixture estimation is a novel ground for quantum estimation theory [5–8], which is one of the basic tools in the field of quantum information and has been continuously developing since the late 1970s. Many problems have been addressed, ranging from the estimation of a single parameter—as, e.g., a phase [8], or the losses of a quantum channel [9]—to full tomography. Quantum estimation theory finds also many applications in quantum metrology [10]—such as improvement of frequency standards [11], gravitational-wave detection [12,13], and clock synchronization [14,15]—and it is often a key ingredient in other quantum computation [16] and communication topics, e.g., quantum benchmarks for teleportation experiments [17]. The recent problem studied by Konrad *et al.* [18] can be viewed as a quantum finite mixture estimation in a simplified context. In the present paper we address the issue in full generality. This, in passing, will enable us to answer most of the questions posed there.

The paper is organized as follows. In Sec. II we introduce the general framework and give the main results for both, the single- and multiple-copy scenarios. The asymptotic limit of large number of copies is addressed in Sec. III. The two sections conclude with a discussion on unidentifiability of mixtures and its consequences. Additionally, in each of these sections, we provide examples to illustrate the use of the techniques that we introduce. Section IV is devoted to two-component mixtures, where closed expressions can be given for rather general situations. The conclusions are in Sec. V and several technical details can be found in the Appendices, which also include an example of a two-step adaptive local strategy that is optimal.

## II. ESTIMATION OF WEIGHTS IN FINITE MIXTURES

### A. General framework

As already mentioned in the Introduction, a quantum finite mixture is defined to be the convex combination in Eq. (1), where $\boldsymbol{\lambda}$ belongs to the unit $(M-1)$ simplex (i.e., the set $\{\boldsymbol{\lambda} : \lambda_r \geqslant 0, \sum_{r=1}^{M} \lambda_r = 1\}$). By quantum finite mixture estimation we mean the following: assume we have been provided with a copy of the average state $\rho_{\boldsymbol{\lambda}}$ (or with several identical and independent copies of it; i.e., with $\rho_{\boldsymbol{\lambda}}^{\otimes N}$), of which we know nothing about the actual value of $\boldsymbol{\lambda}$ but that it has been drawn from a (prior) probability distribution $\pi(\boldsymbol{\lambda})$. Assume also that we are allowed to perform generalized measurements on the copy (or copies) of $\rho_{\boldsymbol{\lambda}}$. Our task is to determine $\boldsymbol{\lambda}$ (or, maybe, some linear combinations of its components $\lambda_r$; namely, $a = \boldsymbol{a}^t \boldsymbol{\lambda}$, where $\boldsymbol{a}$ is some vector of constants $a_r$). This has necessarily to be based on the output(s) of our measurement(s) on $\rho_{\boldsymbol{\lambda}}$ ($\rho_{\boldsymbol{\lambda}}^{\otimes N}$). Due to the inherent nature of quantum measurements, the determination of $\boldsymbol{\lambda}$ cannot be perfect and we can only hope to obtain an estimate within some accuracy. Our goal is to obtain the best estimate.

To give a precise meaning to the term 'best estimate' we take a Bayesian approach and introduce as cost function the covariance-type error matrix

$$\boldsymbol{\Delta} = \langle (\boldsymbol{\lambda} - \boldsymbol{\lambda}_\chi)(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\chi)^t \rangle, \tag{2}$$

where $\boldsymbol{\lambda}_\chi$ is our estimate of $\boldsymbol{\lambda}$ based on the outcome $\chi$ of our measurement and $\langle \cdot \rangle$ stands for averaging over $\boldsymbol{\lambda}$ and $\chi$. More precisely, the averaging is performed over the joint probability distribution $p(\chi, \boldsymbol{\lambda}) = p(\chi|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda})$, where $p(\chi|\boldsymbol{\lambda})$ is the probability of obtaining the outcome $\chi$ conditioned to the actual value of $\boldsymbol{\lambda}$. In Quantum Mechanics this conditional probability is given by Born's rule: $p(\chi|\boldsymbol{\lambda}) = \mathrm{tr}\, E_\chi \rho_{\boldsymbol{\lambda}}$, where $\{E_\chi\}$ is the positive operator-valued measure (POVM) that defines our generalized quantum measurement. The trace of the error matrix $\boldsymbol{\Delta}$ gives the total mean square error (MSE), $E = \mathrm{tr}\, \boldsymbol{\Delta}$, while the expectation value $E_a = \boldsymbol{a}^t \boldsymbol{\Delta} \boldsymbol{a}$ gives the mean square error in the estimation of $a$.

In order to analyze one-copy and multiple-copy estimation in a unified framework we have found it convenient to define quantum finite mixtures, Eq. (1), in a slightly more general form, allowing for nonlinear mixtures of the type

$$\rho_{\boldsymbol{\lambda}} = \sum_\alpha c_\alpha(\boldsymbol{\lambda}) \rho_\alpha, \tag{3}$$

where the coefficient functions satisfy $\sum_\alpha c_\alpha(\boldsymbol{\lambda}) = 1$ for all $\boldsymbol{\lambda}$ [but not necessarily $c_\alpha(\boldsymbol{\lambda}) \geqslant 0$], and the range of values for $\alpha$ may not coincide with that for $r$ in Eq. (1). As for linear finite mixtures, our goal still is to best estimate $\boldsymbol{\lambda}$ (we assume that the functional dependence of the coefficient functions $c_\alpha$ on $\boldsymbol{\lambda}$ is known).

The error matrix $\boldsymbol{\Delta}$ can be written as

$$\boldsymbol{\Delta} = \sum_\chi p(\chi) \langle (\boldsymbol{\lambda} - \boldsymbol{\lambda}_\chi)(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\chi)^t \rangle_\chi, \tag{4}$$

where $p(\chi)$ is the marginal of $p(\chi, \boldsymbol{\lambda})$ and $\langle \cdot \rangle_\chi$ indicates averaging over the conditional probability $p(\boldsymbol{\lambda}|\chi) = p(\chi, \boldsymbol{\lambda})/p(\chi)$ (Bayes rule). More explicitly, $p(\chi) = \int d\boldsymbol{\lambda}\, p(\chi, \boldsymbol{\lambda})$, where we use the shorthand notation $d\boldsymbol{\lambda} = \delta(\sum_r \lambda_r - 1) \prod_r d\lambda_r$. Note that the Dirac $\delta$ function, along with $\lambda_r \geqslant 0$, guarantees that $\boldsymbol{\lambda}$ is a point in the unit $(M-1)$ simplex (hereafter, simplex for brevity). Equation (4) can be cast as

$$\boldsymbol{\Delta} = \sum_\chi p(\chi) \{ \langle (\boldsymbol{\lambda} - \langle \boldsymbol{\lambda} \rangle_\chi)(\boldsymbol{\lambda} - \langle \boldsymbol{\lambda} \rangle_\chi)^t \rangle_\chi + \boldsymbol{\delta}_\chi \boldsymbol{\delta}_\chi^t \}, \tag{5}$$

with $\boldsymbol{\delta}_\chi = \langle \boldsymbol{\lambda} \rangle_\chi - \boldsymbol{\lambda}_\chi$. Note that all dependence on our particular choice of the estimator $\boldsymbol{\lambda}_\chi$ is contained in $\boldsymbol{\delta}_\chi$. Since the matrix $\boldsymbol{\delta}_\chi \boldsymbol{\delta}_\chi^t$ is manifestly positive semidefinite, the estimator that minimizes our cost function $\boldsymbol{\Delta}$ is

$$\boldsymbol{\lambda}_\chi = \langle \boldsymbol{\lambda} \rangle_\chi = \int d\boldsymbol{\lambda}\, p(\boldsymbol{\lambda}|\chi) \boldsymbol{\lambda} = \frac{\int d\boldsymbol{\lambda}\, \pi(\boldsymbol{\lambda}) \boldsymbol{\lambda}\, \mathrm{tr}\, E_\chi \rho_{\boldsymbol{\lambda}}}{\int d\boldsymbol{\lambda}\, \pi(\boldsymbol{\lambda})\, \mathrm{tr}\, E_\chi \rho_{\boldsymbol{\lambda}}}. \tag{6}$$

(Note that the components of $\boldsymbol{\lambda}_\chi$ are non-negative and add up to one; i.e., $\boldsymbol{\lambda}_\chi$ is a probability vector.) Hereafter, we will only consider this optimal estimator, which gives the smallest error matrix. We will denote this matrix by the same symbol $\boldsymbol{\Delta}$ to simplify the notation. Hence, we may write

$$\boldsymbol{\Delta} = \sum_\chi p(\chi) \boldsymbol{\Delta}_\chi. \tag{7}$$

By rearranging the remaining terms in Eq. (5) one can further simplify the expression of the error matrix to obtain

$$\boldsymbol{\Delta} = \langle \boldsymbol{\lambda}\boldsymbol{\lambda}^t \rangle - \sum_\chi p(\chi) \langle \boldsymbol{\lambda} \rangle_\chi \langle \boldsymbol{\lambda}^t \rangle_\chi, \qquad (8)$$

where it is important to note that the first average is over the prior distribution $\pi(\boldsymbol{\lambda})$ alone, i.e., independent of the measurements we may perform on the average state. As to the second term, we may write the average value of $\boldsymbol{\lambda}$ as

$$\langle \boldsymbol{\lambda} \rangle_\chi = \sum_\alpha \frac{\tilde{\boldsymbol{\Lambda}}_\alpha \, \mathrm{tr} E_\chi \rho_\alpha}{p(\chi)} + \langle \boldsymbol{\lambda} \rangle, \qquad (9)$$

where we have defined $\tilde{\boldsymbol{\Lambda}}_\alpha = \langle \boldsymbol{\lambda} c_\alpha(\boldsymbol{\lambda}) \rangle - \langle \boldsymbol{\lambda} \rangle \langle c_\alpha(\boldsymbol{\lambda}) \rangle$ and used that $\sum_\alpha \langle c_\alpha(\boldsymbol{\lambda}) \rangle \mathrm{tr}\, E_\chi \rho_\alpha = p(\chi)$. Inserting this result in Eq. (8) we find

$$\boldsymbol{\Delta} = \boldsymbol{\Lambda} - \sum_\chi \frac{\left( \sum_\alpha \tilde{\boldsymbol{\Lambda}}_\alpha \, \mathrm{tr} E_\chi \rho_\alpha \right) \left( \sum_\beta \tilde{\boldsymbol{\Lambda}}_\beta^t \, \mathrm{tr} E_\chi \rho_\beta \right)}{p(\chi)}, \qquad (10)$$

where $\boldsymbol{\Lambda} = \langle \boldsymbol{\lambda}\boldsymbol{\lambda}^t \rangle - \langle \boldsymbol{\lambda} \rangle \langle \boldsymbol{\lambda}^t \rangle$ is the covariance matrix of the unknown weights, i.e., its elements are the second-order moments of the prior distribution $\pi(\boldsymbol{\lambda})$. In order to interpret the second term in this equation, we define an *effective* state $\sigma_{\boldsymbol{\lambda}}$ that combines information relative to the prior distribution of $\boldsymbol{\lambda}$ with the quantum states $\rho_\alpha$:

$$\sigma_{\boldsymbol{\lambda}} = \langle \rho_{\boldsymbol{\lambda}} \rangle + (\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}})^t \left( \sum_\alpha \tilde{\boldsymbol{\Lambda}}_\alpha \rho_\alpha \right), \qquad (11)$$

where we have defined $\bar{\boldsymbol{\lambda}} = \langle \boldsymbol{\lambda} \rangle$. It is shown in Appendix A that this equation defines a proper density matrix. Let $p_{\boldsymbol{\lambda}}(\chi)$ be the probability distribution of the outcomes obtained when performing the POVM measurement $\{E_\chi\}$ on this effective state, namely $p_{\boldsymbol{\lambda}}(\chi) = \mathrm{tr}\, E_\chi \sigma_{\boldsymbol{\lambda}}$. Then, Eq. (10) can be written in a very appealing form as

$$\boldsymbol{\Delta} = \boldsymbol{\Lambda} - \boldsymbol{F}(\bar{\boldsymbol{\lambda}}), \qquad (12)$$

where $\boldsymbol{F}(\boldsymbol{\lambda})$ is the Fisher information matrix of the probability distribution $p_{\boldsymbol{\lambda}}(\chi)$, whose elements are defined by

$$F_{rs}(\boldsymbol{\lambda}) = \sum_\chi \frac{\partial_r p_{\boldsymbol{\lambda}}(\chi) \partial_s p_{\boldsymbol{\lambda}}(\chi)}{p_{\boldsymbol{\lambda}}(\chi)}, \qquad (13)$$

and we use the compact notation $\partial_r = \partial/\partial \lambda_r$. Some comments are in order. Note that the error matrix $\boldsymbol{\Delta}$ has two distinct contributions: (i) the intrinsic 'error' of the random variable $\boldsymbol{\lambda}$ (that one would obtained by just guessing the weights of the quantum finite mixture without performing any measurement whatsoever), which is given by the covariance matrix $\boldsymbol{\Lambda}$; and (ii) the Fisher information of the effective state $\sigma_{\boldsymbol{\lambda}}$, which represents the information gathered from the outcomes of the measurement on the average state $\rho_{\boldsymbol{\lambda}}$. Naturally, this information reduces the uncertainty on the actual value of $\boldsymbol{\lambda}$, which explains the minus sign in Eq. (12). Despite this very natural interpretation, one might be somehow surprised to find the Fisher information matrix in the context of Eq. (12). It usually appears in connection to the Cramér-Rao bound (see Sec. III A below), where it provides lower bounds to the MSE in estimation problems. Typically these lower bounds are attained only in the asymptotic limit of many identical and independent copies. Note however that relation (12) is an exact expression.

More interestingly for our purposes here, relation (12) enables us to apply known results [5,6] concerning the Fisher information. In particular, the Braunstein and Caves inequality [19], which states that the Fisher information is upper bounded by the so-called quantum Fisher information (QFI) matrix $\boldsymbol{H}(\boldsymbol{\lambda})$. Thus,

$$\boldsymbol{\Delta} \geqslant \boldsymbol{\Lambda} - \boldsymbol{H}(\bar{\boldsymbol{\lambda}}). \qquad (14)$$

Before proceeding, we recall the definition of $\boldsymbol{H}(\boldsymbol{\lambda})$. Its matrix elements, which depend only on the family of states $\sigma_{\boldsymbol{\lambda}}$, are given by

$$H_{rs}(\boldsymbol{\lambda}) = \mathrm{Re}\, \mathrm{tr}\, [L_r(\boldsymbol{\lambda}) L_s(\boldsymbol{\lambda}) \sigma_{\boldsymbol{\lambda}}], \qquad (15)$$

where the matrix $L_r(\boldsymbol{\lambda})$ is the symmetric logarithmic derivative (SLD), (implicitly) defined as

$$\tfrac{1}{2}[L_r(\boldsymbol{\lambda})\sigma_{\boldsymbol{\lambda}} + \sigma_{\boldsymbol{\lambda}} L_r(\boldsymbol{\lambda})] = \partial_r \sigma_{\boldsymbol{\lambda}}. \qquad (16)$$

Although Eqs. (15) and (16) are particularized to the case under consideration, they also apply to a general situation where $\sigma_{\boldsymbol{\lambda}}$ represents an arbitrary family of states, such as that defined by $\rho_{\boldsymbol{\lambda}}$. We also recall that the SLD is most easily computed in the basis that diagonalizes $\sigma_{\boldsymbol{\lambda}}$. A simple calculation leads to

$$\boldsymbol{L}(\boldsymbol{\lambda}) = 2 \sum_{n,m} \frac{\langle \phi_n | \sum_\alpha \tilde{\boldsymbol{\Lambda}}_\alpha \rho_\alpha |\phi_m\rangle}{v_m + v_n} |\phi_n\rangle\langle\phi_m|, \qquad (17)$$

where $\{|\phi_n\rangle\}$ and $v_n$ are the eigenvectors and eigenvalues of $\sigma(\bar{\boldsymbol{\lambda}}) = \langle \rho_{\boldsymbol{\lambda}} \rangle$, respectively.

Let us go back to Eq. (14). Since $\boldsymbol{H}(\bar{\boldsymbol{\lambda}})$ is independent of the measurement (as pointed out above, it only depends on the effective state $\sigma_{\boldsymbol{\lambda}}$), Eq. (14) provides an absolute lower bound to the error matrix $\boldsymbol{\Delta}$.

In those cases where this lower bound is attainable [such as the dimension two case, where $\boldsymbol{\lambda} = (\lambda, 1 - \lambda)^t$, or when the SLD matrices commute with one another], the QFI matrix further provides us with the optimal measurement. In those cases $\{E_\chi\}$ can be chosen to be the projectors onto the eigenspaces of $L_r(\bar{\boldsymbol{\lambda}})$. An important instance is the estimation of the linear combination $a = \boldsymbol{a}^t \boldsymbol{\lambda}$. In this case the optimal measurement is given by the projector onto the eigenspaces of $L_a = \sum_r a_r L_r(\bar{\boldsymbol{\lambda}})$, and the minimal error $E_a$ is exactly given by

$$E_a = \boldsymbol{a}^t \boldsymbol{\Lambda} \boldsymbol{a} - 2 \sum_{n,m} \frac{\left| \langle \phi_m | \sum_\alpha \boldsymbol{a}^t \tilde{\boldsymbol{\Lambda}}_\alpha \rho_\alpha |\phi_n\rangle \right|^2}{v_m + v_n}, \qquad (18)$$

which comes from sandwiching Eq. (12) with $\boldsymbol{a}^t$ and $\boldsymbol{a}$. In particular, the MSE on a single weight $\lambda_r$ is given by

$$\Delta_{rr} = \Lambda_{rr} - 2 \sum_{n,m} \frac{\left| \langle \phi_m | \sum_\alpha \tilde{\Lambda}_{\alpha r} \rho_\alpha |\phi_n\rangle \right|^2}{v_m + v_n}. \qquad (19)$$

Quantum finite mixtures of orthogonal states ($\rho_\alpha \rho_\beta = 0$) is yet another instance where the bound (14) is attainable. In this case, one can easily check that the MSE is simply given by

$$E^\perp = \mathrm{tr}\boldsymbol{\Delta} = \mathrm{tr}\boldsymbol{\Lambda} + \sum_\alpha \frac{\tilde{\boldsymbol{\Lambda}}_\alpha^t \tilde{\boldsymbol{\Lambda}}_\alpha}{\langle c_\alpha(\boldsymbol{\lambda}) \rangle}. \qquad (20)$$

### B. Estimation with multiple copies

Let us assume that we are given an arbitrary number $N$ of identical and independent copies of the average state $\rho_\lambda$ in Eq. (1). The global state of the $N$ copies can be written as

$$\rho_\lambda^{\otimes N} = N! \sum_{\boldsymbol{k}} \prod_r \frac{\lambda_r^{k_r}}{k_r!} \mathcal{S}\big(\rho_1^{\otimes k_1} \otimes \cdots \otimes \rho_M^{\otimes k_M}\big), \quad (21)$$

where the components of the 'occupation number' vector $\boldsymbol{k}$ satisfy $\sum_{r=1}^{M} k_r = N$, and $\mathcal{S}$ indicates averaging over all permutations of the $N$ copies, which produces a proper (normalized) state. From this equation we note that the state $\rho_\lambda^{\otimes N}$ can be written in the form (3) with $\boldsymbol{k}$ playing the role of $\alpha$ and $c_{\boldsymbol{k}}(\lambda) = N! \prod_r \lambda_r^{k_r}/k_r!$, $\rho_{\boldsymbol{k}} = \mathcal{S}(\rho_1^{\otimes k_1} \otimes \cdots \otimes \rho_M^{\otimes k_M})$. Because of this, the results of the previous section can be applied to multiple copies.

For arbitrary prior distributions $\pi(\lambda)$ that is about all we can say concerning the multiple copy scenario. However, more explicit expression can be derived if a flat distribution of weights can be assumed. This is the most conservative scenario, and also the situation when nothing is known a priori about the weights $\lambda$. Appendix B collects useful formulas for computing integrals and averages on the simplex when $\pi(\lambda)$ is flat (constant). From this appendix one can easily obtain

$$\Lambda_{rs} = \frac{\delta_{rs} - 1/M}{M(M+1)}, \quad (22)$$

$$\tilde{\Lambda}_{\boldsymbol{k}r} = \frac{k_r - N/M}{(N+M)} \langle c_{\boldsymbol{k}}(\lambda) \rangle, \quad (23)$$

for the matrix elements of $\boldsymbol{\Lambda}$ and $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{k}}$, respectively, where

$$\langle c_{\boldsymbol{k}}(\lambda) \rangle = \binom{N+M-1}{N}^{-1}. \quad (24)$$

Hence, the lower bound on the MSE follows:

$$\text{tr}\boldsymbol{\Delta} \geqslant \frac{M-1}{M(M+1)} - 2 \sum_{n,m} \frac{\big|\langle \phi_m| \sum_{\boldsymbol{k}} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{k}} \rho_{\boldsymbol{k}} |\phi_n\rangle\big|^2}{\nu_m + \nu_n}. \quad (25)$$

This is as far as one can get for mixtures of arbitrary states $\{\rho_r\}$. In the case of mixtures of orthogonal states we can substitute Eqs. (22) to (24) in Eq. (20) and find a closed expression for the MSE for multiple copies:

$$E_N^\perp = \text{tr}\boldsymbol{\Delta} = \frac{M-1}{(M+1)(M+N)}, \quad (26)$$

where we have used the summation formula in Appendix B. Note that the error $E_N^\perp$ vanishes as $N$ goes to infinity.

### C. Identifiability

A mixture is *identifiable* if there exists a one-to-one correspondence between $\lambda$ and $\rho_\lambda$. That is to say, if and only if given $\rho_\lambda$, there is no other vector of weights $\lambda$ satisfying Eq. (1). In a general situation, though, different vectors $\lambda$ can give rise to the same density matrix ($\rho_\lambda = \rho_{\lambda'}$ for some $\lambda \neq \lambda'$) and, therefore, identifiability cannot be taken for granted. Necessary and sufficient conditions for identifiability of classical finite mixtures, $p_\lambda(i) = \sum \lambda_r p_r(i)$, were established more than four decades ago by Teicher [20]. These conditions are equivalent to $\{p_r\}_{r=1}^{M}$ being a linearly independent set. Similarly, the linear independence of the (density) matrices in a quantum ensemble $\{\rho_r\}_{r=1}^{M}$, constitutes a necessary and sufficient condition for the identifiability of quantum finite mixtures: states lying in the convex hull of a linearly independent set of density matrices will be identifiable, while all states in the convex hull of a linearly dependent set will necessarily be unidentifiable, except for possibly some states on the boundary.

Identifiability is usually assumed in (classical) mixture estimation (see, e.g., [21]), since unidentifiable models often give rise to ill-defined estimation procedures and their asymptotic theories break down. In contrast, our approach leads to sensible results for the estimation of quantum finite mixtures *even in unidentifiable scenarios*. The above results for single-copy case, as well as the derivation of the effective model for finite number of copies, can be directly applied without taking notice of identifiability considerations. Care must be taken, however, when applying the asymptotic methods of the next section to unidentifiable mixtures. Such methods assume that the errors go to zero as the number of copies increases, which cannot be guaranteed if mixtures are unidentifiable. We will revisit unidentifiability at the end of Sec. III, where we will introduce ways to circumvent this difficulty.

### III. ESTIMATION OF WEIGHTS IN THE ASYMPTOTIC LIMIT

In the preceding sections we have presented protocols to optimally estimate quantum finite mixtures and have obtained bounds on their accuracy using a covariance-type error matrix as a cost functions. We have also identified situations where these bounds are attainable and provided the corresponding optimal measurements; all this, in the framework of single- and multiple-copy estimation. In this section, we focus on the latter, in the asymptotic limit when a large number $N$ of copies is available for the experiment. Although the approach of the preceding sections can be carried out also in this case, asymptotic expansions become involved, with a few exceptions where a closed expression can be found for arbitrary $N$ [see, e.g., Eq. (26)]. Our aim here is to provide more straightforward means to obtain asymptotically optimal estimation protocols in general situations and compute the corresponding MSE. For this, we can resort on the well-known Cramér-Rao (CR) theory and its quantum extension, which we briefly discuss next, particularized to finite mixture estimation. A very powerful result, known as Holevo bound, will be also presented in the next section along with a simple example of use. A more detailed and comprehensive presentation, which includes a discussion on the relationship between this theory and the Bayesian approach of the preceding sections, can be found in [22].

In this framework, to which we will refer as 'pointwise,' one focusses on a fix point in parameter space, i.e., the unit simplex in our case, and restrict oneself to consider locally unbiased (LU) estimators: those for which $\langle \lambda_\chi \rangle_\lambda = \lambda$ in some open set, where, in the same spirit of previous notation, $\langle \cdot \rangle_\lambda$ indicates averaging over the conditional probability $p(\chi|\lambda)$ at the fixed point $\lambda$. We define the error matrix $\boldsymbol{\Delta}(\lambda)$ as

$$\boldsymbol{\Delta}(\lambda) = \langle (\lambda_\chi - \lambda)(\lambda_\chi - \lambda)^t \rangle_\lambda = \sum_\chi p(\chi|\lambda)(\lambda - \lambda_\chi)(\lambda - \lambda_\chi)^t. \quad (27)$$

It depends on the measurement and on the estimator, i.e., on the particular way one associates $\lambda_\chi$ to a given outcome $\chi$ of the chosen measurement. For the sake of simplicity in most of this section we will assume that the mixtures are identifiable. The problem of dealing with unidentifiable mixtures will be postponed to the last subsection (Sec. III C).

### A. The Cramér Rao bound

A first important result of the theory is the so-called CR bound [23,24]. It states that the error matrix of a LU estimator at $\lambda$ is lower bounded by the inverse of the Fisher information defined in Eq. (13) with $p_\lambda(\chi) = p(\chi|\lambda) = \text{tr} \, E_{\chi,\lambda} \rho_\lambda$ [note that in this theory the POVM may depend on the vector $\lambda$; see the comments after Eq. (30)], namely,

$$\mathbf{\Delta}(\lambda) \geqslant \boldsymbol{F}^{-1}(\lambda). \tag{28}$$

Assume now that the same measurement is performed on several independent copies; i.e., on the average state $\rho_\lambda^{\otimes N}$. Due to the additivity of the Fisher information, in this multiple-copy scenario one has

$$\mathbf{\Delta}(\lambda) \geqslant \frac{\boldsymbol{F}_1^{-1}(\lambda)}{N}, \tag{29}$$

where the subscript 1 refers to the one-copy model $\rho_\lambda$. This inequality expresses the fact that the MSE of the estimation scales with the inverse of the number of copies, and the accuracy by which we are able to estimate $\lambda$ with just a copy sets the scale. It is well-known that under some regularity conditions the maximum likelihood estimator achieves the CR bound asymptotically.

In spite of its fundamental character, the CR bound has the drawback that the bound it provides refers to a particular measurement, not necessarily optimal. To go around this difficulty, we invoke the Braunstein and Caves inequality, already discussed in the Sec. II A, and obtain

$$\mathbf{\Delta}(\lambda) \geqslant \frac{\boldsymbol{H}_1^{-1}(\lambda)}{N}. \tag{30}$$

Recall, however, that this bound is not always attainable but, when it is, the projectors onto the eigenspaces of the SLD $L_r(\lambda)$ define the optimal measurement. It is important to point out here that practical use of this approach requires a two-step measurement in order to saturate the bound. This is necessary because this optimal measurement, and thus the estimator, depend themselves on $\lambda$, which we do not know beforehand. To overcome this difficulty, one can take an asymptotically vanishing fraction of copies, say $\sqrt{N}$, and make an initial estimate of the weights $\lambda_\text{ini}$. Then, on the remaining copies one can perform the measurement that is optimal at $\lambda_\text{ini}$, i.e., project on the eigenspaces of $L_r(\lambda_\text{ini})$ (see Appendix C for an explicit example of this procedure). Thus, this two-step adaptive measurement, which is *independent* of $\lambda$, approaches the optimal one in the asymptotic limit at leading order in $1/N$, and one may write

$$\mathbf{\Delta} = \int d\lambda \, \pi(\lambda) \mathbf{\Delta}(\lambda) + o(N^{-1}). \tag{31}$$

This equation establishes a bridge between the asymptotic pointwise theory of this section and the Bayesian approach discussed in the first part of this paper. With all this in mind,

we conclude that for sufficiently smooth priors $\pi(\lambda)$ it holds that

$$\mathbf{\Delta} \geqslant \frac{1}{N} \int d\lambda \, \pi(\lambda) \boldsymbol{H}_1^{-1}(\lambda) + o(N^{-1}). \tag{32}$$

So far in this section we have overlooked the fact that not all the components of $\lambda$ are independent, as $\lambda$ must lie on the unit simplex. One could circumvent this by simply using the constraint $\sum_r \lambda_r = 1$ to write a particular component, say $\lambda_M$, in terms of the remaining $M-1$ as $\lambda_M = 1 - \sum_{r=1}^{M-1} \lambda_r$. This possibility, however, introduces a huge asymmetry in the calculation which may result in difficulties to invert the Fisher information matrix $\boldsymbol{H}_1$ and compute the bound (32). Note that inside the unit simplex the variations of $\lambda$ are constrained by $\Delta\lambda \cdot \boldsymbol{u} = 0$, where $\boldsymbol{u} = (1, 1, \dots, 1)$. A fully symmetric way of dealing with this issue is to project the information matrices $\boldsymbol{F}$ and $\boldsymbol{H}$ onto the orthogonal complement of span$\{\boldsymbol{u}\}$, which we call $S$. Thus, the CR bound, Eq. (29), takes the form [25]

$$\boldsymbol{P}_S \mathbf{\Delta}(\lambda) \boldsymbol{P}_S \geqslant \frac{1}{N} [\boldsymbol{P}_S \boldsymbol{F}_1(\lambda) \boldsymbol{P}_S]^{-1}, \tag{33}$$

and similarly for its quantum version in Eq. (30), where $\boldsymbol{P}_S$ stands for the projector on $S$ and the inverse, $[\cdot]^{-1}$, is restricted to the support of $\boldsymbol{P}_S$.

As an example, let us consider again the mixture of $M$ orthogonal states and compute the asymptotic expression of $E_N^\perp$, introduced in Eq. (26). Applying the definition of SLD in Eq. (16) to the one-copy family $\rho_\lambda$ it is straightforward to obtain that $L_r(\lambda) = P_r/\lambda_r$, where $P_r$ is the projector onto the support of $\rho_r$. Applying now the definition of the QFI, Eq. (15), to the same family we obtain $[H_1(\lambda)]_{rs} = \delta_{rs}/\lambda_r$. For brevity, we omit the arguments and write $\boldsymbol{H}_1'$ for the projection of $\boldsymbol{H}_1$ onto $S$, i.e., $\boldsymbol{H}_1' = \boldsymbol{P}_S \boldsymbol{H}_1 \boldsymbol{P}_S$, and similarly for other matrices. Let us start by computing det $\boldsymbol{H}_1'$ (here the zero eigenvalue corresponding to the kernel of the projection is, of course, removed from det). Since (i) the determinant of a $d \times d$ matrix is a homogeneous polynomial of degree $d$ in its matrix elements and (ii) the vector $\boldsymbol{u}$ has the same projection on each eigenspace of $\boldsymbol{H}_1$, it follows that (i) det $\boldsymbol{H}_1'$ must also be a homogeneous polynomial of degree $M-1$ in $1/\lambda_r$, i.e., in the eigenvalues of $\boldsymbol{H}_1$, and (ii) it must be a symmetric function of these eigenvalues. We also note that det $\boldsymbol{H}_1'$ must vanish if any two or more of these eigenvalues are set equal to zero, since in this case $S$ necessarily contains a null subspace of $\boldsymbol{H}_1$ [in doing so, the condition $\lambda_r \leqslant 1$ is temporarily lifted, which is legitimate, since the result we are after, Eq. (34) below, is an algebraic relation that holds for generic $\{\lambda_r\}$ regardless whether they are probabilities or not]. Hence,

$$\det \boldsymbol{H}_1' = \frac{1}{M} \sum_{\{r_i\}} \prod_{i=1}^{M-1} \frac{1}{\lambda_{r_i}} = \frac{1}{M} \prod_{i=1}^{M} \frac{1}{\lambda_i}, \tag{34}$$

where the sum extends to all subsets of $M-1$ indexes drawn from $1, 2, \dots, M$, and the prefactor $1/M$ can be easily computed by considering the particular case where all $\lambda_r$ are equal. Reasoning along the same lines, we conclude that

$$\det \boldsymbol{H}_1' \, \text{tr}(\boldsymbol{H}_1')^{-1} = \frac{2}{M} \sum_{\{r_j\}} \prod_{i=1}^{M-2} \frac{1}{\lambda_{r_i}}. \tag{35}$$

[Note that on the left-hand side of this last equation $\det \boldsymbol{H}'_1[(\boldsymbol{H}'_1)^{-1}]_{st}$ is the $(s,t)$ cofactor of $\boldsymbol{H}'_1$, i.e., the signed determinant of the matrix $\boldsymbol{H}'_1$ with row $s$ and column $t$ removed. It follows that $\det \boldsymbol{H}'_1 \operatorname{tr}(\boldsymbol{H}'_1)^{-1}$ is a homogeneous and symmetric polynomial in $1/\lambda_r$ of degree $M-2$ that vanishes if three or more eigenvalues of $\boldsymbol{H}_1$ are set equal to zero.] Combining Eq. (35) with Eq. (34), and after some algebra, we obtain

$$\operatorname{tr}(\boldsymbol{H}'_1)^{-1} = \left(\sum_r \lambda_r\right)^{-1} \sum_{r \neq s} \lambda_r \lambda_s = 1 - \sum_r \lambda_r^2. \quad (36)$$

The averaging over the flat prior can be easily performed with the help of Appendix B1, obtaining

$$\int d\boldsymbol{\lambda}\ \pi_{\text{flat}} \operatorname{tr}(\boldsymbol{H}'_1)^{-1} = \frac{M-1}{M+1}. \quad (37)$$

Taking into account (32) and $E_N^\perp = \operatorname{tr}\boldsymbol{\Delta} = \operatorname{tr}\boldsymbol{\Delta}'$ up to $o(N^{-1})$, we finally find that $E_N^\perp = (M-1)/[(M+1)N] + o(N^{-1})$, which indeed agrees with Eq. (26) for large $N$.

### B. Holevo bound

The quantum CR bound is a matrix inequality which is in general nonattainable [a few remarkable exceptions are those discussed in Sec. II A, in the paragraph after Eq. (17), and the example above]. However there is a related bound that one can expect to be saturated asymptotically: the Holevo bound. Indeed for qubit systems asymptotic attainability has been proved by Hayashi and Matsumoto in [26] and the general proof for finite dimensional systems follows from a recent paper by Kahn and Guţă [27]. We note that attainability here, as in the CR bound, is proven in a pointwise approach and hence makes implicit use of the two step adaptive measurement that we mentioned above. An important difference here is that at the second step the measurement attaining the Holevo bound will in general be a collective measurement that can not be implemented by local measurements on each copy.

Let us briefly introduce the Holevo bound for quantum finite mixture estimation (see also [22]). Let $G$ be a positive semidefinite matrix and

$$C_\lambda^N(G) = \min_{\substack{[(\{E_\chi\},\{\lambda_\chi\})] \\ \text{LU}}} \operatorname{tr} G\boldsymbol{\Delta}(\lambda), \quad (38)$$

where the minimization is over all pairs $(\{E_\chi\}, \{\lambda_\chi\})$ of measurements on $\rho_\lambda^{\otimes N}$ and estimators for which the latter is LU at $\boldsymbol{\lambda}$ (the unbiasedness of an estimator depends on the measurement through its outcome probability distribution). Equation (38) is relevant to the problem we are dealing with because its right hand side gives, e.g., the smallest MSE, $\operatorname{tr}\boldsymbol{\Delta}(\lambda)$, if $G = \mathbb{1}$, i.e., $C_\lambda^N(\mathbb{1})$ is the MSE of the optimal $N$-copy estimation scheme.

In Ref. [6] Holevo proved the following bound:

$$C_\lambda^1(G) \geqslant C_\lambda^H(G), \quad (39)$$

where

$$C_\lambda^H(G) = \min_{X \in \Xi_\lambda} \{\operatorname{tr} G\ \operatorname{Re} Z[X] + \operatorname{tr}|\sqrt{G}\ \operatorname{Im} Z[X]\sqrt{G}|\}. \quad (40)$$

In this expression $X = (X_1, X_2, \ldots, X_{M-1})$ are Hermitian matrices, one for each *independent* parameter (thus, for quan-

tum mixtures, we will choose $\lambda_M = 1 - \sum_r^{M-1} \lambda_r$), satisfying the following relations:

$$\operatorname{tr} \rho_\lambda X = 0, \quad (41)$$

$$\operatorname{tr} \partial_r \rho_\lambda X_s = \delta_{rs}, \quad 1 \leqslant r, s \leqslant M-1. \quad (42)$$

The minimization in Eq. (40) is over the set $\Xi_\lambda$ of all such $X$. Finally, $Z[X]$ is the matrix whose elements are given by

$$Z_{rs}[X] = \operatorname{tr} \rho_\lambda X_r X_s; \quad 1 \leqslant r, s \leqslant M-1. \quad (43)$$

Although the Holevo bound (39) is not in general attainable, it is attainable for the class of Gaussian models. The recent work [27] on asymptotic normality shows the asymptotic (local) equivalence between the many-copy states of finite-dimensional systems and a Gaussian model and thereby proves the asymptotic attainability of the Holevo bound for finite dimensional systems, i.e.,

$$\lim_{N \to \infty} N C_\lambda^N(G) = C_\lambda^H(G). \quad (44)$$

To relate the above with the Bayesian approach of the preceding sections, we need to average over $\pi(\boldsymbol{\lambda})$: asymptotically, we have $\operatorname{tr} G\boldsymbol{\Delta} = N^{-1} \int d\boldsymbol{\lambda}\pi(\boldsymbol{\lambda}) C_\lambda^H(G) + o(N^{-1})$. Thus, for instance, the MSE $E$ can be computed as

$$E = \operatorname{tr}\boldsymbol{\Delta} = \frac{1}{N} \int d\boldsymbol{\lambda}\pi(\boldsymbol{\lambda}) C_\lambda^H(\mathbb{1}) + o(N^{-1}). \quad (45)$$

As to whether or not this averaging is legitimate and the resulting bound on the averaged cost function is attainable, there exist very good heuristic arguments, as well as various examples [22], that this should be the case, but no rigorous proof. Thus, this last equation should be taken with a grain of salt.

To illustrate the use of the Holevo bound in finite mixture estimation, let us assume that $\rho_r$, $1 \leqslant r \leqslant 4$ are four pure qubit states whose Bloch vectors $\vec{n}_r$ form the vertices of a regular tetrahedron:

$$\vec{n}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \quad \vec{n}_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix},$$

$$\vec{n}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \quad \vec{n}_4 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \quad (46)$$

With this, the Bloch vector of the finite mixture is $\vec{r}_\lambda = \vec{n}_4 + \sum_{r=1}^3 \lambda_r(\vec{n}_r - \vec{n}_4)$. With full generality we may write $X_r = a_r + \vec{b}_r \cdot \vec{\sigma}$, where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the standard Pauli matrices. Conditions (41) and (42) are equivalent to $a_r = -\vec{b}_r \cdot \vec{r}_\lambda$ and $(\vec{n}_r - \vec{n}_4) \cdot \vec{b}_s = \delta_{rs}, 1 \leqslant r, s \leqslant 3$. This very last equation can be inverted (this will be always the case if the mixture is identifiable) and we obtain $\vec{b}_r = (3/4)\vec{n}_r$, $1 \leqslant r, s \leqslant 3$. With this,

$$X_r = \frac{1 - 4\lambda_r + 3\vec{n}_r \cdot \vec{\sigma}}{4}, \quad r = 1, 2, 3. \quad (47)$$

We see that Eqs. (41) and (42) determine $X$ uniquely and no minimization is required in Eq. (40). A straightforward calculation leads to

$$(\operatorname{Re} Z[X])_{rr} = \frac{(1 + 2\lambda_r)(1 - \lambda_r)}{2}; \quad (48)$$

$$(\text{Re } Z[X])_{rs} = -\frac{3 + (1 - 4\lambda_r)(1 - 4\lambda_s)}{16}, \quad r \neq s; \quad (49)$$

$$(\text{Im } Z[X])_{rs} = \frac{\sqrt{3}}{4}\epsilon_{rst}(\lambda_4 - \lambda_t); \quad (50)$$

where $\epsilon_{rst}$ is the (fully antisymmetric) Levi-Civita tensor in three dimensions. To compute the MSE we need the following:

$$\text{tr Re } Z[X] = \frac{1}{2}\left[3 + \sum_{r=1}^{3}\lambda_r(1 - 2\lambda_r)\right], \quad (51)$$

$$\text{tr }|\text{Im } Z[X]| = \frac{\sqrt{3}}{2}\left[\sum_{r=1}^{3}(\lambda_4 - \lambda_r)^2\right]^{1/2}, \quad (52)$$

and, averaging over $\pi_{\text{flat}}(\boldsymbol{\lambda})$, Eq. (45):

$$E = \frac{1}{N}\left(\frac{63}{40} + 0.43\right) + o(N^{-1}) = \frac{2.01}{N} + o(N^{-1}), \quad (53)$$

where the first (second) figure in the parenthesis comes from the real (imaginary) part of $Z[X]$ in Eq. (51) [Eq. (52)]. It is interesting to note that for this example, the quantum CR bound is not attainable. Indeed, one can check that the SLD $L_r(\boldsymbol{\lambda})$ is given by

$$L_r(\boldsymbol{\lambda}) = \frac{\vec{n}_r \cdot \vec{r}_\lambda}{|\vec{r}_\lambda|^2 - 1} + \left(\vec{n}_r - \frac{\vec{n}_r \cdot \vec{r}_\lambda}{|\vec{r}_\lambda|^2 - 1}\vec{r}_\lambda\right) \cdot \vec{\sigma}, \quad (54)$$

where $\vec{r}_\lambda = \sum_{r=1}^{4}\lambda_r\vec{n}_r$ is the Bloch vector of the averaged state $\rho_\lambda$ and $1 \leqslant r \leqslant 4$ (we now treat all components of $\boldsymbol{\lambda}$ as independent, in accordance with the approach developed in Sec. III A). One can immediately check that the commutator of the SLDs does not vanish, and the quantum CR bound is not saturated. Just for the sake of completeness, the quantum Fisher information matrix is given by

$$(H_1)_{rs} = \vec{n}_r \cdot \vec{n}_s + \frac{(\vec{n}_r \cdot \vec{r}_\lambda)(\vec{n}_s \cdot \vec{r}_\lambda)}{1 - |\vec{r}_\lambda|^2}, \quad (55)$$

for $1 \leqslant r \leqslant 4$. Projecting on $S$ with

$$\boldsymbol{P}_S = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (56)$$

and (pseudo)inverting, one obtains the relation

$$\boldsymbol{H}_1^{-1} = \text{Re } Z[X]. \quad (57)$$

After averaging, we observe from Eq. (53) that the quantum CR bound

$$E > \frac{63}{40N} + o(N^{-1}) \quad (58)$$

cannot be saturated.

## C. Unidentifiable mixtures in the asymptotic limit

In the preceding section, we were required to assume that estimation errors become vanishingly small as the number of copies increases. This assumption does not necessarily hold if mixtures are unidentifiable. In order to be able to apply the asymptotic techniques introduced above, we make a useful observation. If a quantum finite mixture is unidentifiable there

necessarily exists an orthogonal transformation $\boldsymbol{\lambda}' = \mathcal{O}\boldsymbol{\lambda}$ such that the states $\rho_\lambda$ depend solely on a reduced number of parameters $\{\xi_r = \lambda_r'\}_{r=1}^{m}$, with $m < M$, and are independent of the redundant parameters $\{\eta_r = \lambda_r'\}_{r=m+1}^{M}$. The error matrix $\boldsymbol{\Delta}$ of the original parameters $\boldsymbol{\lambda}$ is, of course, related to the error matrix $\boldsymbol{\Delta}'$ of the new ones, $\boldsymbol{\xi}, \boldsymbol{\eta}$, by the similarity transformation $\boldsymbol{\Delta}' = \mathcal{O}\boldsymbol{\Delta}\mathcal{O}^t$. Any measurement performed on the state $\rho_\lambda$ will only give information about the parameters $\boldsymbol{\xi}$, whereas the components of $\boldsymbol{\eta}$ have to be guessed independently of the measurement outcomes (e.g., by random choice). The optimal choice for $\boldsymbol{\eta}$ is, actually, $\langle\boldsymbol{\eta}\rangle$, and leads to an error that is, of course, independent of the number of copies. This means that in unidentifiable quantum mixture estimation there will always be an intrinsic error associated to the uncertainty in the redundant parameters $\boldsymbol{\eta}$, which remains constant regardless of the number of copies one is provided with. In the asymptotic limit, one can apply the bounds of the preceding sections to the block of $\boldsymbol{\Delta}'$ corresponding to the relevant components $\boldsymbol{\xi}$.

To illustrate this let us consider the unidentifiable qubit mixture defined by

$$\begin{aligned}\rho_\lambda &= \lambda_1|0\rangle\langle0| + \lambda_2|1\rangle\langle1| + \lambda_3|+\rangle\langle+| + \lambda_4|-\rangle\langle-| \\ &= \frac{1}{2}[\mathbb{1} + (\lambda_1 - \lambda_2)\sigma_z + (\lambda_3 - \lambda_4)\sigma_x], \quad (59)\end{aligned}$$

where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. If we perform the following rotation $\mathcal{O}$ in parameter space:

$$\begin{pmatrix}\xi_1 \\ \xi_2 \\ \eta_1 \\ \eta_2\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1\end{pmatrix}\begin{pmatrix}\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4\end{pmatrix}, \quad (60)$$

we have

$$\rho_\lambda = \frac{1}{2}(\mathbb{1} + \sqrt{2}\xi_1\sigma_z + \sqrt{2}\xi_2\sigma_x). \quad (61)$$

This shows that $\eta_1$ and $\eta_2$ are redundant parameters, and measurements will give us no information about them. For the simple model in Eq. (61), it is straightforward to obtain the Holevo bound. We first check that $X = (X_1, X_2)^t$, with

$$X_1 = \frac{\sigma_z}{\sqrt{2}} - \xi_1\mathbb{1}, \quad X_2 = \frac{\sigma_x}{\sqrt{2}} - \xi_2\mathbb{1}, \quad (62)$$

is the solution to conditions (41) and (42) that minimizes Eq. (40). It follows that

$$\text{Re } Z[X] = \begin{pmatrix}\frac{1}{2} - \xi_1^2 & -\xi_1\xi_2 \\ -\xi_1\xi_2 & \frac{1}{2} - \xi_2^2\end{pmatrix}, \quad \text{Im } Z[X] = 0, \quad (63)$$

and Eq. (40) gives

$$C_\xi^H(\mathbb{1}) = 1 - \xi_1^2 - \xi_2^2. \quad (64)$$

In the limit $N \to \infty$, we can compute the error coming from the estimation of $\boldsymbol{\xi}$ through Eq. (45), i.e.,

$$\begin{aligned}\sum_{r=1}^{2}\Delta_{rr}^{(\xi)} &= \frac{1}{N}\int d\boldsymbol{\lambda}\pi_{\text{flat}}(\boldsymbol{\lambda})C_{\xi(\lambda)}^H(\mathbb{1}) + o(N^{-1}) \\ &= \frac{9}{10N} + o(N^{-1}), \quad (65)\end{aligned}$$

which is asymptotically vanishing. As for the estimation of $\eta_1$ and $\eta_2$, we make the optimal guess

$$\langle \eta_r \rangle = \int d\lambda \pi_{\text{flat}}(\lambda)\eta_r(\lambda) = \frac{1}{2\sqrt{2}}, \tag{66}$$

thus

$$\sum_{r=1}^{2} \Delta_{rr}^{(\eta)} = \int d\lambda \pi_{\text{flat}}(\lambda) \sum_{r=1}^{2} [\eta_r(\lambda) - \langle \eta_r \rangle]^2 = \frac{1}{20} \tag{67}$$

[according to the notation introduced in the paragraph below Eq. (10), this quantity could also be denoted by $\sum_{r=1}^{2} \Lambda_{rr}^{(\eta)}$].

Putting all pieces together, the estimation error is

$$E = \text{tr}\boldsymbol{\Delta} = \text{tr}\boldsymbol{\Delta}' = \sum_{r=1}^{2} \Delta_{rr}^{(\eta)} + \sum_{r=1}^{2} \Delta_{rr}^{(\xi)}$$
$$= \frac{1}{20} + \frac{9}{10N}. \tag{68}$$

In conclusion, this explicit example shows that unidentifiable mixtures will lead to a non-vanishing estimation error even in the asymptotic limit.

## IV. ESTIMATION OF TWO-COMPONENT MIXTURES

In this section we dwell on the simplest quantum mixture scenario, where the average state $\rho_\lambda$ belongs to the one-simplex

$$\rho_\lambda = \lambda\rho_1 + (1 - \lambda)\rho_2, \quad 0 \leqslant \lambda \leqslant 1 \tag{69}$$

(hence $\lambda_1 = \lambda$, $\lambda_2 = 1 - \lambda$). Although the error matrix is $2 \times 2$, only one of its entries, say $\Delta_{11}$, contains independent information about the accuracy in the estimation of the mixture (69). Therefore, in the following we simply drop the remaining three entries and write $\Delta$ [and likewise for the Fisher information matrix $F(\lambda)$, the quantum Fisher $H(\lambda)$, etc., to which we will refer as $F(\lambda)$, $H(\lambda)$, etc.]. Since there is only an independent parameter, we may also drop the vector notation and write $\lambda$ instead of $\boldsymbol{\lambda}$.

### A. Single-shot estimation

The single-copy version of this problem was considered recently in [18], though the optimal measurements and minimal estimation error were only determined when $\rho_1$ and $\rho_2$ are qubit and/or pure states. Our results in Sec. II show that for the two-component mixture in Eq. (69), the attainability conditions are fulfilled for *any* $\rho_1$ and $\rho_2$, and the optimal measurements, along with their minimal estimation error, can always be determined in both single- and multiple-copy scenarios. In particular, it follows from our results that the optimal protocol consists of a projective measurement, where the projectors are those onto the eigenspaces of the SLD $L(\bar{\lambda})$. Our results in the present paper thus provide answers to various open questions posed in [18].

Let us focus first on the single-copy estimation. By choosing the optimal estimator (6), the MSE is given by Eq. (19). For the mixture (69) this equation can be cast as

$$\Delta = \Lambda - 2\Lambda^2 \sum_{nm} \frac{|\langle \phi_m | \rho_1 - \rho_2 | \phi_n \rangle|^2}{\nu_m + \nu_n}. \tag{70}$$

[In the case under consideration here, $\tilde{\Lambda}_{\alpha 1} = \langle \lambda \lambda_\alpha \rangle - \langle \lambda \rangle \langle \lambda_\alpha \rangle$. Thus, $\tilde{\Lambda}_{11} = \langle \lambda^2 \rangle - \langle \lambda \rangle^2 \equiv \Lambda$, and $\tilde{\Lambda}_{21} = \langle \lambda(1 - \lambda) \rangle - \langle \lambda \rangle \langle 1 - \lambda \rangle = -\Lambda$.] The bound (70) is attained with the measurement characterized by the eigenprojectors of the SLD [see Eq. (17)]

$$L(\bar{\lambda}) = 2\Lambda \sum_{nm} \frac{\langle \phi_m | \rho_1 - \rho_2 | \phi_n \rangle}{\nu_m + \nu_n} |\phi_n\rangle\langle\phi_m|, \tag{71}$$

where we recall that $\{|\phi_n\rangle\}$ ($\nu_n$) are the eigenvectors (eigenvalues) of $\langle \rho_\lambda \rangle$. One can readily check by explicitly solving Eq. (16) that for a uniform prior $\pi_{\text{flat}}(\lambda) = 1$, and for pure (or for qubit states with the same purity) $\rho_1$ and $\rho_2$ we have $L(\bar{\lambda} = 1/2) \propto (\rho_1 - \rho_2)$, in agreement with [18]. Accordingly, in this situation $\Delta = [2 + \text{tr}(\rho_1\rho_2)]/36$.

### B. Multiple-copy estimation and the asymptotic limit

Although a straightforward exercise, computing $\Delta$ for $N > 1$ copies of $\rho_\lambda$ is a tedious task even for two-component mixtures. In most cases, the resulting expressions cannot be written in closed form for arbitrary $N$ and are thus not very revealing. So, rather than attempting to present a general case, we have selected a particular example, which we will later use to illustrate the connection between the Bayesian and the asymptotic pointwise approaches.

Assume $\rho_1$ and $\rho_2$ are commuting non-orthogonal qubit states. Let us further assume that $\rho_2$ is pure and that the prior is flat. Then, we can choose basis so that

$$\rho_\lambda = \lambda \begin{pmatrix} 1 - \epsilon & 0 \\ 0 & \epsilon \end{pmatrix} + (1 - \lambda) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{72}$$

Proceeding as in Sec. II B, the $N$-copy state $\rho_\lambda^{\otimes N}$ can be cast in the form (3) with

$$c_k(\lambda) = \binom{N}{k}(\lambda\epsilon)^k(1 - \lambda\epsilon)^{N-k}, \quad 0 \leqslant k \leqslant N, \tag{73}$$

and $\rho_k = \mathcal{S}[|1\rangle\langle 1|^{\otimes k} \otimes |0\rangle\langle 0|^{\otimes(N-k)}]$. Hence, using Eq. (19) we have that the minimum error is given by [recall that we are assuming the flat prior $\pi_{\text{flat}}(\lambda) = 1$]

$$\Delta^{\text{comm}} = \frac{1}{12} - \frac{1}{2} \sum_{n,m} \frac{|\langle \phi_m | \sum_k B_k \rho_k |\phi_n\rangle|^2}{\nu_m + \nu_n}, \tag{74}$$

where

$$B_k \equiv 2\tilde{\Lambda}_k = 2[\langle \lambda c_k(\lambda) \rangle - \langle \lambda \rangle \langle c_k(\lambda) \rangle]$$
$$= \binom{N}{k} \int_0^1 d\lambda(2\lambda - 1)(\epsilon\lambda)^k(1 - \epsilon\lambda)^{N-k}, \tag{75}$$

and where now

$$\{|\phi_n\rangle\} = \text{perms}\{|1\rangle^{\otimes k} \otimes |0\rangle^{\otimes(N-k)}\}_{k=0}^{N} \tag{76}$$

are the $2^N$ eigenvectors of $\langle \rho_\lambda^{\otimes N} \rangle$ (perms$\{\cdot\}$ stands for the set of distinct permutations of the set $\{\cdot\}$). Defining

$$A_k \equiv \binom{N}{k} \int_0^1 d\lambda(\epsilon\lambda)^k(1 - \epsilon\lambda)^{N-k}, \tag{77}$$

the eigenvalues of $\langle \rho_\lambda^{\otimes N} \rangle$ are $\nu_k = A_k / \binom{N}{k}$, and have multiplicity $\binom{N}{k}$. Therefore,

$$\Delta^{\mathrm{comm}} = \frac{1}{12} - \frac{1}{4} \sum_k \frac{B_k^2}{A_k}. \tag{78}$$

As shown in Appendix B3, the terms of the sum above can be written as ratios of regularized incomplete beta functions thus providing a more compact expression for the error. However, we can only give a closed form for $\Delta$ in the asymptotic limit of very large number of copies. This requires evaluating the sum in Eq. (78) up to order $1/N$:

$$S \equiv \sum_k \frac{B_k^2}{A_k} = \frac{1}{3} + \frac{1}{N}\left(\frac{4}{3} - \frac{2}{\epsilon}\right) + o(N^{-1}) \tag{79}$$

(details of this evaluation are also given in Appendix B). Plugging this expression into Eq. (78) we obtain

$$\Delta^{\mathrm{comm}} = \frac{1}{N}\left(\frac{1}{2\epsilon} - \frac{1}{3}\right) + o(N^{-1}). \tag{80}$$

With the asymptotic techniques introduced in Sec. III A the previous evaluation can be simplified a great deal. Moreover, these techniques enable us to give closed-form expressions of $\Delta$ for rather more general two-component mixtures. As already mentioned, the attainability of the CR bound is guaranteed for these (one-parameter) mixtures and its application is particularly simple. From our discussion in Sec. III A, Eq. (32), we can write

$$\Delta = \frac{1}{N} \int d\lambda \, \pi(\lambda) H_1^{-1}(\lambda) + o(1/N), \tag{81}$$

where we recall that $H_1$ is the QFI of the one-copy model (69). As it can be simply read off from Eq. (70),

$$H_1(\lambda) = 2 \sum_{nm} \frac{|\langle \phi_m | \rho_1 - \rho_2 | \phi_n \rangle|^2}{\nu_m + \nu_n}. \tag{82}$$

Note, however, that $\{|\phi_n\rangle\}$ $(\nu_n)$ are now the eigenvectors (eigenvalues) of $\rho_\lambda$, rather than of $\langle \rho_\lambda \rangle$, and the QFI is thus a function of $\lambda$. In the Bloch representation we can write

$$\rho_r = \frac{1}{2}(\mathbb{I} + \vec{r}_r \cdot \vec{\sigma}), \quad r = 1, 2, \tag{83}$$

which holds when $\rho_1$ and $\rho_2$ are both qubit states, but also when they are pure states in arbitrary dimensions. Since in these cases the two density matrices can be taken to be real, it suffices to consider $\vec{\sigma} = (\sigma_x, \sigma_z)$. The eigenvalues and eigenvectors of $\rho_\lambda$ can be written as

$$|\phi_\pm\rangle\langle\phi_\pm| = \frac{1}{2}\left(\mathbb{I} \pm \frac{\vec{r}_\lambda \cdot \vec{\sigma}}{r_\lambda}\right), \quad \nu_\pm = \frac{1 \pm r_\lambda}{2}, \tag{84}$$

where, as in previous examples, $\vec{r}_\lambda = \lambda \vec{r}_1 + (1-\lambda)\vec{r}_2$ is the Bloch vector of $\rho_\lambda$, and we have defined $r_\lambda = |\vec{r}_\lambda|$. After some algebra one finds

$$H_1(\lambda) = |\vec{r}_1 - \vec{r}_2|^2 + \frac{[(\vec{r}_1 - \vec{r}_2) \cdot \vec{r}_\lambda]^2}{1 - r_\lambda^2}. \tag{85}$$

For pure states, $\rho_1 = |\varphi_1\rangle\langle\varphi_1|$ and $\rho_2 = |\varphi_2\rangle\langle\varphi_2|$ (i.e., $r_1 = r_2 = 1$) one can further simplify this expression and

write

$$H_1^{\mathrm{pure}}(\lambda) = \frac{1 - |\langle\varphi_1|\varphi_2\rangle|^2}{\lambda(1-\lambda)}. \tag{86}$$

If the prior is assumed to be flat, $\pi_{\mathrm{flat}}(\lambda) = 1$, a trivial integration leads to

$$\Delta^{\mathrm{pure}} = \frac{1}{6N(1 - |\langle\varphi_1|\varphi_2\rangle|^2)} + o(1/N). \tag{87}$$

If $\rho_1$ and $\rho_2$ are not pure, the Bloch representation (83) holds only for qubit states. Assuming the flat prior, after a lengthy calculation one finds (to leading order in $1/N$)

$$\begin{aligned}\Delta^{\mathrm{qubit}} &= \frac{1}{6N} \frac{6 - |\vec{r}_1 + \vec{r}_2|^2 - r_1^2 - r_2^2}{|\vec{r}_1 - \vec{r}_2|^2 - r_1^2 r_2^2 + (\vec{r}_1 \cdot \vec{r}_2)^2} \\ &= \frac{1}{6N} \frac{3 - \mathrm{tr}\,\rho_1^2 - \mathrm{tr}\,\rho_2^2 - \mathrm{tr}\,\rho_1\rho_2}{\mathrm{tr}\,\rho_1^2 + \mathrm{tr}\,\rho_2^2 - \mathrm{tr}\,\rho_1^2\mathrm{tr}\,\rho_2^2 - (2 - \mathrm{tr}\,\rho_1\rho_2)\mathrm{tr}\,\rho_1\rho_2}.\end{aligned} \tag{88}$$

Recall that for the cases at hand there exist adaptive measurement that attain the above bound. The reader is referred to Appendix C for a specific illustration of this general result.

Before ending this section, we come back to the two commuting states example in Eq. (72), for which the estimation error, Eq. (80), was worked out entirely in the Bayesian framework and the limit $N$ was taken afterwards. The same estimation error can be obtained applying the pointwise CR result (88). It is straightforward to check that this much less costly procedure leads to the same result (80), as it should. Recall, however, that it leads to sensible results only if the number $N$ of copies is exceedingly large, whereas the Bayesian approach works for any $N$.

## V. CONCLUSIONS

Quantum ensembles embody what in classical statistics is known as finite mixtures, and can thus be viewed as their quantum counterpart. More precisely, we have a quantum finite mixture whenever a signal can be characterized by a density matrix that is the average of a set of *known* states (pure or mixed), as is often the case in quantum communication. In these situations, one wishes to find the probability law that best describes the signal, or in other words, the weights that define the quantum ensemble. This has been the subject of the present paper, where we have relied on quantum estimation theory, but also broadened the field by proposing certain applications and tools.

The topics addressed in this paper include: the precise definition of quantum finite mixtures, as an extension of finite mixtures to the quantum domain; optimal estimation (of their weights) when a given number of copies of the average state is available for measurement; optimal estimation in the asymptotic regime of large number of copies; and characterization of the (un)identifiability of quantum mixtures. For each of these topics we have answered the relevant questions and provided useful results, of which we also give some examples of application.

Going into more detail, we have approached optimality from both the Bayesian and the 'pointwise' points of view. In the former, one minimizes an averaged cost function, which we have chosen to be the covariance-type error matrix of the estimation, over a joint probability involving the measurement

outcomes as well as the prior knowledge of the weights. Our key result is $\boldsymbol{\Delta} = \boldsymbol{\Lambda} - \boldsymbol{F}$ [see Eq. (12)]. It states that the error matrix is the intrinsic uncertainty of the weights minus the Fisher information matrix, which quantifies the information gained in the measurement process. This *exact* relation, valid for any number of copies, is linear in the Fisher information matrix, in contrast to the Cramér-Rao bound, where the error is lower bounded by the inverse of the Fisher information matrix. From our relation one obtains a measurement independent lower bound on the error matrix in terms of the quantum Fisher information. In those cases where the Braunstein-Caves inequality (which states that the Fisher information matrix is upper bounded by the quantum Fisher information) is saturated our bound is attainable *for any number of copies*. When this holds (e.g., two-component mixtures), we give the optimal measurement protocol, which turns out to be of von Neumann type.

As to the pointwise approach to quantum mixture estimation, we have briefly introduced the quantum Cramér-Rao and the Holevo bounds in the specific context at hand. We have next applied these tools to obtain lower bounds for the error matrix of the weights when the number of copies of the average state is asymptotically large. In those situations, the Bayesian approach becomes rather involved and it is advisable to switch to the tools under discussion. Although the quantum Cramér-Rao and the Holevo bounds can be applied to unidentifiable mixtures, its use requires some technicalities that we have commented upon and illustrated with an example. As one would expect, the accuracy of the weight estimation for such mixtures does not vanish even if an infinite number of copies were available. A discussion on the relationship between the Bayesian and pointwise approaches has been also given, as well as an example illustrating that the two approaches give consistent results.

Among the examples one can find in this paper, we would like to highlight that of a mixture of a number of orthogonal states, which is relevant in the context of channel estimation. For this problem, and assuming a flat prior distribution of weights we have been able to write the minimal square error in a closed form, valid for any number of orthogonal states and any number of copies of the average state.

This paper is mostly devoted to the formalism and general results concerning quantum finite mixtures and the estimation of their weights. The examples are chosen for the sake of illustration, rather than for their practical relevance. As mentioned in the introduction, real applications of our work are, e.g., the characterization of signals in relevant quantum communication problems and the estimation of probabilities with which various errors occur in a given channel. We have shown that in some instances the bounds we give are attainable by local two-step adaptive measurements. It remains an open question to establish whether or not collective measurements are necessary in the general case. Future extensions of our work also include the estimation of mixtures of continuous variable systems.

## ACKNOWLEDGMENTS

## APPENDIX A: THE $\sigma_\lambda$ ARE PHYSICAL STATES

It is clear from its definition in Eq. (11) that $\operatorname{tr} \sigma_\lambda = 1$. So $\sigma_\lambda$ is a proper density matrix if $\sigma_\lambda \geqslant 0$. To prove this inequality, we take any state $|\psi\rangle$ and define $p_\lambda^\psi = \langle\psi|\rho_\lambda|\psi\rangle \geqslant 0$. Recalling Eqs. (3) and (11), we see that the relation $\langle\psi|\sigma_\lambda|\psi\rangle \geqslant 0$ is equivalent to

$$\langle p_\lambda^\psi\rangle \left[1 + \sum_r (\lambda_r - \langle\lambda_r\rangle)\left(\frac{\langle\lambda_r p_\lambda^\psi\rangle}{\langle p_\lambda^\psi\rangle} - \langle\lambda_r\rangle\right)\right] \geqslant 0. \quad \text{(A1)}$$

[Note that $\langle\lambda_r p_\lambda^\psi\rangle/\langle p_\lambda^\psi\rangle \geqslant 0$ and $\sum_r \langle\lambda_r p_\lambda^\psi\rangle/\langle p_\lambda^\psi\rangle = 1$.] But Eq. (A1) immediately follows from the inequality

$$(\boldsymbol{x} - \boldsymbol{z}) \cdot (\boldsymbol{y} - \boldsymbol{z}) \geqslant -1/2, \quad \text{(A2)}$$

where $\boldsymbol{x} = \{x_r\}$, $\boldsymbol{y} = \{y_r\}$ and $\boldsymbol{z} = \{z_r\}$ stand for any three probability vectors.

For the sake of completeness, we also prove Eq. (A2). We just need to notice that $(\boldsymbol{x} - \boldsymbol{z}) \cdot (\boldsymbol{y} - \boldsymbol{z})$ as a function of $\boldsymbol{z}$ has a minimum at $\boldsymbol{z}_0 = (\boldsymbol{x} + \boldsymbol{y})/2$. For any $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ we can thus write

$$(\boldsymbol{x} - \boldsymbol{z}) \cdot (\boldsymbol{y} - \boldsymbol{z}) \geqslant (\boldsymbol{x} - \boldsymbol{z}_0) \cdot (\boldsymbol{y} - \boldsymbol{z}_0)$$
$$= -\frac{|\boldsymbol{x} - \boldsymbol{y}|^2}{4} \geqslant -\frac{1}{2}, \quad \text{(A3)}$$

which is the inequality (A2).

## APPENDIX B: USEFUL FORMULAE

### 1. Averages with a flat prior

Recall our notation: $d\boldsymbol{\lambda} = \delta(\sum_r \lambda_r - 1)\prod_r d\lambda_r$. Then, one can prove the following useful result:

$$\int d\boldsymbol{\lambda} \prod_r \lambda_r^{k_r} = \frac{k_1!\cdots k_M!}{(M-1+\sum_r k_r)!}, \quad \text{(B1)}$$

where the integration is restricted to positive values of $\lambda_r$, $r = 1, \ldots, M$. Although we use this integral for $k_r$ being positive integers, the result can be generalized to complex $k_r$ by simply replacing the factorials by Euler Gamma functions: $k! \rightarrow \Gamma(k+1)$.

In particular, Eq. (B1) and the normalization condition $\int d\boldsymbol{\lambda}\pi(\boldsymbol{\lambda}) = 1$ imply that the flat distribution is given by

$$\pi_{\text{flat}}(\boldsymbol{\lambda}) = (M-1)!. \quad \text{(B2)}$$

### 2. Sums

Some results in Sec. II B require computing sums of the form $\sum_{\boldsymbol{k},r} f(k_r)$, where $k_r$ are the components of the vector $\boldsymbol{k}$. They are positive integers that add up to $N$, and the sum extends over all the

$$\binom{M+N-1}{M-1} \quad \text{(B3)}$$

such vectors. To compute this sums, we first note that

$$\sum_{\boldsymbol{k},r} f(k_r) = \sum_{r=1}^{M} \left( \sum_{\boldsymbol{k}} f(k_r) \right) = M \sum_{\boldsymbol{k}} f(k_1), \quad \text{(B4)}$$

where we have used that the sum in parenthesis is independent of $r$. This is so because the set of all vectors $\boldsymbol{k}$ is invariant under $k_r \to k_{\sigma(r)}$, where $\sigma$ is any permutation of the symmetric group $S_M$, and thus $\sum_{\boldsymbol{k}} f(k_r) = \sum_{\boldsymbol{k}} f(k_{\sigma(r)})$. We next note that any vector $\boldsymbol{k}$ whose first component is fixed to be $k_1$ gives the same contribution, $f(k_1)$, to the last sum in (B4). The number of such vectors follows from Eq. (B3) by simply making the substitutions $M \to M - 1$ and $N \to N - k_1$. Hence,

$$\sum_{\boldsymbol{k},r} f(k_r) = M \sum_{k_1=0}^{N} \binom{M + N - k_1 - 2}{M - 2} f(k_1). \quad \text{(B5)}$$

For the particular case we need in Sec. II B, $f(x) = x^2$ and the corresponding sum gives

$$\sum_{\boldsymbol{k},r} k_r^2 = \frac{2N + M - 1}{M + 1} \frac{(M + N - 1)!}{(M - 1)!(N - 1)!}. \quad \text{(B6)}$$

### 3. Evaluation of the sum (79)

Recalling the definitions (75) and (77), and after some algebra, we have

$$S = \frac{1}{N+1} \frac{4}{\epsilon^3} \sum_{k=0}^{N} \left( \frac{k+1}{N+2} \right)^2 \frac{I_\epsilon^2(k+2, \bar{k}+1)}{I_\epsilon(k+1, \bar{k}+1)} - 1$$

$$\equiv \frac{R(\epsilon)}{\epsilon^3} - 1, \quad \text{(B7)}$$

where $\bar{k} \equiv N - k$ and $I_x(a, b)$ is the regularized incomplete beta function,

$$I_x(a, b) = \frac{B_x(a, b)}{B(a, b)} = \frac{1}{B(a, b)} \int_0^x dt \ t^{a-1}(1-t)^{b-1}. \quad \text{(B8)}$$

To obtain Eq. (B7) we have also used that

$$\sum_{k=0}^{N} I_\epsilon(k+1, \bar{k}+1) = (N+1)\epsilon \quad \text{(B9)}$$

and

$$\sum_{k=0}^{N} \frac{k+1}{N+2} I_\epsilon(k+1, \bar{k}+1) = (N+1)\frac{\epsilon^2}{2}, \quad \text{(B10)}$$

which both follow immediately from the definition in Eq. (B8). Recall also that $B_1(a, b) \equiv B(a, b)$, where $B(a, b)$ is the standard (complete) beta function, $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$. According to the Euler-MacLaurin formula, the sum in Eq. (B7) can be approximated by an integral which, after differentiating with respect to $\epsilon$, can be cast as

$$R'(\epsilon) = \frac{4N}{N+1} \int_0^1 dx \left( \frac{Nx+1}{N+2} \right)^2 \frac{I_\epsilon(Nx+2, N\bar{x}+1)}{I_\epsilon(Nx+1, N\bar{x}+1)}$$

$$\times \left\{ \frac{2(N+2)\epsilon}{Nx+1} - \frac{I_\epsilon(Nx+2, N\bar{x}+1)}{I_\epsilon(Nx+1, N\bar{x}+1)} \right\}$$

$$\times \frac{\epsilon^{Nx}(1-\epsilon)^{N\bar{x}}}{B(Nx+1, N\bar{x}+1)}, \quad \text{(B11)}$$

where $\bar{x} \equiv 1 - x$. The last factor peaks at $x = \epsilon$ as $N$ becomes large and can be replaced by the Gaussian

$$\frac{N+1}{N} \sqrt{\frac{N}{2\pi\epsilon(1-\epsilon)}} \exp\left\{ -N\frac{(x-\epsilon)^2}{2\epsilon(1-\epsilon)} \right\}.$$

Since we are interested only in terms that vanish asymptotically as $N^{-1}$, we can drop those that vanish exponentially, and approximate Eq. (B11) by

$$R'(\epsilon) = 4 \int_{-\infty}^{\infty} dx \left( \frac{Nx+1}{N+2} \right)^2 \left\{ \frac{2(N+2)\epsilon}{Nx+1} - 1 \right\}$$

$$\times \sqrt{\frac{N}{2\pi\epsilon(1-\epsilon)}} \exp\left\{ -N\frac{(x-\epsilon)^2}{2\epsilon(1-\epsilon)} \right\}. \quad \text{(B12)}$$

For the same reason, we can expand the first line in Eq. (B12) up to first order in $u^2 \equiv (x - \epsilon)^2$ and write

$$R'(\epsilon) = 4 \int_{-\infty}^{\infty} du \ (\epsilon^2 - u^2)$$

$$\times \sqrt{\frac{N}{2\pi\epsilon(1-\epsilon)}} \exp\left\{ -N\frac{u^2}{2\epsilon(1-\epsilon)} \right\}. \quad \text{(B13)}$$

The remaining integral gives

$$R'(\epsilon) = 4\epsilon^2 - \frac{4}{N}\epsilon(1-\epsilon). \quad \text{(B14)}$$

Hence,

$$R(\epsilon) = R(0) + \int_0^\epsilon ds \, R'(s)$$

$$= \frac{4}{3}\epsilon^3 - \frac{2}{3N}\epsilon^2(3 - 2\epsilon), \quad \text{(B15)}$$

from which the final result follows.

### APPENDIX C: TWO-STEP ADAPTIVE MEASUREMENT IN THE ASYMPTOTIC LIMIT

In this appendix we give an explicit example of the two-step adaptive measurement protocol that attains the Cramér-Rao bound asymptotically [see Sec. III A, the paragraph after Eq. (30)]. To ease the calculation we choose the simplest instance: that of a mixture of two pure states, $\rho_\lambda = \lambda\rho_1 + (1 - \lambda)\rho_2$. This mixture has been already considered in Sec. IV, in the paragraph after Eq. (85). Here we stick to the same notation. If $\rho_r$ ($r = 1, 2$) are pure, without loss of generality they can be chosen to be

$$\rho_r = \tfrac{1}{2}[\mathbb{1} + \sigma_z \cos\theta + (-1)^{r+1}\sigma_x \sin\theta] \quad \text{(C1)}$$

(as if they were qubit states on the equator of the Bloch sphere), where $\cos\theta = \sqrt{\text{tr}\,\rho_1\rho_2} = |\langle\varphi_1|\varphi_2\rangle|$ is the overlap.

Let us assume that we are given $N$ copies of the state $\rho_\lambda$. On a first stage of the protocol, we take $\sqrt{N}$ of these copies and perform on each of them a same measurement, with the aim of obtaining an initial, rough estimate of $\lambda$, which we denote by $\lambda_{\text{ini}}$. Since these measurements use uncorrelated copies and are themselves independent, we expect to benefit from the well understood statistical improvement that results from averaging over the $\sqrt{N}$ samples. Thus, we can assume that, in average,

$$(\lambda - \lambda_{\text{ini}})^2 \sim \alpha/\sqrt{N}, \quad \text{(C2)}$$

where $\alpha$ is some constant whose value depends on the precise measurement that we perform.

On a second stage, we refine the rough estimation obtained in the preceding stage by performing a (nearly optimal) measurement on the remaining $N - \sqrt{N}$ copies. As discussed in Sec. III A (see also Sec. II A), the optimal measurement is described by the set of projector, $\{P_\chi(\lambda)\}$ (it is a von Neumann measurement), onto the different eigenspaces of the SLD, $L(\lambda)$, of our model evaluated at $\lambda$. For our example, one can readily find that

$$L(\lambda) = \frac{(1 - 2\lambda)(\mathbb{1} + \sigma_z \cos\theta) + \sigma_x \sin\theta}{2\lambda(1 - \lambda)}. \qquad \text{(C3)}$$

However, since we do not know the true value of $\lambda$, we choose the measurement to be given by $\{P_\chi(\lambda_{\text{ini}})\}$, and hope this change will not affect optimality. Let us check that this is indeed the case. To this end, we diagonalize Eq. (C3), obtain $\{P_\chi(\lambda_{\text{ini}})\}$ and, in turn, compute its Fisher information defined

in Eq. (13). We obtain

$$F_1 = \frac{\sin^2\theta}{\lambda(1 - \lambda) + (\lambda_{\text{ini}} - \lambda)^2 \cos^2\theta}. \qquad \text{(C4)}$$

(recall that the subscript 1 refers to one copy). Thus, the error of performing this measurement on the $N - \sqrt{N}$ copies is

$$\Delta(\lambda) = \frac{1}{N - \sqrt{N}} \left\{ \frac{\lambda(1 - \lambda)}{\sin^2\theta} + (\lambda_{\text{ini}} - \lambda)^2 \cot^2\theta \right\}. \qquad \text{(C5)}$$

For sufficiently large $N$ (so that $\sqrt{N}$ itself is also very large), Eq. (C2) holds in average, and

$$\begin{aligned}
\Delta(\lambda) &= \frac{\lambda(1 - \lambda)\,\text{cosec}^2\theta}{N(1 - N^{-1/2})} + \frac{\alpha \cot^2\theta}{N^{3/2}(1 - N^{-1/2})} \\
&= \frac{\lambda(1 - \lambda)\,\text{cosec}^2\theta}{N} + O(N^{-3/2}),
\end{aligned} \qquad \text{(C6)}$$

thus attaining the optimal bound, as can be read off from Eq. (86).

---

[1] G. McLachlan and D. Peel, *Finite Mixture Models* (Wiley Interscience, New York, 2000).

[2] K. Pearson, Philos. Trans. R. Soc. London A **185**, 71 (1894).

[3] E. Schrödinger, Proc. Camb. Phil. Soc. **32**, 446 (1936); L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).

[4] O. Oreshkov and J. Calsamiglia, Phys. Rev. A **79**, 032336 (2009).

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

[6] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[7] M. Hayashi, *Quantum Information: An Introduction* (Springer, New York, 2006).

[8] M. Paris and J. Rehacek, *Quantum State Estimation* (Springer, New York, 2004).

[9] M. Hotta, T. Karasawa, and M. Ozawa, Phys. Rev. A **72**, 052334 (2005); A. Fujiwara, *ibid.* **63**, 042304 (2001); A. Monras and M. G. A. Paris, Phys. Rev. Lett. **98**, 160401 (2007).

[10] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[11] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Phys. Rev. Lett. **79**, 3865 (1997).

[12] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).

[13] K. Goda, O. Miyakawa, E. E. Mikhailov, S. Saraf, R. Adhikari, K. McKenzie, R. Ward, S. Vass, A. J. Weinstein, and N. Mavalvala, Nat. Phys. **4**, 472 (2008).

[14] R. Jozsa, D. S. Abrams, J. P. Dowling, and C. P. Williams, Phys. Rev. Lett. **85**, 2010 (2000).

[15] M. de Burgh and S. D. Bartlett, Phys. Rev. A **72**, 042301 (2005).

[16] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Proc. R. Soc. London A **454**, 339 (1998).

[17] S. L. Braunstein, C. A. Fuchs, and H. J. Kimble, J. Mod. Opt. **47**, 267 (2000); K. Hammerer, M. M. Wolf, E. S. Polzik, and J. I. Cirac, Phys. Rev. Lett. **94**, 150503 (2005); G. Adesso and G. Chiribella, *ibid.* **100**, 170503 (2008); M. Owari, M. B. Plenio, E. S. Polzik, A. Serafini, and M. M. Wolf, New J. Phys. **10**, 113014 (2008); J. Calsamiglia, M. Aspachs, R. Munoz-Tapia, and E. Bagan, Phys. Rev. A **79**, 050301(R) (2009).

[18] T. Konrad, O. Gühne, J. Audretsch, and H. J. Briegel, Phys. Rev. A **75**, 062101 (2007).

[19] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[20] H. Teicher, Ann. Math. Stat. **34**, 1265 (1963).

[21] D. C. Boes, Ann. Math. Stat. **37**, 177 (1966).

[22] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Munoz-Tapia, Phys. Rev. A **73**, 032301 (2006).

[23] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).

[24] C. Rao, Bull. Calcutta Math. Soc. **37**, 81 (1945).

[25] J. D. Gorman and A. O. Hero, IEEE Trans. Inf. Theory **26**, 1285 (1990).

[26] M. Hayashi and K. Matsumoto, J. Math. Phys. **49**, 102101 (2008).

[27] J. Kahn and M. Guţă, Comm. Math. Phys. **289**, 597 (2009).