Quantum state estimation with informationally overcomplete measurements

Huangjun Zhu*

Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada N2L 2Y5;
Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore;
and NUS Graduate School for Integrative Sciences and Engineering, Singapore 117597, Singapore

(Received 14 April 2014; published 28 July 2014)

We study informationally overcomplete measurements for quantum state estimation so as to clarify their tomographic significance as compared with minimal informationally complete measurements. We show that informationally overcomplete measurements can improve the tomographic efficiency significantly over minimal measurements when the states of interest have high purities. Nevertheless, the efficiency is still too limited to be satisfactory with respect to figures of merit based on monotone Riemannian metrics, such as the Bures metric and quantum Chernoff metric. In this way, we also pinpoint the limitation of nonadaptive measurements and motivate the study of more sophisticated measurement schemes. In the course of our study, we introduce the best linear unbiased estimator and show that it is equally efficient as the maximum likelihood estimator in the large sample limit. This estimator may significantly outperform the canonical linear estimator for states with high purities. It is expected to play an important role in experimental designs and adaptive quantum state tomography besides its significance to the current study.

DOI: 10.1103/PhysRevA.90.012115 PACS number(s): 03.65.Wj, 03.67.−a

I. INTRODUCTION

Quantum state estimation is a procedure for inferring the state of a quantum system from generalized measurements [1,2]. A central problem in quantum state estimation is to determine the state of a quantum system as efficiently as possible with suitable measurements and data processing. In practice, the set of accessible measurements is usually determined by experimental settings, which are not easy to modify. Given an ensemble of identically prepared quantum systems, the simplest measurement schemes consist of identical and independent measurements on individual copies. A measurement is *informationally complete* (IC) if every state is determined completely by the measurement statistics [3–5]. Such a measurement has at least d^2 outcomes for a d-level quantum system. An IC measurement is *minimal* if it has exactly d^2 outcomes and *informationally overcomplete* (IOC) otherwise. A prominent example of minimal IC measurements are *symmetric informationally complete* (SIC) measurements [6–9], whereas measurements composed of complete sets of *mutually unbiased bases* (MUB) [10–12] are IOC. Note, however, that the later measurements are minimal IC among combinations of projective measurements. Another example of IOC measurements is the *covariant measurement*, whose outcomes consist of all pure states weighted by the Haar measure. The efficiencies of minimal IC measurements and special IOC measurements, such as mutually unbiased measurements, have been studied extensively in the literature [9–28]. Still we often hear the basic question: which one is more efficient for state estimation, SIC or MUB? Even little is known about general IOC measurements [21,29]. In particular, it is not so clear what is the efficiency limit of IOC measurements, whether such measurements are useful in improving the tomographic efficiency over minimal IC measurements, and when and to what extent if the answer is positive. These general questions are the main motivations behind the present study, which extends some recent work presented in the author's thesis [27].

To answer the questions raised in the previous paragraph, we need to choose suitable figures of merit and estimators. Among common choices of figures of merit are the mean-square error (MSE) with respect to the Hilbert-Schmidt (HS) distance and its generalization—weighted mean-square errors (WMSEs), which include the mean-square Bures distance (MSB) as a special case. In traditional linear state tomography, the estimator is constructed in terms of measurement frequencies and reconstruction operators [1,17,18,26,27]. The set of canonical reconstruction operators is optimal if these operators are required to be independent of the measurement statistics [17,26,27]. However, such a choice generally cannot make full use of the information provided by an IOC measurement. To make a fair comparison among various measurements entails considering reconstruction operators that are optimal in the pointwise sense, which may depend on the measurement statistics. A similar problem has been addressed by D'Ariano and Perinotti [30] (see also Refs. [31,32]), who derived the set of optimal reconstruction operators with respect to the MSE in estimating certain observables. The situation is not so clear concerning other figures of merit, such as the WMSE corresponding to a generic weighting matrix, say, the MSB. Furthermore, several basic questions are not well understood. For example, by how much can the efficiency be improved with the optimal reconstruction operators instead of the canonical choice?

In this paper, we determine the set of optimal reconstruction operators in the pointwise sense and derive the *best linear unbiased estimator* (BLUE), using the MSE matrix as a benchmark. The BLUE is as efficient as the *maximum likelihood estimator* (MLE) [1,33–35] in the large sample limit. Compared with the ML approach, our approach has the merit that it is parametrization independent and is thus often much easier to work with and easier for deriving analytical results. Also, it can help clarify the differences between canonical

*hzhu@pitp.ca

state reconstruction and optimal reconstruction since the two alternatives are treated in a unified framework. Our approach is simpler than the one studied in Ref. [30], but the result has wider applicability. In particular, it is applicable to study tomographic efficiencies with respect to a variety of figures of merit, including various WMSEs, such as the MSE and MSB, as well as the volume of the uncertainty ellipsoid, which is pertinent to constructing good region estimators [36–38]. Furthermore, the current work provides a stepping stone for exploring quantum state estimation with more sophisticated measurement schemes, such as adaptive measurements [27]. What is more remarkable is that certain results presented here prove to be useful for studying information theoretic analogs of uncertainty and complementarity relations [39].

Based on the above work, we show that covariant measurements are optimal among all nonadaptive measurements in minimizing the average WMSE based on any unitarily invariant distance, including the MSE and the MSB. Compared with minimal IC measurements, covariant measurements can improve the tomographic efficiency significantly when the states of interest have high purities. However, the efficiency is still too limited to be satisfactory with respect to the scaled MSB, which diverges at the boundary of the state space in the large-sample limit. This divergence is also persistent for any scaled WMSE based on a monotone Riemannian metric [40–42] as long as the measurement is nonadaptive, in sharp contrast with the intuitive belief that states with high purities are easier to estimate than states with low purities. These general conclusions are further corroborated by extensive study of qubit state estimation with IOC measurements. Our work not only clarifies the power of IOC measurements compared with minimal IC measurements, but also pinpoints the limitation of nonadaptive measurements, thereby motivating the exploration of more sophisticated measurement schemes, which we hope to address in the future.

The rest of the paper is organized as follows. In Sec. II, we discuss optimal state reconstruction for IOC measurements in comparison with canonical reconstruction and illustrate the matter with SIC and MUB measurements. In Sec. III, we clarify the efficiency advantage of IOC measurements over minimal IC measurements as well as the limitation of nonadaptive measurements. In Sec. IV, we focus on qubit state estimation with IOC measurements. Section V summarizes this paper.

II. OPTIMAL STATE RECONSTRUCTION FOR INFORMATIONALLY OVERCOMPLETE MEASUREMENTS

In this section we study optimal state reconstruction for general IC measurements with emphasis on IOC measurements, in preparation for the discussions in the rest of the paper. In particular, we determine the BLUE for any IC measurement and show that it is as efficient as the MLE in the large sample limit as long as the states of interest are not on the boundary of the state space. As an application of this result, we clarify the relative merits of SIC and MUB measurements in quantum state estimation. To this end, we first need to review the basic framework of linear state tomography [1,17,18,26,27].

A. Linear state tomography

A generalized measurement is composed of a set of outcomes represented mathematically by positive operators Π_ξ that sum up to the identity 1 [43] (this simplified description is enough for us since we are only concerned with the measurement statistics, not the state after the measurement). Given an unknown state ρ, the probability of obtaining the outcome Π_ξ is given by the Born rule: $p_\xi = \mathrm{tr}(\Pi_\xi \rho)$. Following the convention in Refs. [26,27] (see also Refs. [30,44]), the probability can be expressed as an inner product $\langle\langle \Pi_\xi | \rho \rangle\rangle$ between the operator kets $|\Pi_\xi\rangle\rangle$ and $|\rho\rangle\rangle$, where the double ket notation is used to distinguish them from ordinary kets. A measurement is IC if every state is determined by the measurement statistics, namely, the set of probabilities p_ξ. This amounts to the requirement that the *frame superoperator*

$$\mathcal{F} = d \sum_\xi \frac{|\Pi_\xi\rangle\rangle\langle\langle\Pi_\xi|}{\mathrm{tr}(\Pi_\xi)} \tag{1}$$

is invertible [17,26,30], where the factor d is introduced for the convenience of later discussions.

For an IC measurement, we can find a set of reconstruction operators Θ_ξ with the property $\sum_\xi |\Theta_\xi\rangle\rangle\langle\langle\Pi_\xi| = \mathbf{I}$, where $\mathbf{I}$ is the identity superoperator. Then any state can be recovered from the set of probabilities p_ξ as $\rho = \sum_\xi p_\xi \Theta_\xi$. In practice, the probabilities p_ξ need to be replaced by the frequencies f_ξ since the number N of measurements is finite. The estimator based on these frequencies $\hat{\rho} = \sum_\xi f_\xi \Theta_\xi$ is thus different from the true state. Nevertheless, the requirement $\sum_\xi |\Theta_\xi\rangle\rangle\langle\langle\Pi_\xi| = \mathbf{I}$ on the reconstruction operators guarantees that the estimator is unbiased, that is, $E(\hat{\rho}) = \rho$. In general, these frequencies obey a multinomial distribution with the scaled covariance matrix (that is the covariance matrix multiplied by the number of measurements) $\Sigma_{\xi\zeta} = p_\xi \delta_{\xi\zeta} - p_\xi p_\zeta$. The scaled MSE matrix (or covariance matrix) of the estimator $\hat{\rho}$ is then determined by the formula of error propagation [26],

$$\mathcal{C}(\rho) = \sum_{\xi,\zeta} |\Theta_\xi\rangle\rangle \Sigma_{\xi\zeta} \langle\langle\Theta_\zeta| = \sum_\xi |\Theta_\xi\rangle\rangle p_\xi \langle\langle\Theta_\xi| - |\rho\rangle\rangle\langle\langle\rho|. \tag{2}$$

Denote by $\Delta\rho = \sqrt{N}(\hat{\rho} - \rho)$ the scaled deviation of the estimator from the true state. Then the scaled MSE with respect to the HS distance reads

$$\mathcal{E}(\rho) := E\big(\,\|\Delta\rho\|_{\mathrm{HS}}^2\,\big) = \mathrm{Tr}\{\mathcal{C}(\rho)\} = \sum_\xi p_\xi \,\mathrm{tr}\big(\Theta_\xi^2\big) - \mathrm{tr}(\rho^2). \tag{3}$$

Here "Tr" denotes the trace of a superoperator, and "tr" of an ordinary operator.

The set of reconstruction operators is not unique except for a minimal IC measurement, such as a SIC measurement. In linear state tomography, usually the set of reconstruction operators, once chosen, is independent of the measurement statistics. In that case, the set of *canonical reconstruction operators*

$$|\Theta_\xi\rangle\rangle = \frac{d\mathcal{F}^{-1}|\Pi_\xi\rangle\rangle}{\mathrm{tr}(\Pi_\xi)} \tag{4}$$

is the best choice in the sense of minimizing the MSE averaged over unitarily equivalent true states [17,18,26,27].

The resulting estimator is called *canonical linear estimator* (CLE). The situation is different if reconstruction operators are allowed to depend on the measurement statistics, which is the focus of the next section.

B. Best linear unbiased estimator

In this section we determine the set of optimal reconstruction operators in the pointwise sense and derive the BLUE.

The following lemma is crucial to achieving our goal. Its proof is relegated to Appendix A.

Lemma 1. Suppose A and B are two $m \times n$ matrices such that $AB^\dagger$ is the projector on the support of $B^\dagger$ (that is the range of B). Then $AA^\dagger \geqslant (BB^\dagger)^+$, and the inequality is saturated if and only if $A = B^{\dagger +} = (BB^\dagger)^+ B$. If, in addition, $AB^\dagger = 1$, then $AA^\dagger \geqslant (BB^\dagger)^{-1}$, and the inequality is saturated if and only if $A = (BB^\dagger)^{-1}B$.

Here A^+ denotes the (Moore-Penrose) pseudoinverse of A (the arithmetics of pseudoinverses can be found in Ref. [45]).

Given Eq. (2), Lemma 1 applied to the matrices $(|\Theta_1\rangle\rangle p_1^{1/2}, |\Theta_2\rangle\rangle p_2^{1/2}, \ldots)$ and $(|\Pi_1\rangle\rangle p_1^{-1/2}, |\Pi_2\rangle\rangle p_2^{-1/2}, \ldots)$ with respect to a suitable operator basis yields

$$\mathcal{C}(\rho) \geqslant \mathcal{F}(\rho)^{-1} - |\rho\rangle\rangle\langle\langle\rho|, \tag{5}$$

where

$$\mathcal{F}(\rho) = \sum_\xi |\Pi_\xi\rangle\rangle \frac{1}{p_\xi} \langle\langle\Pi_\xi| \tag{6}$$

is also called the frame superoperator, which generalizes the definition in Eq. (1). To avoid unnecessary technicality, we assume that ρ has full rank and thus $p_\xi > 0$ for all ξ; rank-deficient states can be treated in suitable limits. The inequality is saturated if and only if the reconstruction operators are of the form

$$|\Theta_\xi\rangle\rangle = p_\xi^{-1} \mathcal{F}(\rho)^{-1} |\Pi_\xi\rangle\rangle, \tag{7}$$

in which case we get the BLUE along with the scaled MSE matrix

$$\mathcal{C}(\rho) = \mathcal{F}(\rho)^{-1} - |\rho\rangle\rangle\langle\langle\rho|. \tag{8}$$

According to the Aitken theorem, a generalization of the Gauss-Markov theorem, the BLUE is a special instance of *weighted linear least square estimators* for which the weighting matrix is the inverse of the covariance matrix of the measurement statistics [46] (note that the weighting matrix here is different from the one in the definition of WMSE).

The scaled WMSE of the BLUE for a given weighting matrix $\mathcal{W}$ reads

$$\mathcal{E}_\mathcal{W}(\rho) = \mathrm{Tr}\{\mathcal{W}\mathcal{F}(\rho)^{-1}\} - \langle\langle\rho|\mathcal{W}|\rho\rangle\rangle. \tag{9}$$

It reduces to the scaled MSE (with respect to the HS distance) when $\mathcal{W}$ is the identity,

$$\mathcal{E}(\rho) = \mathrm{Tr}\{\mathcal{F}(\rho)^{-1}\} - \mathrm{tr}(\rho^2). \tag{10}$$

The volume of the scaled uncertainty ellipsoid is given by

$$\begin{aligned}\mathcal{V}(\rho) &= V_{d^2-1}\sqrt{\bar{\mathrm{Det}}\{\mathcal{C}(\rho)\}} \\ &= V_{d^2-1}\sqrt{\bar{\mathrm{Det}}\{\mathcal{F}(\rho)^{-1} - |\rho\rangle\rangle\langle\langle\rho|\}},\end{aligned} \tag{11}$$

where

$$V_{d^2-1} = \frac{\pi^{(d^2-1)/2}}{\Gamma\left(\frac{d^2+1}{2}\right)} \tag{12}$$

is the volume of the (d^2-1)-dimensional unit ball, and $\bar{\mathrm{Det}}(\mathcal{O})$ denotes the determinant of the restriction of $\mathcal{O}$ onto the space of traceless Hermitian operators. All superoperators in this paper of which we need to evaluate $\bar{\mathrm{Det}}$ are supported on this space. In particular, this is the case for $\mathcal{C}(\rho)$, as we shall see shortly.

The inequality in Eq. (5) implies that the BLUE is optimal not only in minimizing the MSE but also in minimizing any other cost function that is monotonic increasing in the MSE matrix, such as various WMSEs and the volume of the uncertainty ellipsoid. This observation is crucial to investigating the efficiency advantage of the optimal state reconstruction over canonical reconstruction. It is also indispensable for clarifying the questions of whether and to what extent IOC measurements are helpful in improving the tomographic efficiency over minimal IC measurements. Furthermore, the formulas for the BLUE and its associated MSE matrix can serve as a benchmark for selecting more efficient measurement schemes, thereby providing a stepping stone for studying experimental designs and adaptive quantum state tomography [27].

When ρ is the completely mixed state, Eqs. (6) and (7) reduce to Eqs. (1) and (4), respectively, and it follows that the set of canonical reconstruction operators and the CLE are optimal. This observation implies that the canonical reconstruction is optimal in minimizing the WMSE averaged over unitarily equivalent states as long as the weighting matrix is state independent. In the case the weighting matrix is a constant matrix, this conclusion reduces to the one of Scott that the set of canonical reconstruction operators is optimal in minimizing the average MSE [17] (see Sec. II A).

Meticulous readers may have noticed that the optimal reconstruction operators depend on the true state, which is usually unknown. To remedy this problem, we may replace the true state in the relevant formulas with an estimator obtained from another reconstruction scheme, canonical reconstruction for instance. Alternatively, we may just replace probabilities p_ξ with frequencies f_ξ in Eqs. (6) and (7). In that case, the final estimator is no longer linear in the frequencies. So strictly speaking, the BLUE is not a linear estimator in the usual sense. Nevertheless, the resulting estimator is almost as good as the theoretical BLUE as long as N is not too small. To see this, note that for an IC measurement, any reasonable estimator, such as the CLE, will converge to the true state in the large-N limit. Therefore, intuitively, the reconstruction operators based on the estimator will also converge to the theoretical optimal reconstruction operators. Numerical calculation indicates that the MSE between the approximate BLUE and the theoretical BLUE decreases approximately as $1/N^2$, in sharp contrast with the scaling law $1/N$ of the MSE between each estimator and the true state. For most values of N of practical interest, there is almost no difference between the two estimators, as illustrated in Fig. 1 along with the CLE and MLE (cf. Sec. II D and Appendix C). Therefore, the BLUE is useful not only to theoretical study but also to practical applications.

For the convenience of subsequent discussions, here we collect several basic properties of the frame superoperator and

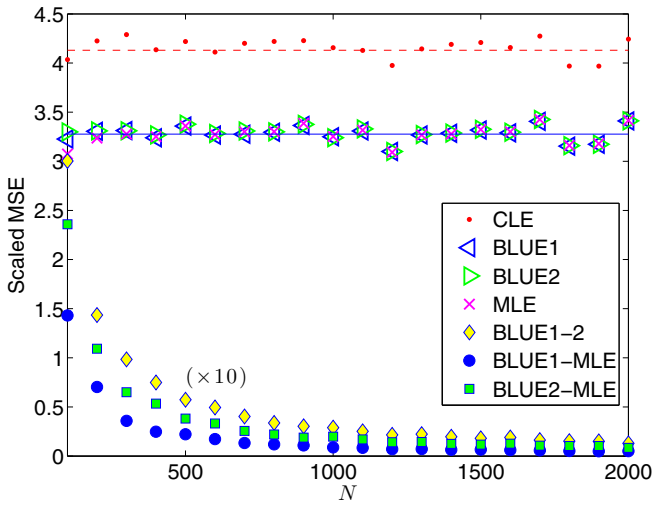

FIG. 1. (Color online) Tomographic efficiencies of the CLE, BLUE, and MLE. The scaled MSEs of these estimators are determined by numerical simulation of the cube measurement (cf. Sec. IV) on a qubit state with random Bloch vector $\boldsymbol{s} = (0.6886, 0.1137, -0.5025)$. Each data point is an average over 1000 repetitions. BLUE1 assumes the knowledge of the true state in computing the reconstruction operators, while BLUE2 uses frequencies instead of probabilities in relevant formulas. The theoretical scaled MSEs of the CLE and BLUE are shown as dashed line and solid line, respectively. Also plotted are pairwise scaled MSEs (multiplied by a factor of 10 for ease of viewing) among BLUE1, BLUE2, and MLE. The figure indicates that the three estimators are almost equally efficient as long as N is not too small.

the optimal reconstruction operators,

$$\mathcal{F}(\rho)|\rho\rangle\!\rangle = |1\rangle\!\rangle, \quad \mathcal{F}(\rho)^{-1}|1\rangle\!\rangle = |\rho\rangle\!\rangle, \tag{13a}$$

$$\mathrm{tr}(\Theta_\xi) = 1, \tag{13b}$$

$$\sum_\xi \mathrm{tr}(\Pi_\xi)\Theta_\xi = 1. \tag{13c}$$

Equation (13a) follows from the definition of $\mathcal{F}(\rho)$; Eq. (13b) can be derived by multiplying both sides of Eq. (7) with $\langle\!\langle 1|$ and applying Eq. (13a); Eq. (13c) follows from the requirement $\sum_\xi |\Theta_\xi\rangle\!\rangle\langle\!\langle \Pi_\xi| = \mathbf{I}$ and thus holds for any set of reconstruction operators, regardless of whether it is optimal or not.

According to Eqs. (8) and (13a), $|1\rangle\!\rangle$ is a null eigenvector of $\mathcal{C}(\rho)$; that is, $\mathcal{C}(\rho)$ is supported on the space of traceless Hermitian operators as claimed before. Let $\bar{\mathbf{I}}$ denote the projector onto this space and define $\bar{\mathcal{F}}(\rho)$ as the projection of $\mathcal{F}(\rho)$ onto this space,

$$\bar{\mathcal{F}}(\rho) := \bar{\mathbf{I}}\mathcal{F}(\rho)\bar{\mathbf{I}} = \sum_\xi |\bar{\Pi}_\xi\rangle\!\rangle \frac{1}{p_\xi} \langle\!\langle \bar{\Pi}_\xi|, \tag{14}$$

where $\bar{\Pi}_\xi = \Pi_\xi - \mathrm{tr}(\Pi_\xi)/d$. Then we can deduce from Eq. (13) that $\mathcal{C}(\rho)\bar{\mathcal{F}}(\rho) = \bar{\mathbf{I}}$, which implies that $\mathcal{C}(\rho)$ is the inverse of $\bar{\mathcal{F}}(\rho)$ in the space of traceless Hermitian operators. Consequently,

$$\mathcal{C}(\rho) = \bar{\mathcal{F}}(\rho)^+, \quad \mathcal{E}_{\mathcal{W}}(\rho) = \mathrm{Tr}\{\mathcal{W}\bar{\mathcal{F}}(\rho)^+\},$$
$$\mathcal{E}(\rho) = \mathrm{Tr}\{\bar{\mathcal{F}}(\rho)^+\}, \quad \mathcal{V}(\rho) = V_{d^2-1}[\bar{\mathrm{Det}}\{\bar{\mathcal{F}}(\rho)\}]^{-1/2}. \tag{15}$$

Comparison with Eq. (8) yields

$$\bar{\mathcal{F}}(\rho)^+ = \mathcal{F}(\rho)^{-1} - |\rho\rangle\!\rangle\langle\!\langle\rho|. \tag{16}$$

This simple formula is quite useful in later study.

In the rest of this section, we briefly discuss the problem of state reconstruction when the measurement is not IC [47]. This problem is also relevant to studying IOC measurements, such as mutually unbiased measurements, since many of them are combinations of informationally incomplete measurements.

For an informationally incomplete measurement, it is generally impossible to infer the true state accurately even if the sample size is arbitrarily large. Nevertheless, the projection of the true state onto the reconstruction subspace, the space spanned by the Π_ξ, can be determined in the asymptotic limit. Let ρ_{R} and $\mathcal{C}_{\mathrm{R}}(\rho)$ be the restrictions of the true state and the scaled MSE matrix onto the reconstruction subspace. Then using a similar argument that leads to Eq. (5), we find

$$\mathcal{C}_{\mathrm{R}}(\rho) \geqslant \mathcal{F}(\rho)^+ - |\rho_{\mathrm{R}}\rangle\!\rangle\langle\!\langle\rho_{\mathrm{R}}| = \bar{\mathcal{F}}(\rho)^+. \tag{17}$$

The inequality is saturated if and only if the reconstruction operators are given by

$$|\Theta_\xi\rangle\!\rangle = p_\xi^{-1}\mathcal{F}(\rho)^+|\Pi_\xi\rangle\!\rangle, \tag{18}$$

when restricted to the reconstruction subspace.

To illustrate the above idea, let us consider a rank-one projective measurement $\{\Pi_\xi\}$ for example. Noticing that the outcomes Π_ξ are orthogonal projectors and $\rho_{\mathrm{R}} = \sum_\xi p_\xi \Pi_\xi$, we get

$$\mathcal{C}_{\mathrm{R}}(\rho) = \sum_\xi |\Pi_\xi\rangle\!\rangle p_\xi \langle\!\langle \Pi_\xi| - \sum_{\xi,\zeta} |\Pi_\xi\rangle\!\rangle p_\xi p_\zeta \langle\!\langle \Pi_\zeta|,$$
$$\mathcal{E}_{\mathrm{R}}(\rho) = \mathrm{Tr}\{\mathcal{C}_{\mathrm{R}}(\rho)\} = 1 - \sum_\xi p_\xi^2. \tag{19}$$

C. Illustration with SIC and MUB measurements

To illustrate the improvement of the BLUE over the CLE and to answer a question raised in Sec. I, here we consider state estimation with SIC measurements and complete sets of mutually unbiased measurements. Although the main results concerning SIC and MUB presented in this section were known before, they were derived under various different assumptions scattered in the literature, and a coherent account is still lacking. We hope to bridge this gap by stating the conclusion explicitly and precisely within a unified framework.

In a d-dimensional Hilbert space, a SIC measurement is composed of d^2 subnormalized projectors onto pure states $\Pi_\xi = |\psi_\xi\rangle\langle\psi_\xi|/d$ with equal pairwise fidelity [6,7],

$$|\langle\psi_\xi|\psi_\zeta\rangle|^2 = \frac{d\delta_{\xi\zeta} + 1}{d+1}; \tag{20}$$

see Refs. [8,9,27] for the latest developments. Two bases $\{|\psi_j\rangle\}$ and $\{|\phi_k\rangle\}$ are mutually unbiased if all the transition probabilities $|\langle\psi_j|\phi_k\rangle|^2$ across their basis elements are equal to $1/d$ [10–12]. In a d-dimensional Hilbert space, there exist at most $d+1$ MUB; such a maximal set, if it exists, is called complete. When d is a prime power, a complete set of MUB can be constructed explicitly [10,11]; see Ref. [12] for a review. Two (rank-one) projective measurements are mutually unbiased if their measurement bases are mutually unbiased.

Applications of SIC and MUB to quantum state estimation have been studied extensively in the literature [9–12,14–28].

For a minimal IC measurement, the optimal reconstruction is identical with the canonical reconstruction. The scaled MSE averaged over unitarily equivalent states is bounded below by

$$\overline{\mathcal{E}(\rho)} \geqslant d^2 + d - 1 - \mathrm{tr}(\rho^2), \tag{21}$$

and the lower bound is saturated if and only if the measurement is SIC [9,17,18,26,27]. For a SIC measurement, the scaled MSE is unitarily invariant, so we have

$$\mathcal{E}(\rho) = \overline{\mathcal{E}(\rho)} = d^2 + d - 1 - \mathrm{tr}(\rho^2). \tag{22}$$

The bound in Eq. (21) is also applicable to IOC measurements, such as mutually unbiased measurements if canonical reconstruction is applied. The lower bound is saturated if and only if the measurement is composed of subnormalized pure states that form a weighted 2-design [7,9,17,26,27], that is, $\Pi_\xi = |\psi_\xi\rangle w_\xi \langle\psi_\xi|$ with $\sum_\xi w_\xi = d$ and

$$\sum_\xi w_\xi (|\psi_\xi\rangle\langle\psi_\xi|)^{\otimes 2} = \frac{2}{d+1} P_\mathrm{s}, \tag{23}$$

where P_s is the projector onto the bipartite symmetric subspace. Such a measurement is called *tight IC*. In that case, the canonical reconstruction operators have a very simple form,

$$\Theta_\xi = |\psi_\xi\rangle(d+1)\langle\psi_\xi| - 1, \tag{24}$$

and the scaled MSE is also unitarily invariant [17,18,26,27]. Since both MUB and SIC form 2-designs, it follows that they are equally efficient with respect to the MSE under canonical reconstruction.

The situation is different if the optimal reconstruction is employed. Now the scaled MSE achievable with MUB is given by [16,18,27]

$$\mathcal{E}(\rho) = \overline{\mathcal{E}(\rho)} = (d+1)[d - \mathrm{tr}(\rho^2)]. \tag{25}$$

Therefore, MUB is more efficient than SIC under the optimal reconstruction, especially for states with high purities. This example shows that the optimal reconstruction is crucial to unleashing the full potential of IOC measurements and to making sensible comparison among various measurement schemes. In addition, it demonstrates that IOC measurements can indeed improve the tomographic efficiency over minimal IC measurements in quantum state estimation, as discussed in more detail in Secs. III and IV.

D. Connection with the maximum-likelihood method

To elucidate the connection between the BLUE and the MLE [1,35], we need to introduce a suitable parametrization for the quantum state space. A convenient choice is the affine parametrization

$$\rho(\theta) = \frac{1}{d} + \sum_{j=1}^{d^2-1} \theta_j E_j, \tag{26}$$

where the E_j form an orthonormal basis in the space of traceless Hermitian operators. Now the Fisher information matrix takes on the form (see Appendix B)

$$I_{jk}(\theta) = \sum_\xi \frac{\langle\!\langle E_j|\Pi_\xi\rangle\!\rangle\langle\!\langle\Pi_\xi|E_k\rangle\!\rangle}{p_\xi} = \langle\!\langle E_j|\mathcal{F}(\rho)|E_k\rangle\!\rangle$$
$$= \langle\!\langle E_j|\bar{\mathcal{F}}(\rho)|E_k\rangle\!\rangle. \tag{27}$$

This equation clearly indicates that the superoperator $\bar{\mathcal{F}}(\rho)$ is essentially the Fisher information matrix in disguise and that the BLUE is as efficient as the MLE in the large-N limit as long as the true state is not on the boundary of the state space (see Fig. 1 for an illustration). Recall that the MSE matrix of any unbiased estimator is lower bounded by the inverse of the Fisher information matrix and that the bound can be saturated asymptotically with the MLE [33,34,48,49] (see Appendix B). This observation implies that the BLUE is optimal not only among linear unbiased estimators but also among all unbiased estimators in the asymptotic limit.

Alternatively, we can clarify the connection between the BLUE and the MLE by inspecting the likelihood functional $\mathcal{L}(\rho)$ (see Appendix C) in the large-N limit. According to Eq. (C2),

$$\frac{\partial^2 \ln \mathcal{L}(\rho)}{\partial\theta_j\partial\theta_k} = -N\sum_\xi \frac{f_\xi}{p_\xi^2}\,\mathrm{tr}(E_j\Pi_\xi)\,\mathrm{tr}(\Pi_\xi E_k)$$
$$\approx -N\sum_\xi \frac{1}{p_\xi}\,\mathrm{tr}(E_j\Pi_\xi)\,\mathrm{tr}(\Pi_\xi E_k)$$
$$= -N\langle\!\langle E_j|\mathcal{F}(\rho)|E_k\rangle\!\rangle = -N\langle\!\langle E_j|\bar{\mathcal{F}}(\rho)|E_k\rangle\!\rangle. \tag{28}$$

Suppose that the likelihood functional is maximized at $\tilde{\theta}$. Let $\Delta\theta = \theta - \tilde{\theta}$; then

$$\frac{1}{N}\ln\mathcal{L}(\rho) \approx c - \frac{1}{2}\sum_{j,k}\Delta\theta_j\,\Delta\theta_k\langle\!\langle E_j|\bar{\mathcal{F}}(\rho)|E_k\rangle\!\rangle, \tag{29}$$

where c is a constant. Again, we find that $\bar{\mathcal{F}}(\rho)$ plays the role of the Fisher information matrix.

Compared with the ML method, our approach is independent of the parametrization and is thus often more convenient to work with. In particular, it allows deriving analytical results more easily, thereby elucidating the dependence of the cost function on various parameters, such as the dimension of the Hilbert space and the purity. Also, our approach can better clarify the differences between canonical state reconstruction and optimal reconstruction as well as the differences between minimal IC measurements and IOC measurements. In addition, it is quite helpful for studying adaptive measurements and quantum precision limit [27]. The drawback of our approach is that the optimal reconstruction operators need to be chosen adaptively, and it is not easy to take into account naturally the positivity constraint on the density operators. Depending on the situation, one alternative may be preferable to the other, and a judicious choice is crucial to simplifying the problem.

III. TOMOGRAPHIC SIGNIFICANCE AND LIMITATION OF IOC MEASUREMENTS

In this section we investigate the tomographic efficiency of IOC measurements in comparison with minimal IC measurements, so as to answer the questions of when and to what

extent IOC measurements are advantageous over minimal IC measurements. Our study also clarifies the limitation of nonadaptive measurements for quantum state estimation. As we shall see shortly, covariant measurements play a crucial role in understanding the tomographic significance of IOC measurements, although it is not practical to implement them in practice. Most previous studies on covariant measurements focused on pure-state models [50]. Our study fills the gap in the case of mixed states.

A. Optimality of the covariant measurement

Suppose $\bar{\mathcal{F}}_1(\rho)$ and $\bar{\mathcal{F}}_2(\rho)$ are the Fisher information matrices associated with two given IC measurements. If the two measurements are performed with probabilities p_1 and $p_2 = 1 - p_1$, then the Fisher information matrix is a convex combination,

$$\bar{\mathcal{F}}(\rho) = p_1\bar{\mathcal{F}}_1(\rho) + p_2\bar{\mathcal{F}}_2(\rho). \tag{30}$$

Since the function $1/x$ is operator convex over the interval $(0,\infty)$ [51], it follows that

$$\mathcal{C}(\rho) \leqslant p_1\mathcal{C}_1(\rho) + p_2\mathcal{C}_2(\rho), \quad \mathcal{E}(\rho) \leqslant p_1\mathcal{E}_1(\rho) + p_2\mathcal{E}_2(\rho). \tag{31}$$

Taking average over unitarily equivalent states yields

$$\overline{\mathcal{E}(\rho)} \leqslant p_1\overline{\mathcal{E}_1(\rho)} + p_2\overline{\mathcal{E}_2(\rho)}. \tag{32}$$

As a consequence, $\overline{\mathcal{E}(\rho)} \leqslant \overline{\mathcal{E}_1(\rho)} = \overline{\mathcal{E}_2(\rho)}$ if the two given measurements are unitarily equivalent. In other words, the average MSE never increases by combining unitarily equivalent measurements. Given that the set of optimal measurements contains at least one measurement that is composed of subnormalized pure states, we conclude that the average MSE is minimized by the covariant measurement. By the same token, so is the average WMSE based on any unitarily invariant distance, such as the Bures distance.

In addition to minimizing average WMSEs based on various unitarily invariant distances, the covariant measurement is also optimal in minimizing the average log volume of the uncertainty ellipsoid. To see this,

$$\begin{aligned}\ln \mathcal{V}(\rho) &= \ln V_{d^2-1} + \tfrac{1}{2}\ln \bar{\mathrm{D}}\mathrm{et}\{\mathcal{C}(\rho)\} \\ &= \ln V_{d^2-1} - \tfrac{1}{2}\ln \bar{\mathrm{D}}\mathrm{et}\{\bar{\mathcal{F}}(\rho)\} \\ &= \ln V_{d^2-1} - \tfrac{1}{2}\bar{\mathrm{T}}\mathrm{r}\,\ln\{\bar{\mathcal{F}}(\rho)\},\end{aligned} \tag{33}$$

where "$\bar{\mathrm{T}}\mathrm{r}$" denotes the trace on the space of traceless Hermitian operators. Observing that the function $\ln(x)$ is operator concave [51], we deduce

$$\ln \mathcal{V}(\rho) \leqslant p_1 \ln \mathcal{V}_1(\rho) + p_2 \ln \mathcal{V}_2(\rho). \tag{34}$$

Now our claim follows from the same reasoning as in the previous paragraph.

B. Efficiency of the covariant measurement with canonical reconstruction

As we have seen in the previous section, the covariant measurement sets the efficiency limit to nonadaptive measurements, so it is crucial to understand its tomographic efficiency. Before investigating its performance under the optimal reconstruction, it is instructive to consider the situation under the canonical reconstruction. The covariant measurement is a special instance of *isotropic measurements*, whose outcomes form not only (weighted) 2-designs, but also 3-designs [26,27]. Under canonical reconstruction, isotropic measurements share the same covariant MSE matrix and are thus equally efficient with respect to any figure of merit that is a function of the MSE matrix, including various WMSEs and the volume of the uncertainty ellipsoid. So the conclusions in this section also apply to any isotropic measurement.

To evaluate the tomographic efficiency of the covariant measurement, which is unitarily invariant, without loss of generality, we may assume that ρ is diagonal with eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_d$. Under canonical reconstruction, the scaled MSE matrix associated with the covariant measurement is given by [cf. Eq. (2)]

$$\begin{aligned}\mathcal{C}(\rho) &= d\int d\mu(\psi)(|\Theta_\psi\rangle\!\rangle\langle\psi|\rho|\psi\rangle\langle\!\langle\Theta_\psi| - |\rho\rangle\!\rangle\langle\!\langle\rho| \\ &= \sum_{j,k}\mathcal{Q}_{jk}(|E_{jj}\rangle\!\rangle\langle\!\langle E_{kk}|) \\ &\quad + \frac{d+1}{d+2}\sum_{j\neq k}(1+\lambda_j+\lambda_k)(|E_{jk}\rangle\!\rangle\langle\!\langle E_{jk}|),\end{aligned} \tag{35}$$

where the $\Theta_\psi = |\psi\rangle(d+1)\langle\psi| - 1$ are reconstruction operators [see Eq. (24)], $d\mu(\psi)$ is the normalized Haar measure, $E_{jk} = |j\rangle\langle k|$, and

$$\mathcal{Q}_{jk} = \frac{(d+1)(1+2\lambda_j)\delta_{jk} - 1 - \lambda_j - \lambda_k - (d+2)\lambda_j\lambda_k}{d+2}. \tag{36}$$

Define

$$\begin{aligned}E^+_{jk} &:= \frac{1}{\sqrt{2}}(|j\rangle\langle k| + |k\rangle\langle j|), \\ E^-_{jk} &:= -\frac{i}{\sqrt{2}}(|j\rangle\langle k| - |k\rangle\langle j|).\end{aligned} \tag{37}$$

Then $E^\pm_{jk}$ with $j \neq k$ are eigenvectors of $\mathcal{C}(\rho)$ with eigenvalues $(d+1)(1+\lambda_j+\lambda_k)/(d+2)$.

The scaled MSE agrees with Eq. (22) as expected for a tight IC measurement. The scaled MSB reads

$$\begin{aligned}\mathcal{E}_{\mathrm{SB}}(\rho) &= \frac{2d^3+2d^2-3d-2}{4(d+2)} \\ &\quad + \frac{1}{4(d+2)}\left[\sum_j\frac{d}{\lambda_j} + \sum_{j\neq k}\frac{2(d+1)}{\lambda_j+\lambda_k}\right],\end{aligned} \tag{38}$$

where we have applied the formula for the Bures metric derived by Hübner [52],

$$D^2_{\mathrm{B}}(\rho,\rho+d\rho) = \frac{1}{2}\sum_{j,k}\frac{|\langle j|d\rho|k\rangle|^2}{\lambda_j+\lambda_k}. \tag{39}$$

Note that the scaled MSB diverges at the boundary of the state space. The same is true for the scaled WMSE based on any monotone Riemannian metric because the Bures metric is minimal among such metrics [40–42]. To see this explicitly, observe that up to a multiplicative constant a generic monotone

Riemannian metric has the form

$$D_c^2(\rho,\rho+d\rho)=\sum_j\frac{|\langle j|d\rho|j\rangle|^2}{4\lambda_j}+\sum_{j\neq k}\frac{c(\lambda_j,\lambda_k)}{4}|\langle j|d\rho|k\rangle|^2, \tag{40}$$

where $c(x,y)$ is a Morozova-Chentsov function [40–42]. The corresponding scaled WMSE is given by

$$\mathcal{E}_c(\rho)=\frac{2d^2-d-2}{4(d+2)}+\frac{d}{4(d+2)}\sum_j\frac{1}{\lambda_j}+\frac{d+1}{4(d+2)}\sum_{j\neq k}(1+\lambda_j+\lambda_k)c(\lambda_j,\lambda_k). \tag{41}$$

This equation reduces to Eq. (38) if $c(x,y)=2/(x+y)$, which corresponds to the Bures metric. For the quantum Chernoff metric [53], we have $c(x,y)=4/(\sqrt{x}+\sqrt{y})^2$ and

$$\mathcal{E}_c(\rho)=\frac{2d^2-d-2}{4(d+2)}+\frac{d}{4(d+2)}\sum_j\frac{1}{\lambda_j}+\frac{d+1}{d+2}\sum_{j\neq k}\frac{1+\lambda_j+\lambda_k}{(\sqrt{\lambda_j}+\sqrt{\lambda_k})^2}. \tag{42}$$

C. Efficiency of the covariant measurement with optimal reconstruction

Now let us turn to the optimal state reconstruction based on the covariant measurement. According to Eq. (6), the frame superoperator is given by

$$\mathcal{F}(\rho)=d\int d\mu(\psi)\frac{1}{\langle\psi|\rho|\psi\rangle}(|\Pi_\psi\rangle\!\rangle\langle\!\langle\Pi_\psi|), \tag{43}$$

where $\Pi_\psi=|\psi\rangle\langle\psi|$. In general, it is not easy to derive an explicit formula for $\mathcal{F}(\rho)$. To understand its state dependence, it is instructive to consider those states that are convex combinations of the completely mixed state and a projector state of rank r,

$$\rho_r(s)=\frac{s}{r}\sum_{j=1}^r|j\rangle\langle j|+(1-s)\frac{1}{d},\qquad 1\leqslant r\leqslant d-1,\ 0\leqslant s\leqslant 1. \tag{44}$$

Note, however, that we do not assume this knowledge in state reconstruction. In this case, $\mathcal{F}(\rho_r(s))$ has the form

$$\mathcal{F}(\rho_r(s))=a\mathcal{P}_1+b\mathcal{P}_2+c\mathcal{P}_3+\sum_{j,k}\mathcal{M}_{jk}|E_{jj}\rangle\!\rangle\langle\!\langle E_{kk}|, \tag{45}$$

where $\mathcal{P}_1,\mathcal{P}_2,\mathcal{P}_3$ are projectors,

$$\mathcal{P}_1=\sum_{j\neq k=1}^r|E_{jk}\rangle\!\rangle\langle\!\langle E_{jk}|,\quad \mathcal{P}_3=\sum_{j\neq k=r+1}^d|E_{jk}\rangle\!\rangle\langle\!\langle E_{jk}|,$$

$$\mathcal{P}_2=\sum_{j=1}^r\sum_{k=r+1}^d(|E_{jk}\rangle\!\rangle\langle\!\langle E_{jk}|+|E_{kj}\rangle\!\rangle\langle\!\langle E_{kj}|), \tag{46}$$

and

$$\mathcal{M}_{jk}=\begin{cases}(1+\delta_{jk})a & \text{if}\quad 1\leqslant j,k\leqslant r,\\ (1+\delta_{jk})c & \text{if}\quad r+1\leqslant j,k\leqslant d,\\ b & \text{otherwise.}\end{cases} \tag{47}$$

The three parameters a,b, and c are determined by the formulas $a=g_{20}$, $b=g_{11}$, and $c=g_{02}$, where

$$g_{jk}=\frac{2dr\Gamma(d+1)}{\Gamma(r+j)\Gamma(d-r+k)}\times\int_0^{\pi/2}d\alpha\frac{(\cos\alpha)^{2r-1+2j}(\sin\alpha)^{2d-2r-1+2k}}{ds(\cos\alpha)^2+r(1-s)}, \tag{48}$$

which can be evaluated by applying the formula

$$\int_0^{\pi/2}d\alpha\frac{\cos\alpha(\sin\alpha)^{2m+1}}{(\cos\alpha)^2+u}=\frac{1}{2}(1+u)^m\ln\frac{1+u}{u}-\frac{1}{2}\sum_{n=0}^{m-1}\frac{(1+u)^n}{m-n},\quad u>0 \tag{49}$$

after replacing $(\cos\alpha)^2$ with $1-(\sin\alpha)^2$. The Fisher information matrix $\bar{\mathcal{F}}(\rho_r(s))$ has the same form as $\mathcal{F}(\rho_r(s))$, except that $\mathcal{M}$ is replaced by $\bar{\mathcal{M}}:=\bar{\mathbf{I}}\mathcal{M}\bar{\mathbf{I}}$.

Calculation shows that $\bar{\mathcal{M}}$ has $r-1$ eigenvalues equal to a, $d-r-1$ eigenvalues equal to c, and one eigenvalue equal to

$$\beta=\frac{(r+1)(d-r)a+r(d-r+1)c-2r(d-r)b}{d}. \tag{50}$$

Note that $E_{jk}^{\pm}$ for $j\neq k$ are eigenvectors of $\mathcal{F}$ and $\bar{\mathcal{F}}$, and that the common eigenvalue is one of the three choices a,b,c depending on the values of j and k. We deduce that $\bar{\mathcal{F}}$ has four distinct eigenvalues a,b,c, and β with multiplicities r^2-1, $2r(d-r)$, $(d-r)^2-1$, and 1, respectively (the eigenvalue corresponding to the null eigenvector $|1\rangle\!\rangle$ is excluded here).

According to Eq. (15), the scaled MSE is given by

$$\mathcal{E}(\rho_r(s))=\frac{r^2-1}{a}+\frac{2r(d-r)}{b}+\frac{(d-r)^2-1}{c}+\frac{1}{\beta}. \tag{51}$$

The scaled MSB can be determined by virtue of Eq. (39) with the result

$$\mathcal{E}_{\mathrm{SB}}(\rho_r(s))=\frac{1}{4}\left(\frac{r^2-1}{a\lambda_1}+\frac{4r(d-r)}{b(\lambda_1+\lambda_2)}+\frac{(d-r)^2-1}{c\lambda_2}+\frac{d-r}{d\beta\lambda_1}+\frac{r}{d\beta\lambda_2}\right), \tag{52}$$

where $\lambda_1=(s/r)+(1-s)/d$ and $\lambda_2=(1-s)/d$ are the two distinct eigenvalues of ρ. The scaled WMSEs with respect to other monotone Riemannian metrics can be derived in a similar manner. The volume (with respect to the HS metric) of the scaled uncertainty ellipsoid is given by

$$\mathcal{V}(\rho_r(s))=V_{d^2-1}\left(a^{r^2-1}b^{2r(d-r)}c^{(d-r)^2-1}\beta\right)^{-1/2}, \tag{53}$$

along with its logarithm

$$\ln\mathcal{V}(\rho_r(s))=\ln V_{d^2-1}-\tfrac{1}{2}\{(r^2-1)\ln a+[2r(d-r)]\ln b+[(d-r)^2-1]\ln c+\ln\beta\}. \tag{54}$$

Figure 2 illustrates the scaled MSE and MSB in the case $r=1$ and $d=2,3,4,5,6$. Compared with canonical linear state tomography or minimal state tomography, optimal state estimation with covariant measurements can improve the efficiency significantly when the states of interest have high

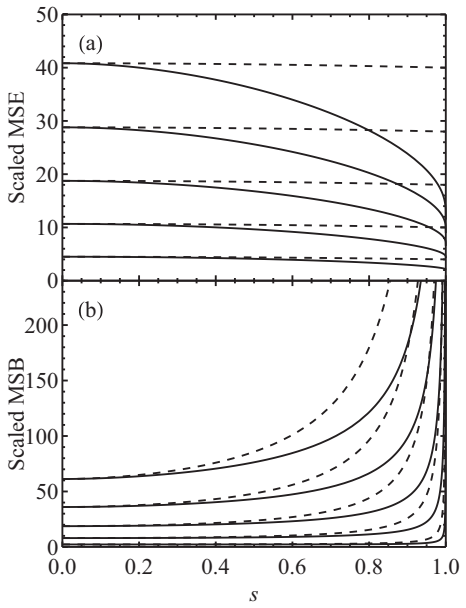

FIG. 2. Tomographic efficiencies of covariant measurements. The true states have the form in Eq. (44) with $r = 1$ and $d = 2,3,\ldots,6$ (from bottom to top). The scaled MSB diverges in the limit $s \to 1$, in which case the states are rank deficient. For comparison, the dashed lines show the performances of covariant measurements under canonical linear reconstruction. In plot (a), they also represent the performances of the optimal minimal IC measurements (that is SIC measurements) with respect to the scaled MSE.

purities. Nevertheless, the efficiency is still too limited to be satisfactory when the scaled MSB is chosen as the figure of merit.

As s approaches 1, the state $\rho_r(s)$ turns into a subnormalized projector of rank r. When $r \geqslant 2$, the three parameters a,b,c have well-defined limits $a = r/(r+1), b = 1, c = r/(r-1)$, and so does the scaled MSE,

$$\mathcal{E}(\rho_r(1)) = d^2 + 2d - 1 - \frac{d^2}{r} - \frac{1}{r}. \tag{55}$$

When $r = 1$, the parameters a and b still have well-defined limits, whereas c diverges as $\ln[d/(1-s)]$. The formula for the scaled MSE is still applicable, except that the derivative of $\mathcal{E}(\rho_r(s))$ with respect to s can diverge. In the pure-state limit, the scaled MSE $2(d-1)$ achieved by the covariant measurement is equal to the corresponding value for the pure-state model [50]. Compared with the scaled MSE $d^2 + d - 2$ [17,26,27] that is achievable with minimal state tomography, it is smaller by $(d+2)/2$ times. Furthermore, it is minimal not only in the Bayesian sense but also in the pointwise sense by saturating a quantum analog of the Cramér-Rao bound; see Ref. [54] as well as Secs. 5.3.3 and 6.2.2 of Ref. [27].

In the pure-state limit, the scaled MSE matrix can be determined based on Eqs. (8) and (45), with the result

$$\mathcal{C}(|1\rangle\langle 1|) = \sum_{j=2}^{d} (|E_{1j}^{+}\rangle\!\rangle\langle\!\langle E_{1j}^{+}| + |E_{1j}^{-}\rangle\!\rangle\langle\!\langle E_{1j}^{-}|). \tag{56}$$

It is a rank-$2(d-1)$ projector, in contrast with the scaled MSE matrix associated with canonical reconstruction, which has full rank in the space of traceless Hermitian operators [see Eq. (35)]. The scaled deviation $\Delta\rho$ has the form

$$\Delta\rho = \sum_{j=2}^{d} (x_j E_{1j}^{+} + y_j E_{1j}^{-}), \tag{57}$$

where x_j, y_j obey a $2(d-1)$-dimensional standard isotropic Gaussian distribution. Since $\Delta\rho$ has only two nonzero eigenvalues $\pm\sqrt{\sum_{j=2}^{d}(x_j^2 + y_j^2)/2}$, its trace norm is proportional to the HS norm, $\|\Delta\rho\|_{\mathrm{tr}} = \|\Delta\rho\|_{\mathrm{HS}}/\sqrt{2}$. The scaled mean errors (not MSE) with respect to the trace distance and the HS distance are given by

$$\mathcal{E}_{\mathrm{tr}}(\rho) = \frac{1}{\sqrt{2}}\mathcal{E}_{\mathrm{HS}}(\rho) = \frac{\Gamma\left(d-\frac{1}{2}\right)}{\Gamma(d-1)} \approx \sqrt{d-1}. \tag{58}$$

Compared with the result achievable with minimal tomography [26,27], the scaled mean trace distance is approximately smaller by a factor of $4d/3\pi$ when $d \gg 2$. Therefore, the efficiency advantage of IOC measurements is more substantial with respect to the mean trace distance in comparison with the MSE. The contrast is even more dramatic with respect to the volume of the scaled uncertainty ellipsoid: the average volume vanishes in the pure-state limit for the covariant measurement but remains finite for any minimal IC measurement or any set of mutually unbiased measurements.

In sharp contrast, the scaled MSB diverges in the limit $s \to 1$. Consequently, with respect to the Bures metric, the volume of the scaled uncertainty ellipsoid also diverges. This seemingly surprising phenomenon can be explained as follows: the entries of $\bar{\mathcal{F}}$ are either finite or logarithmically divergent in this limit, while the entries of the weighting matrix diverge much faster according to Eq. (39). Recalling that the covariant measurement minimizes the average scaled MSB among all nonadaptive measurements, we conclude that the average scaled MSB diverges at the boundary of the state space for all nonadaptive measurements. From the Bayesian perspective, our analysis implies that the MSB generally decreases more slowly than the scaling law $1/N$ expected from common statistical consideration once the prior weight near pure states is non-negligible. For single qubit, this phenomenon was noticed in Ref. [55]. The same conclusion also holds for any WMSE based on a monotone Riemannian metric since the Bures metric is minimal among all such metrics [40–42]. These observations reveal a severe limitation of nonadaptive measurements for quantum state estimation and the importance of exploring more sophisticated strategies, which deserve further study [27].

IV. QUBIT STATE ESTIMATION WITH INFORMATIONALLY OVERCOMPLETE MEASUREMENTS

In this section we exemplify our general approach on IOC measurements with qubit state estimation. Our main goal is to elucidate with this simple example the efficiency limit of IOC measurements and the extent to which they are advantageous over minimal IC measurements with respect

to various figures of merit, such as the MSE, MSB, and the volume of the uncertainty ellipsoid. To be concrete, our discussions focus on the covariant measurement and measurements constructed out of platonic solids inscribed on the Bloch sphere. Nevertheless, our approach applies equally well to other measurements. There are already many studies on this subject [15,20–22], but most theoretical works are based on numerical simulations. We have derived several analytical results on canonical linear state tomography in Ref. [26]. Here we turn to the optimal state reconstruction in comparison with the canonical reconstruction.

A. Canonical reconstruction

Following the convention in Refs. [26,27], to each platonic solid inscribed on the Bloch sphere, we can construct a generalized measurement whose outcomes correspond to the vertices of the platonic solid. Given a platonic solid with n vertices represented by n unit vectors $\boldsymbol{v}_k$, the outcomes of the corresponding measurement are given by $\Pi_k = (1 + \boldsymbol{v}_k \cdot \boldsymbol{\sigma})/n$. Suppose the qubit state ρ is parametrized by the Bloch vector $\boldsymbol{s} = (x, y, z)$; then reconstructing the state ρ is equivalent to reconstructing its Bloch vector $\boldsymbol{s}$.

Under canonical reconstruction, the reconstruction operators take on the form $\Theta_k = (1 + 3\boldsymbol{v}_k \cdot \boldsymbol{\sigma})/2$ according to Eq. (24) since the measurement corresponding to any platonic solid is tight IC. The scaled MSE matrix of the estimator $\hat{s}$ of the Bloch vector has the form [26,27]

$$C(\boldsymbol{s}) = 3 - \boldsymbol{s}\boldsymbol{s} + \frac{9}{n}\sum_{k=1}^{n} (\boldsymbol{v}_k \cdot \boldsymbol{s})\, \boldsymbol{v}_k \boldsymbol{v}_k, \tag{59}$$

where $\boldsymbol{s}\boldsymbol{s}$ is the dyadic composed of the vector $\boldsymbol{s}$ and itself. For any measurement constructed from a platonic solid other than the regular tetrahedron, the last term in the equation vanishes due to symmetry, which yields

$$C^{\mathrm{Iso}}(\boldsymbol{s}) = 3 - \boldsymbol{s}\boldsymbol{s}. \tag{60}$$

More generally, all isotropic measurements [26,27] share the same scaled MSE matrix and are equally efficient under canonical reconstruction. The scaled MSE with respect to the HS distance is equal to

$$\mathcal{E}(\rho) = \frac{1}{2}\,\mathrm{tr}\{C(\boldsymbol{s})\} = \frac{9 - s^2}{2}. \tag{61}$$

Here the factor $1/2$ accounts for the difference between the HS distance and the distance on the Bloch ball. The scaled MSE is independent of the orientation of the Bloch vector, regardless of the platonic solid under consideration, as expected for any rank-one tight IC measurement.

The weighting matrix corresponding to the Bures metric is one-fourth of the quantum Fisher information matrix and takes on the form

$$W(\boldsymbol{s}) = \frac{1}{4} + \frac{\boldsymbol{s}\boldsymbol{s}}{4(1 - s^2)}. \tag{62}$$

The scaled MSB is thus given by

$$\begin{aligned}\mathcal{E}_{\mathrm{SB}}(\rho) &= \mathrm{tr}\{W(\boldsymbol{s})C(\boldsymbol{s})\} \\ &= \frac{9}{4} + \frac{s^2}{2(1 - s^2)} + \frac{9}{4n(1 - s^2)}\sum_k (\boldsymbol{s} \cdot \boldsymbol{v}_k)^3.\end{aligned} \tag{63}$$

Except for the SIC (tetrahedron) measurement, the last term vanishes, and we have

$$\mathcal{E}_{\mathrm{SB}}(\rho) = \frac{9}{4} + \frac{s^2}{2(1 - s^2)}. \tag{64}$$

To derive an explicit formula for the SIC measurement, we assume that the cube (also the octahedron) takes the standard orientation and that the tetrahedron is composed of four vertices of the cube including $(1,1,1)/\sqrt{3}$. In that case,

$$\mathcal{E}_{\mathrm{SB}}^{\mathrm{SIC}}(\rho) = \frac{9}{4} + \frac{s^2 + 3\sqrt{3}xyz}{2(1 - s^2)}. \tag{65}$$

Unlike the scaled MSE, which is unitarily invariant, the scaled MSB for given s is maximized when the Bloch vector of the true state is parallel to one leg of the outcomes and minimized in the opposite situation. The last term in the above equation vanishes after taking average over unitarily equivalent states. Therefore, all measurements constructed from platonic solids are equally efficient with respect to the average scaled MSB under canonical reconstruction. This conclusion is not as obvious as the corresponding statement concerning the scaled MSE.

The volume (with respect to the HS metric) of the scaled uncertainty ellipsoid is given by

$$\mathcal{V}(\rho) = \frac{4\pi}{3}\sqrt{\frac{\det\{C(\boldsymbol{s})\}}{8}}; \tag{66}$$

note that $V_3 = 4\pi/3$. Here the factor $1/8$ accounts for the difference between the HS distance and the distance on the Bloch ball as before. For isotropic measurements, it reduces to

$$\mathcal{V}^{\mathrm{Iso}}(\rho) = \pi\sqrt{2(3 - s^2)}. \tag{67}$$

For the SIC measurement, we have

$$\begin{aligned}\mathcal{V}^{\mathrm{SIC}}(\rho) = \sqrt{\frac{2}{3}}\pi[&2(x^4 + y^4 + z^4) + 8\sqrt{3}xyz \\ &- s^4 - 6s^2 + 9]^{1/2}.\end{aligned} \tag{68}$$

B. Optimal reconstruction

Now let us turn to the optimal reconstruction. In terms of the Bloch vector, the Fisher information matrix takes on the form

$$I(\boldsymbol{s}) = \frac{1}{n}\sum_k \frac{1}{1 + \boldsymbol{v}_k \cdot \boldsymbol{s}}\boldsymbol{v}_k \boldsymbol{v}_k. \tag{69}$$

The scaled MSE matrix $C(\boldsymbol{s})$ is the inverse of $I(\boldsymbol{s})$. For the SIC measurement, it is still given by Eq. (59). For the MUB measurement, we have

$$C(\boldsymbol{s}) = 3\ \mathrm{diag}(1 - x^2, 1 - y^2, 1 - z^2). \tag{70}$$

It is smaller than the scaled MSE matrix $3 - \boldsymbol{s}\boldsymbol{s}$ under the canonical reconstruction [cf. Eq. (60)], but is no longer invariant under unitary transformations of the measurement outcomes. The differences between the two reconstruction methods are clearly reflected in the uncertainty ellipses, as

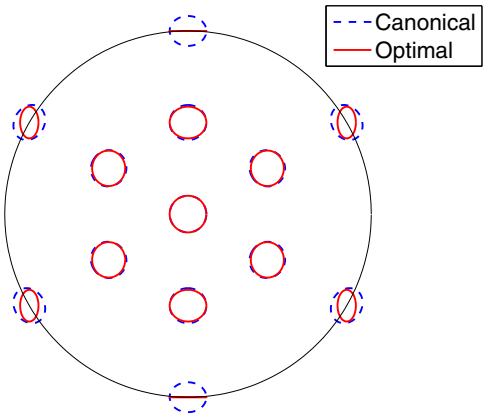

FIG. 3. (Color online) Uncertainty ellipses of the canonical reconstruction and the optimal reconstruction. The uncertainty ellipses are associated with the marginal distributions on the x-z plane of the Bloch ball resulting from mutually unbiased measurements on a family of states, each repeated 300 times. The optimal reconstruction reduces the sizes of the uncertainty ellipses at the prize of losing the covariance property.

illustrated in Fig. 3. The situations are quite similar for measurements constructed from platonic solids other than the tetrahedron, although the expressions of $C(\boldsymbol{s})$ can be much more complicated.

The scaled MSEs of the measurements constructed from the tetrahedron, octahedron, and cube are respectively given by

$$\mathcal{E}^{\mathrm{SIC}}(\rho) = \frac{9-s^2}{2}, \quad \mathcal{E}^{\mathrm{MUB}}(\rho) = \frac{3(3-s^2)}{2},$$
$$\mathcal{E}^{\mathrm{Cube}}(\rho) = \frac{27-18s^2+s^4+2(x^4+y^4+z^4)}{2(3-s^2)}. \tag{71}$$

The scaled MSE is unitarily invariant for the SIC (tetrahedron) measurement and the MUB (octahedron) measurement, as mentioned in Sec. II C. This is not the case for the cube measurement, although it is a combination of two tetrahedron measurements and is seemingly more symmetric than a single tetrahedron measurement. For given s, the minimal scaled MSE $(9-s^2)(9-5s^2)/6(3-s^2)$ is attained when $\boldsymbol{s}$ is parallel to one of the diagonals of the cube, and the maximum $3(3-s^2)/2$ is attained when $\boldsymbol{s}$ is parallel to one of the axes. The average is

$$\overline{\mathcal{E}^{\mathrm{Cube}}(\rho)} = \frac{135-90s^2+11s^4}{10(3-s^2)}. \tag{72}$$

The formulas for the MSEs of the dodecahedron measurement and icosahedron measurement are too complicated to convey a clear meaning; suffice it to mention that the MSEs are not unitarily invariant in both cases, as in the case of the cube measurement. This observation reveals an intriguing feature that seems to be unique to SIC and MUB measurements, which deserves further study.

The scaled MSB for the SIC measurement is still given by Eq. (65). For the MUB and cube measurements, we have

$$\mathcal{E}_{\mathrm{SB}}^{\mathrm{MUB}}(\rho) = \frac{3(3-s^2)}{4} + \frac{3(s^2-x^4-y^4-z^4)}{4(1-s^2)},$$
$$\mathcal{E}_{\mathrm{SB}}^{\mathrm{Cube}}(\rho) = \frac{27-27s^2-2s^4}{12(1-s^2)} + \frac{6(x^4+y^4+z^4)-2(x^6+y^6+z^6)-21x^2y^2z^2}{3(3-s^2)(1-s^2)}. \tag{73}$$

Taking average over unitarily equivalent states yields

$$\overline{\mathcal{E}_{\mathrm{SB}}^{\mathrm{SIC}}(\rho)} = \frac{9}{4} + \frac{s^2}{2(1-s^2)}, \quad \overline{\mathcal{E}_{\mathrm{SB}}^{\mathrm{MUB}}(\rho)} = \frac{9}{4} + \frac{3s^4}{10(1-s^2)},$$
$$\overline{\mathcal{E}_{\mathrm{SB}}^{\mathrm{Cube}}(\rho)} = \frac{945-1260s^2+413s^4-26s^6}{140(3-s^2)(1-s^2)}. \tag{74}$$

The volume of the scaled uncertainty ellipsoid of the SIC measurement is still determined by Eq. (68). For MUB and cube measurements, they are respectively given by

$$\mathcal{V}^{\mathrm{MUB}}(\rho) = \pi\sqrt{6(1-x^2)(1-y^2)(1-z^2)},$$
$$\mathcal{V}^{\mathrm{Cube}}(\rho) = \frac{\pi}{3}\sqrt{\frac{2[3-(x+y-z)^2][3-(x-y+z)^2]}{3-s^2}} \times \sqrt{[3-(-x+y+z)^2][3-(x+y+z)^2]}. \tag{75}$$

They are all equal to $\sqrt{6}\pi$ when $s=0$. The averages of the log volumes over unitarily equivalent states read

$$\overline{\ln \mathcal{V}^{\mathrm{MUB}}(\rho)} = \ln(\sqrt{6}\pi) - 3 + \frac{3}{2}\left[\ln(1-s^2) + \frac{1}{s}\ln\frac{1+s}{1-s}\right],$$
$$\overline{\ln \mathcal{V}^{\mathrm{Cube}}(\rho)} = \ln(3\sqrt{2}\pi) - 4 + \ln\frac{(1-s^2)^2}{\sqrt{3-s^2}} + \frac{2}{s}\ln\frac{1+s}{1-s}. \tag{76}$$

For the SIC measurement, this average can be determined by numerical integration.

For the covariant measurement, the parameters b in Eq. (45) and β in Eq. (50) are now given by

$$b = \frac{2s-(1-s^2)\ln\left(\frac{1+s}{1-s}\right)}{2s^3}, \quad \beta = \frac{-2s+\ln\left(\frac{1+s}{1-s}\right)}{s^3}. \tag{77}$$

Note that the parameters a and c are irrelevant here. The Fisher information matrix takes on the form

$$\bar{\mathcal{F}}(\rho) = b\bar{\mathbf{I}} + \tfrac{1}{2}(\beta-b)|\tilde{\boldsymbol{s}}\cdot\sigma\rangle\rangle\langle\langle\tilde{\boldsymbol{s}}\cdot\sigma|, \tag{78}$$

where $\tilde{\boldsymbol{s}} = \boldsymbol{s}/s$ is the normalized Bloch vector (the ambiguity at $s=0$ does not matter since $b=\beta$ in that case). In terms of

the Bloch vector, it simplifies to

$$I(\boldsymbol{s}) = \tfrac{1}{2}[b + (\beta - b)\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}]. \tag{79}$$

Its inverse is the scaled MSE matrix associated with the optimal reconstruction,

$$C(\boldsymbol{s}) = 2\left[\frac{1}{b} + \left(\frac{1}{\beta} - \frac{1}{b}\right)\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}\right]. \tag{80}$$

The scaled MSE, MSB, and the volume of the scaled uncertainty ellipsoid (with respect to the HS metric) follow from Eqs. (51)–(53), respectively,

$$\mathcal{E}(\rho) = \frac{2}{b} + \frac{1}{\beta}, \quad \mathcal{E}_{\mathrm{SB}}(\rho) = \frac{1}{b} + \frac{1}{2\beta(1-s^2)},$$
$$\mathcal{V}(\rho) = \frac{4\pi}{3b\sqrt{\beta}}. \tag{81}$$

Similarly, the WMSE with respect to the monotone Riemannian metric characterized by the Morozova-Chentsov function $c(x,y)$ is given by

$$\mathcal{E}_c(\rho) = \frac{c(\lambda_+,\lambda_-)}{2b} + \frac{1}{2\beta(1-s^2)}, \tag{82}$$

where $\lambda_\pm = (1 \pm s)/2$ are the eigenvalues of ρ. For the Chernoff metric $c(x,y) = 4/(\sqrt{x} + \sqrt{y})^2$, it reduces to

$$\mathcal{E}_c(\rho) = \frac{2}{b(1+\sqrt{1-s^2})} + \frac{1}{2\beta(1-s^2)}. \tag{83}$$

As comparison, in canonical linear tomography with the covariant measurement, the scaled MSE matrix $C(\boldsymbol{s})$ is equal to $3 - \boldsymbol{s}\boldsymbol{s}$ as in Eq. (60) since the covariant measurement is an isotropic measurement. Accordingly, we have

$$\mathcal{E}(\rho) = \frac{9-s^2}{2}, \quad \mathcal{E}_{\mathrm{SB}}(\rho) = \frac{3}{2} + \frac{3-s^2}{4(1-s^2)},$$
$$\mathcal{V}(\rho) = \pi\sqrt{2(3-s^2)}, \quad \mathcal{E}_c(\rho) = \frac{3}{4}c(\lambda_+,\lambda_-) + \frac{3-s^2}{4(1-s^2)}. \tag{84}$$

Figure 4 shows the tomographic performances of SIC, MUB, cube, and covariant measurements in qubit state estimation with respect to the average scaled MSE, MSB, and log volume of the scaled uncertainty ellipsoid. For all three figures of merit, the tomographic efficiencies of the four measurement schemes are monotonic increasing with the number of outcomes, the more so the higher the purities of the states of interest. By contrast, in canonical linear state tomography, MUB, cube, and covariant measurements are as efficient as the SIC measurement with respect to the average scaled MSE and MSB, and even less efficient with respect to the average log volume. Comparison with the scaled MSE achieved by the optimal adaptive strategy [16,27,56–58] shows that under the optimal reconstruction the covariant measurement is almost optimal in the pointwise sense. However, it should be noted that this is generally not the case with respect to other figures of merit, such as the scaled MSB. Also, the situation can be very different beyond the two-level system (see Chap. 5 in Ref. [27]). Actually, the scaled MSB diverges in the pure-state limit for the covariant measurement, although it is the most efficient among all nonadaptive measurements. The same is true for any WMSE based on a monotone Riemannian metric, as explained in Sec. III.

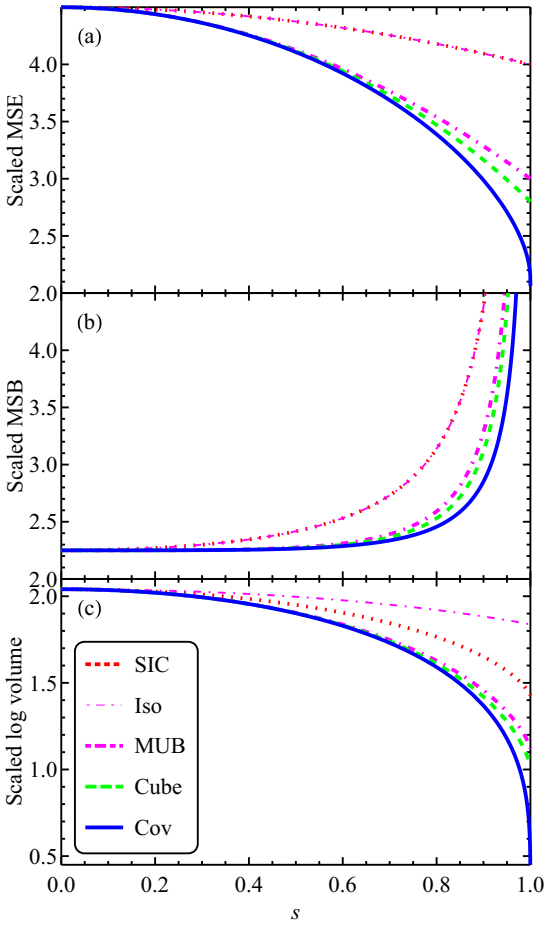

FIG. 4. (Color online) Tomographic efficiencies of the SIC, MUB, cube, and covariant measurements in qubit state estimation. (a) Average scaled MSE; (b) average scaled MSB; (c) average log volume (with respect to the HS metric) of the scaled uncertainty ellipsoid. The average scaled MSBs diverge in the pure-state limit for all four measurements. For the covariant measurement, the log volume diverges to $-\infty$ (that is, the volume vanishes) in the pure-state limit. As comparison, the curve "Iso" shows the common performance of isotropic measurements (including MUB, cube, and covariant measurements) under canonical linear reconstruction. It coincides with the curve "SIC" in plots (a) and (b) since under canonical linear reconstruction isotropic measurements are as efficient as the SIC measurement in qubit state estimation with respect to the average scaled MSE and MSB.

V. SUMMARY

We have studied quantum state estimation with IOC measurements, motivated by the questions of whether and to what extent IOC measurements can improve the tomographic efficiency over minimal IC measurements. To answer these questions and to make fair comparison among various

measurement schemes, we derived the best linear unbiased estimator and showed that it is as efficient as the maximum likelihood estimator in the large sample limit. This estimator may significantly outperform the canonical linear estimator when the states of interest have high purities. This finding is useful not only for studying IOC measurements but also for exploring experimental designs and adaptive quantum state estimation.

Based on the above framework, we showed that the covariant measurement is optimal among all nonadaptive measurements in minimizing the average WMSE based on any unitarily invariant distance, including the MSE and the MSB, as well as the average log volume of the uncertainty ellipsoid. When the states of interest have high purities, IOC measurements can improve the tomographic efficiency significantly and even change the scaling of the cost function with the dimension of the Hilbert space. Nevertheless, the efficiency is still too limited to be satisfactory with respect to the MSB or the WMSE based on any other monotone Riemannian metric as long as the measurement is nonadaptive. On the one hand, our study clarifies the tomographic significance of IOC measurements compared with minimal IC measurements. On the other hand, it pinpoints the limitation of nonadaptive measurements and motivates the study of more sophisticated estimation strategies based on adaptive measurements and collective measurements [27], which deserve further study. In this paper, we only consider ideal measurements. It would be desirable in the future to extend the current work to incorporate imperfection, such as detector inefficiency.

ACKNOWLEDGMENTS

The author is grateful to Berthold-Georg Englert, Masahito Hayashi, and Yong Siah Teo for stimulating discussions and comments on early versions of the manuscript and to Hai Wang for comments on the proof of Lemma 1. The author is also grateful to the referee for comments and suggestions that improve the clarity of the paper. This work is supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation. In the early stage, it was supported by NUS Graduate School (NGS) for Integrative Sciences and Engineering and Centre for Quantum Technologies, which is a Research Centre of Excellence funded by the Ministry of Education and National Research Foundation of Singapore. Figures 2 and 4 have been created using the LevelScheme scientific figure preparation system [59].

APPENDIX A: PROOF OF LEMMA 1

The idea of the proof follows from the proof of Lemma 5.1 in Chap. VI of Ref. [60]. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two $m \times 1$ vectors such that $\boldsymbol{v}$ belongs to the support of $B^\dagger$. Let $\boldsymbol{a} = A^\dagger \boldsymbol{u}$ and $\boldsymbol{b} = B^\dagger \boldsymbol{v}$; then we have

$$\begin{aligned} &\boldsymbol{a}^\dagger \boldsymbol{a} = \boldsymbol{u}^\dagger A A^\dagger \boldsymbol{u}, \quad \boldsymbol{b}^\dagger \boldsymbol{b} = \boldsymbol{v}^\dagger B B^\dagger \boldsymbol{v}, \\ &\boldsymbol{a}^\dagger \boldsymbol{b} = \boldsymbol{u}^\dagger A B^\dagger \boldsymbol{v} = \boldsymbol{u}^\dagger \boldsymbol{v}. \end{aligned} \tag{A1}$$

The Cauchy inequality applied to the equation yields

$$(\boldsymbol{u}^\dagger A A^\dagger \boldsymbol{u})(\boldsymbol{v}^\dagger B B^\dagger \boldsymbol{v}) \geqslant (\boldsymbol{u}^\dagger \boldsymbol{v})^2. \tag{A2}$$

Setting $\boldsymbol{v} = (BB^\dagger)^+ \boldsymbol{u}$ gives rise to

$$\boldsymbol{u}^\dagger A A^\dagger \boldsymbol{u} \geqslant \boldsymbol{u}^\dagger (BB^\dagger)^+ \boldsymbol{u}, \tag{A3}$$

which implies that $AA^\dagger \geqslant (BB^\dagger)^+$. Necessary conditions for saturating the inequality are $A^\dagger \boldsymbol{u} \propto B^\dagger (BB^\dagger)^+ \boldsymbol{u}$ and $|A^\dagger \boldsymbol{u}| = |B^\dagger (BB^\dagger)^+ \boldsymbol{u}|$ for arbitrary $\boldsymbol{u}$; that is, $A^\dagger \propto B^\dagger (BB^\dagger)^+$ and $A \propto (BB^\dagger)^+ B$. Since $AB^\dagger$ is a projector by assumption, it follows that $A = (BB^\dagger)^+ B$, which happens to be the pseudoinverse of $B^\dagger$ [45]. Now the inequality is indeed saturated.

If $AB^\dagger = 1$, then $(BB^\dagger)$ is invertible. The second part of the lemma follows from the fact that $(BB^\dagger)^+ = (BB^\dagger)^{-1}$.

APPENDIX B: FISHER INFORMATION AND CRAMÉR-RAO BOUND

Fisher information [34] and the Cramér-Rao bound [48,49] are two basic ingredients in statistical inference: the former quantifies the amount of information yielded by an observation or a measurement concerning certain parameters of interest, and the latter quantifies the minimal error in estimating these parameters.

Consider a family of probability distributions $p(\xi|\theta)$ parametrized by θ. Our task is to estimate the value of θ as accurately as possible based on the measurement outcomes. Given an outcome ξ, the function $p(\xi|\theta)$ of θ is called the *likelihood function*. The *score* is defined as the partial derivative of the log-likelihood function with respect to θ and reflects the sensitivity of the log-likelihood function with respect to the variation of θ. Its first moment is zero, and the second moment is known as the *Fisher information* [34,61],

$$\begin{aligned} I(\theta) &= \mathrm{Var}\left(\frac{\partial \ln p(\xi|\theta)}{\partial \theta}\right) = \sum_\xi p(\xi|\theta)\left(\frac{\partial \ln p(\xi|\theta)}{\partial \theta}\right)^2 \\ &= \sum_\xi \frac{1}{p(\xi|\theta)}\left(\frac{\partial p(\xi|\theta)}{\partial \theta}\right)^2. \end{aligned} \tag{B1}$$

The Fisher information represents the average sensitivity of the log-likelihood function with respect to the variation of θ. Intuitively, the larger the Fisher information, the better we can estimate the value of the parameter θ.

An estimator $\hat{\theta}(\xi)$ of the parameter θ is *unbiased* if its expectation value is equal to the true parameter; that is,

$$\sum_\xi p(\xi|\theta)[\hat{\theta}(\xi) - \theta] = 0. \tag{B2}$$

Taking the derivative with respect to θ and applying the Cauchy-Schwarz inequality [using the fact that $\sum_\xi p(\xi|\theta) = 1$] yields the well-known *Cramér-Rao bound*

$$C(\theta) = \mathrm{Var}(\hat{\theta}) \geqslant \frac{1}{I(\theta)}, \tag{B3}$$

which states that the MSE or variance of any unbiased estimator is bounded from below by the inverse of the Fisher information [48,49].

In the multiparameter setting, the Fisher information and MSE take on matrix forms,

$$I_{jk}(\theta) = E\left[\left(\frac{\partial \ln p(\xi|\theta)}{\partial\theta_j}\right)\left(\frac{\partial \ln p(\xi|\theta)}{\partial\theta_k}\right)\right],$$
$$C_{jk}(\theta) = E[(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)]. \tag{B4}$$

Accordingly, the Cramér-Rao bound for any unbiased estimator turns out to be a matrix inequality,

$$C(\theta) \geqslant I^{-1}(\theta). \tag{B5}$$

Since the likelihood function is multiplicative, the Fisher information matrix is additive; that is, the total Fisher information matrix of independent measurements is equal to the sum of the respective Fisher information matrices of individual measurements. In particular, the Fisher information matrix of N identical and independent measurements is N times that of one measurement. Accordingly, the MSE matrix of any unbiased estimator based on N measurements satisfies the inequality $C^{(N)}(\theta) \geqslant 1/NI(\theta)$. Thanks to Fisher's theorem [33,34], the lower bound can be saturated asymptotically with the ML estimator under very general assumptions [46]. In the large-sample scenario, the *scaled MSE matrix* $NC^{(N)}(\theta)$ is generally independent of the sample size. It is also denoted by $C(\theta)$ when there is no confusion.

In quantum state estimation, we are interested in the parameters that characterize the state $\rho(\theta)$ of a quantum system. To estimate the values of these parameters, we may perform generalized measurements. Given a measurement Π with outcomes Π_ξ, the probability of obtaining the outcome ξ is $p(\xi|\theta) = \mathrm{tr}\{\rho(\theta)\Pi_\xi\}$. The corresponding Fisher information matrix $I_{jk}(\Pi,\theta)$ is given by

$$I_{jk}(\Pi,\theta) = \sum_\xi \frac{1}{p(\xi|\theta)} \mathrm{tr}\left\{\frac{\partial\rho(\theta)}{\partial\theta_j}\Pi_\xi\right\} \mathrm{tr}\left\{\frac{\partial\rho(\theta)}{\partial\theta_k}\Pi_\xi\right\}. \tag{B6}$$

Once a measurement is chosen, the inverse Fisher information matrix sets a lower bound for the MSE matrix of any unbiased estimator, which can be saturated asymptotically by the ML estimator, as in the case of classical parameter estimation. It should be noted that the bound depends on the specific measurement.

In practice, it is often more convenient to use a single number rather than a matrix to quantify the error. A common choice is the scaled MSE $\mathrm{tr}\{C(\theta)\}$; a more general alternative is the scaled WMSE $\mathrm{tr}\{W(\theta)C(\theta)\}$, where $W(\theta)$ is a positive semidefinite weighting matrix, which may depend on θ. The Cramér-Rao bound implies that $\mathrm{tr}\{W(\theta)C(\theta)\} \geqslant \mathrm{tr}\{W(\theta)I^{-1}(\theta)\}$; again, this bound can be saturated asymptotically with the ML estimator. A drawback with the MSE is that it depends on the parametrization, which is somehow arbitrary. With a suitable choice of the weighting matrix, the WMSE is free from this problem. For example, the WMSEs with respect to the HS distance and Bures distance are parametrization independent. Except when stated otherwise, the MSE concerned in the main text is defined with respect to the HS distance.

APPENDIX C: MAXIMUM-LIKELIHOOD ESTIMATION

In ML estimation, instead of searching for a state that matches the observed frequencies, we seek a state that maximizes the likelihood function (or functional). The principle of ML was proposed by Fisher [33] in the 1920s and has become a basic ingredient in statistical inference. During the past decade, it has found extensive applications in quantum state estimation [1,2,35,62,63]. In addition, it is useful for entanglement detection [64] and characterization [65].

In quantum state estimation, the *likelihood functional* is defined as [1,35]

$$\mathcal{L}(\rho) = \prod_\xi p_\xi^{n_\xi}, \tag{C1}$$

where $p_\xi = \mathrm{tr}(\rho\Pi_\xi)$ and n_ξ are the probability and the number of times of obtaining the outcome ξ given N measurements on the state ρ. In practice, it is often more convenient to work with the *log-likelihood functional*

$$\ln\mathcal{L}(\rho) = \sum_\xi n_\xi \ln p_\xi = N\sum_\xi f_\xi \ln p_\xi. \tag{C2}$$

The ML method consists in choosing a state $\hat{\rho}_{\mathrm{ML}}$ that maximizes the likelihood functional or, equivalently, the log-likelihood functional, as an estimator of the true state [1,2,35,62,63]. If there exists a state that matches the observed frequencies, then the state is also an MLE. This conclusion is an immediate consequence of the inequality

$$\sum_\xi f_\xi \ln p_\xi \leqslant \sum_\xi f_\xi \ln f_\xi. \tag{C3}$$

In general, it is not easy to find a closed formula for the MLE. Fortunately, the estimator can be computed efficiently with an algorithm proposed by Hradil [35].

[1] *Quantum State Estimation*, edited by M. G. A. Paris and J. Řeháček, Lecture Notes in Physics Vol. 649 (Springer, Berlin, 2004).

[2] A. I. Lvovsky and M. G. Raymer, Rev. Mod. Phys. **81**, 299 (2009).

[3] E. Prugovečki, Int. J. Theor. Phys. **16**, 321 (1977).

[4] P. Busch, Int. J. Theor. Phys. **30**, 1217 (1991).

[5] G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, J. Opt. B: Quantum Semiclass. Opt. **6**, S487 (2004).

[6] G. Zauner, Int. J. Quant. Inf. **09**, 445 (2011).

[7] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, J. Math. Phys. **45**, 2171 (2004); supplementary information including the fiducial kets available at http://www.cquic.org/papers/reports/

[8] A. J. Scott and M. Grassl, J. Math. Phys. **51**, 042203 (2010); supplementary information including the fiducial kets available at arXiv:0910.5784.

[9] D. M. Appleby, C. A. Fuchs, and H. Zhu, Quantum Inf. Comput. (to be published), arXiv:1312.0555.

[10] I. D. Ivanović, J. Phys. A: Math. Gen. **14**, 3241 (1981).

[11] W. K. Wootters and B. D. Fields, Ann. Phys. (NY) **191**, 363 (1989).
[12] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, Int. J. Quant. Inf. **08**, 535 (2010).
[13] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
[14] J. Řeháček and Z. Hradil, Phys. Rev. Lett. **88**, 130401 (2002).
[15] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Phys. Rev. A **70**, 052321 (2004).
[16] F. Embacher and H. Narnhofer, Ann. Phys. (NY) **311**, 220 (2004).
[17] A. J. Scott, J. Phys. A: Math. Gen. **39**, 13507 (2006).
[18] A. Roy and A. J. Scott, J. Math. Phys. **48**, 072110 (2007).
[19] J. Du, M. Sun, X. Peng, and T. Durt, Phys. Rev. A **74**, 042341 (2006).
[20] A. Ling, K. P. Soh, A. Lamas-Linares, and C. Kurtsiefer, Phys. Rev. A **74**, 022309 (2006).
[21] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, Phys. Rev. A **78**, 052122 (2008).
[22] A. Ling, A. Lamas-Linares, and C. Kurtsiefer, "Accuracy of minimal and optimal qubit tomography for finite-length experiments," available at arXiv:0807.0991.
[23] R. B. A. Adamson and A. M. Steinberg, Phys. Rev. Lett. **105**, 030406 (2010).
[24] T. Baier and D. Petz, Rep. Math. Phys. **65**, 203 (2010).
[25] Y. S. Teo, H. Zhu, and B.-G. Englert, Opt. Commun. **283**, 724 (2010).
[26] H. Zhu and B.-G. Englert, Phys. Rev. A **84**, 022327 (2011).
[27] H. Zhu, Ph.D. thesis, National University of Singapore, 2012, available at http://scholarbank.nus.edu.sg/bitstream/handle/10635/35247/ZhuHJthesis.pdf
[28] D. Petz and L. Ruppert, Rep. Math. Phys. **69**, 161 (2012).
[29] T. Tasnádi, "Maximal qubit tomography," available at arXiv:0803.1946.
[30] G. M. D'Ariano and P. Perinotti, Phys. Rev. Lett. **98**, 020403 (2007).
[31] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, Phys. Rev. Lett. **102**, 010404 (2009).
[32] A. Bisio, G. Chiribella, G. D'Ariano, S. Facchini, and P. Perinotti, IEEE J. Sel. Top. Quant. Electron. **15**, 1646 (2009).
[33] R. A. Fisher, Philos. Trans. R. Soc. London A **222**, 309 (1922).
[34] R. A. Fisher, Math. Proc. Cambr. Philos. Soc. **22**, 700 (1925).
[35] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
[36] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).
[37] R. Blume-Kohout, "Robust error bars for quantum tomography," arXiv:1202.5270.
[38] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert, New J. Phys. **15**, 123026 (2013).
[39] H. Zhu, "Information complementarity: A new paradigm for decoding quantum incompatibility," arXiv:1406.6898.
[40] D. Petz, Linear Algebra Appl. **244**, 81 (1996).
[41] D. Petz and C. Sudár, J. Math. Phys. **37**, 2662 (1996).
[42] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, UK, 2006).
[43] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
[44] G. M. D'Ariano, P. Lo Presti, and M. F. Sacchi, Phys. Lett. A **272**, 32 (2000).
[45] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory* (Princeton University Press, Princeton, NJ, 2005).
[46] A. van den Bos, *Parameter Estimation for Scientists and Engineers* (John Wiley & Sons, Hoboken, NJ, 2007).
[47] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil, Phys. Rev. Lett. **107**, 020404 (2011).
[48] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).
[49] C. R. Rao, Bull. Calcutta Math. Soc. **37**, 81 (1945).
[50] M. Hayashi, J. Phys. A: Math. Gen. **31**, 4633 (1998).
[51] R. Bhatia, *Matrix Analysis* (Springer, New York, 1997).
[52] M. Hübner, Phys. Lett. A **163**, 239 (1992).
[53] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete, Phys. Rev. Lett. **98**, 160501 (2007).
[54] K. Matsumoto, J. Phys. A: Math. Gen. **35**, 3111 (2002).
[55] E. Bagan, M. A. Ballester, R. D. Gill, R. Muñoz-Tapia, and O. Romero-Isart, Phys. Rev. Lett. **97**, 130501 (2006).
[56] M. Hayashi, in *Quantum Communication, Computing, and Measurement*, edited by O. Hirota, A. S. Holevo, and C. A. Caves (Plenum, New York, 1997), reprinted in Ref. [66].
[57] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[58] M. Hayashi and K. Matsumoto, J. Math. Phys. **49**, 102101 (2008).
[59] M. A. Caprio, Comput. Phys. Commun. **171**, 107 (2005), http://scidraw.nd.edu/levelscheme
[60] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[61] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer, New York, 1998).
[62] J. Řeháček, Z. Hradil, and M. Ježek, Phys. Rev. A **63**, 040303(R) (2001).
[63] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Phys. Rev. A **75**, 042108 (2007).
[64] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk, Phys. Rev. Lett. **105**, 170501 (2010).
[65] L. Chen, H. Zhu, and T.-C. Wei, Phys. Rev. A **83**, 012305 (2011).
[66] *Asymptotic Theory of Quantum Statistical Inference*, edited by M. Hayashi (World Scientific, Singapore, 2005).