

Higher-order corrections to the quantum Cramér-Rao bound

J. R. Hervas¹, A. Z. Goldberg², A. S. Sanz¹, Z. Hradil³, J. Řeháček³, and L. L. Sánchez-Soto^{1,4,5}

¹*Departamento de Óptica, Facultad de Física, Universidad Complutense, 28040 Madrid, Spain*

²*National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1N 5A2, Canada*

³*Department of Optics, Palacký University, 17. listopadu 12, 771 46 Olomouc, Czech Republic*

⁴*Institute for Quantum Studies, Chapman University, Orange, California 92866, USA*

⁵*Max-Planck-Institut für die Physik des Lichts, 91058 Erlangen, Germany*



(Received 8 May 2025; accepted 29 July 2025; published 20 August 2025)

Quantum Fisher information and the associated quantum Cramér-Rao bound (QCRB) are fundamental tools in frequentist quantum metrology, offering both analytical simplicity and practical precision limits for parameter estimation. The QCRB sets a lower bound on the mean square error (MSE) in the idealized limit of infinite measurement trials ($\nu \rightarrow \infty$). Here we perform a systematic expansion in powers of $1/\nu$ to refine MSE estimates in realistic, finite-resource scenarios. These corrections reveal differences between measurements that appear equally optimal under the QCRB. They also help to distinguish among multiple optimal state families for estimating an unknown unitary transformation. Additionally, we explore the Bhattacharyya bound and its quantum counterpart, which constrain these corrections. Our results are relevant for preasymptotic metrology, enabling optimized protocols with limited resources without reliance on numerical simulations.

DOI: [10.1103/PhysRevA.112.022426](https://doi.org/10.1103/PhysRevA.112.022426)

I. INTRODUCTION

Metrology is the science of devising methods to extract the most precise possible estimates of quantities in nature. The ultimate limits in this endeavor are governed by the laws of physics, with quantum mechanics being the most fundamental theoretical framework. By harnessing uniquely quantum phenomena, such as entanglement and squeezing, quantum metrology surpasses the precision of classical methods, making it an ever-growing field of research [1–9].

Parameter estimation has traditionally been approached from two primary perspectives: Frequentist and Bayesian [10]. Rooted in different interpretations of probability, these frameworks provide different insights into parameter estimates and associated uncertainties. While they converge in the asymptotic limit of a large number of measurements, this has led to the mistaken belief that they are interchangeable in quantum metrology. In reality, they embody fundamentally different paradigms.

Our main interest here is in the frequentist scheme, where the quantum Cramér-Rao bound (QCRB) plays a pivotal role, establishing a fundamental lower bound on the mean square error (MSE) of any estimator. It is formulated in terms of the quantum Fisher information (QFI), which quantifies the sensitivity of a quantum state to infinitesimal changes in the parameter, thereby dictating the ultimate precision achievable [11].

The QCRB is widely employed for its elegant blend of analytical tractability, theoretical significance, and practical relevance. It not only drives theoretical advancements but also provides insights into the fundamental limits of estimation precision in real-world experiments.

Nevertheless, despite its prevalence in quantum information, the implications of the QCRB must be taken with a grain of salt. Notably, it is only asymptotically attainable in the limit of infinitely many measurements $\nu \rightarrow \infty$. The performance with limited trials remains largely *terra incognita*, although several intriguing results have recently emerged [12–15].

Another challenge is that the QFI is designed for local estimation rather than global scenarios, where both the location and size of the parameter neighborhood are unknown. It has been realized that the MSE in global estimation can differ significantly from that in local strategies [16–18].

Moreover, while the QCRB indicates that concentrating all resources on a single measurement is the optimal strategy, this approach paradoxically requires increasingly precise prior knowledge of the parameter as resources increase. The scaling predicted by the QCRB is corroborated by a Bayesian analysis, aside from a numerical constant [19,20].

To overcome some of these limitations, alternative bounds have been introduced in classical statistics, including the Hammersley-Chapman-Robbins bound [21], the Bhattacharyya bound [22], the Barankin bound [23], and the more general Abel bounds [24]. Similarly, a hierarchy of lower bounds on estimator variance has been explored by imposing additional unbiasedness constraints in the quantum domain [25].

Recently, we have addressed these challenges by leveraging the theory of higher-order asymptotics [26]. Extending beyond the central limit theorem, this framework provides trial-dependent corrections essential for *bona fide* statistical

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by Max Planck Society.

inference [27]. In this paper, we build upon this approach to explore additional largely unexplored issues.

For instance, in single-parameter estimation, multiple measurements can saturate the QCRB [28], raising the question of which measurement should be preferred over the rest. We shall demonstrate that $O(1/\nu^2)$ corrections serve to distinguish among them.

The QCRB also presents ambiguousness when determining the optimal state for sensing a unitary transformation. A notable example is the case of rotations in $SU(2)$, where a whole family of optimal states exists when considering all possible rotation axes [29–32]. Here, too, higher-order terms allow us to distinguish between states that are otherwise equivalent under the QCRB.

These findings are instrumental for preasymptotic metrology, enabling the design of optimal protocols well before reaching the asymptotic limit. Since the $O(1/\nu^2)$ corrections depend solely on the quantum state, they allow for straightforward analytical calculations without requiring numerical simulations. In a rapid measurement, where $\nu \sim 10$ – 100 samples can be taken before costs incur or the system changes, this type of correction becomes extremely valuable.

The QCRB is founded on the concept of the symmetric logarithmic derivative (SLD) [33]. The Bhattacharyya bound (BB), in both its classical and quantum formulations, seeks to enhance the QCRB by extending the order of the derivative, making it intrinsically connected to these higher-order corrections. We will demonstrate that it establishes a lower bound for such corrections, offering deeper insights into the fundamental limits of precision in quantum estimation.

This paper is organized as follows. Section II presents the essential background in estimation theory, laying the foundations for the discussions that follow. In Sec. III we examine in detail the previously mentioned limitations of the QCRB. Section IV introduces the theory of higher-order asymptotics, which provides a refined framework for analyzing $O(1/\nu^2)$ corrections. In this context, Sec. V explores the BB and its connections with $O(1/\nu^2)$ terms. We then present its quantum counterpart and analyze its relation with the classical case. Finally, we summarize our findings and present our conclusions in Sec. VI.

II. BACKGROUND

To ensure a self-contained presentation, we briefly outline the key concepts of estimation theory relevant to our discussion. Throughout the paper we restrict ourselves to the case of single-parameter estimation. For further details, we refer the reader to the relevant literature [34,35].

A. Classical estimation theory

To extract information from a physical system, one must perform measurements. The results of such an experiment are represented by a random variable X , which follows a family of probability distributions $p(x|\theta)$, parametrized by θ , the quantity to be estimated. We take X as a discrete variable that can take on N possible values, forming the sample space $S = \{x_1, \dots, x_N\}$. The continuous case can be addressed in a similar manner.

Typically, the experiment is repeated ν times, yielding a set of ν outcomes $\{\mathbf{x}_1, \dots, \mathbf{x}_\nu\}$, where each $\mathbf{x}_i \in S$. These trials are assumed to be independent and identically distributed. The assumption of identical distribution means that each trial is governed by the same parameter θ , while independence implies that the joint distribution of all these trials factorizes

$$p(\mathbf{x}|\theta) = \prod_{i=1}^{\nu} p(\mathbf{x}_i|\theta), \quad (2.1)$$

where $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_\nu)^\top$ and the superscript $\top$ denotes the transpose.

If we denote by ν_i the number of times each x_i occurs in $\mathbf{x}$, with $\nu_1 + \nu_2 + \dots + \nu_N = \nu$, then the probability distribution governing $\{\nu_1, \dots, \nu_N\}$ is given by a multinomial distribution

$$p(\nu_1, \dots, \nu_N|\theta) = \frac{\nu!}{\nu_1! \nu_2! \dots \nu_N!} p(x_1|\theta)^{\nu_1} \dots p(x_N|\theta)^{\nu_N}. \quad (2.2)$$

Because $\{\nu_1, \dots, \nu_N\}$ is a sufficient statistic, it retains all the information contained in the full sample $\mathbf{x}$. In practice, this representation is typically preferred for computational purposes.

The goal is to infer the value of θ from the observed data. To achieve this, we employ an estimator $\hat{\theta}(\mathbf{x})$, which is a function of the observed data only. To quantify the performance of this estimator, we need a loss function that measures the error in the estimation. The most common choice is the squared error $[\hat{\theta}(\mathbf{x}) - \theta]^2$, which leads to the concept of the MSE [36]:

$$\text{MSE}[\hat{\theta}(\mathbf{x})] = \langle [\hat{\theta}(\mathbf{x}) - \theta]^2 \rangle = \sum_{\mathbf{x}} p(\mathbf{x}|\theta) [\hat{\theta}(\mathbf{x}) - \theta]^2, \quad (2.3)$$

where $\langle \cdot \rangle$ denotes the expectation value with respect to the probability distribution $p(\mathbf{x}|\theta)$.

Ideally, we seek an estimator $\hat{\theta}(\mathbf{x})$ that minimizes the MSE for all possible values of θ . However, this is generally not achievable because the optimal estimator may explicitly depend on the true value of the parameter θ , which is unknown [34]. To proceed further, it is necessary to constrain the class of estimators we consider. This leads to the concept of an unbiased estimator, which is defined as an estimator whose expectation value is equal to the true value of the parameter: $\langle \hat{\theta}(\mathbf{x}) \rangle = \theta$.

For unbiased estimators, the classical Cramér-Rao bound (CCRB) establishes a lower bound on the MSE (note that in this case, the MSE coincides with the standard notion of variance):

$$\text{MSE}[\hat{\theta}(\mathbf{x})] \geq \frac{1}{\nu F_x(\theta)}, \quad (2.4)$$

where $F_x(\theta)$ is the classical Fisher information (CFI), defined as [37]

$$F_x(\theta) = \sum_{i=1}^N p(x_i|\theta) \left[\frac{\partial_\theta p(x_i|\theta)}{p(x_i|\theta)} \right]^2. \quad (2.5)$$

Geometrically, the CFI defines a curvature on the statistical manifold of probability distributions, quantifying how distinguishable different parametrized distributions are. A higher

curvature enables more precise parameter estimation. Moreover, the CFI is intrinsically linked to the statistical distance between probability distributions. In particular, for two neighboring distributions, $p(x|\theta)$ and $p(x|\theta + d\theta)$, the infinitesimal statistical distance is given by

$$ds^2 = F_x(\theta) d\theta^2. \quad (2.6)$$

This expression underscores the role of the CFI in shaping the local geometry of the parameter space, establishing a natural scale for parameter estimation.

This unbiasedness constraint facilitates significant analytical progress, allowing the identification of estimators that saturate the bound in specific cases. Even if no estimator fully saturates the bound, it serves as a guiding principle for finding estimators with a lower MSE.

However, this progress is often confined to textbook examples. In most real-world scenarios, identifying unbiased estimators might be impractical. Furthermore, even if such estimators were available, the primary objective remains minimizing the MSE. A biased estimator that achieves a substantially lower MSE is preferable to an unbiased one with a higher MSE [38].

The CCRB is not merely a bound for unbiased estimators; it provides a fundamental lower limit on the MSE for any estimator, whether biased or unbiased, in the asymptotic limit $\nu \rightarrow \infty$. This asymptotic optimality is a highly desirable property, ensuring that with sufficient resources, the MSE reaches its theoretical minimum.

B. Quantum estimation theory

In the quantum version, the probe state is represented by a density operator ϱ . The single parameter θ is encoded via a quantum channel $\mathcal{E}_\theta$, whose action on the state ϱ is the transformation $\mathcal{E}_\theta[\varrho] = \varrho_\theta$ [39]. For simplicity, we will restrict our attention to unitary channels, so that $\varrho_\theta = U_\theta \varrho U_\theta^\dagger$, with the unitary operator U_θ expressed as $U_\theta = \exp(-i\theta G)$, where G is a selfadjoint operator that is called the generator of the transformation.

To estimate θ , we perform a measurement that is represented by some positive operator-valued measure (POVM) $\{\Pi_x\}$ [40]. Then the probability of obtaining outcome x_i is given by the time-honored Born rule [41]:

$$p(x_i|\theta) = \text{Tr}(\varrho_\theta \Pi_{x_i}). \quad (2.7)$$

The goal is to optimally select both the POVM $\{\Pi_x\}$ and the initial state ϱ such that the resulting probability distribution $p(x|\theta)$ is highly sensitive to the parameter θ , enabling its precise estimation. This translates into optimizing ϱ and $\{\Pi_x\}$ to achieve the lowest possible CCRB or, equivalently, the highest possible $F_x(\theta)$. Maximizing the CFI ensures that with sufficient resources, we attain the minimum possible MSE.

Given a quantum state ϱ_θ , maximizing the CFI over all possible POVMs leads to the concept of the QFI, which can be expressed as [11]

$$\mathbf{Q}_\varrho(\theta) = \sup_{\{\Pi_x\}} F_x(\theta), \quad (2.8)$$

and the associated QCRB:

$$\text{MSE}[\hat{\theta}(\mathbf{x})] \geq \frac{1}{\nu \mathbf{Q}_\varrho(\theta)}. \quad (2.9)$$

Helstrom, using elementary arguments [33], showed an explicit way to compute $\mathbf{Q}_\varrho(\theta)$:

$$\mathbf{Q}_\varrho(\theta) = \text{Tr}(\varrho_\theta L_\theta^2), \quad (2.10)$$

where L_θ is the SLD, defined implicitly via

$$\frac{\partial \varrho_\theta}{\partial \theta} = \frac{1}{2} \{\varrho_\theta, L_\theta\}, \quad (2.11)$$

and $\{\cdot, \cdot\}$ is the anticommutator $\{A, B\} = AB + BA$.

The QFI is intimately connected with the distinguishability of a probe for small variations of the parameter [42]. The distinguishability between two states, ϱ and ϱ' , can be quantified by the normalized Bures distance [43] $D_B(\varrho|\varrho') = \sqrt{1 - F(\varrho|\varrho')}$, where $F(\varrho|\varrho') = [\text{Tr}(\sqrt{\sqrt{\varrho} \varrho' \sqrt{\varrho}})]^2$ is the fidelity [44]. Then, given two states ϱ_θ and $\varrho_{\theta+d\theta}$, obtained by an infinitesimal change in the parameter, one has [5]

$$D_B^2(\varrho_\theta|\varrho_{\theta+d\theta}) = \frac{1}{8} \mathbf{Q}_\varrho(\theta) d\theta^2, \quad (2.12)$$

except for pointwise differences when ϱ_θ changes rank [45,46]. Therefore, the more distinguishable a state is from neighboring states in the manifold, the greater the QFI and the sensitivity of the state to the parameter θ , much as in the classical case.

The next step after calculating the QFI is to find the initial state ϱ that maximizes it, thereby optimizing both the measurement and the state. In the general case where θ enters through a quantum channel, this leads to the concept of the channel QFI [47]. In the unitary case with pure states, the state that maximizes the QFI is simply a superposition of the eigenstates corresponding to the largest and smallest eigenvalues of the generator [48]:

$$|\psi_{\text{opt}}\rangle = \frac{1}{\sqrt{2}}(|g_{\text{max}}\rangle + |g_{\text{min}}\rangle). \quad (2.13)$$

III. LIMITATIONS OF THE CRAMÉR-RAO BOUND

The first limitation of the QCRB has already been mentioned: It only provides information about the achievable performance in the asymptotic limit of infinite resources, $\nu \rightarrow \infty$. Yet in practical scenarios, resources are always finite. This raises a critical question: How many trials, ν , are required for the QCRB to accurately approximate the actual MSE? More precisely, for a finite number of trials ν , what is the true MSE? We will explore this in Sec. IV. We turn to two additional limitations that, despite their relevance, have received less attention within the community. We will use the lens of rotation sensing for illustration.

A. Ambiguity in optimal states

Let us consider the illustrative example of a rotation $R = \exp(i\theta J_n) \in \text{SU}(2)$, parametrized by its rotation axis $\mathbf{n}$ and angle θ . Here $J_n = \mathbf{J} \cdot \mathbf{n}$, with $\mathbf{J}$ being the generators of the rotation, which satisfy the commutation relations of the Lie algebra $\mathfrak{su}(2)$. The irreducible representations of this algebra are labeled by J and are spanned by the states

$|J, m\rangle$, which are the simultaneous eigenstates of $\mathbf{J}^2$ and J_z . These states form a basis for a $(2J + 1)$ -dimensional Hilbert space, denoted by $\mathcal{H}_J$. In what follows, we assume a pure state $|\psi\rangle \in \mathcal{H}_J$ that undergoes the rotation R , resulting in $|\psi_\theta\rangle = R|\psi\rangle$.

For a fixed $\mathbf{n}$ (which we can take as the z axis without loss of generality) and a fixed J , the states that maximize the QFI (and minimize the QCRB) are the time-honored NOON states [49]:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|J, J\rangle + |J, -J\rangle), \quad (3.1)$$

in agreement with (2.13). In the context of magnetometry, these correspond to the Greenberg-Horne-Zeilinger (GHZ) states [50]. These optimal (pure) states verify $\mathbf{Q}_\theta(\theta) = 4\mathbf{m}_2(\mathbf{n})$, independent of θ , where we adopt throughout the notation

$$\mathbf{m}_k(\mathbf{n}) = \langle \psi | (J_n - \langle J_n \rangle)^k | \psi \rangle \quad (3.2)$$

for the k th central moment of the generator J_n .

When the axis $\mathbf{n}$ is not fixed, an optimal state should perform well for any arbitrary $\mathbf{n}$. A reasonable criterion is to minimize the average QCRB over all rotation axes; i.e.,

$$\frac{1}{4\pi} \int_{S_2} \frac{1}{4\mathbf{m}_2(\mathbf{n})} d\Omega \equiv \frac{1}{4} \overline{\frac{1}{\mathbf{m}_2}}, \quad (3.3)$$

with the bar denoting such an average over the unit sphere S_2 . The states that optimize this average are second-order isotropic, meaning their first two moments are independent of the rotation axis: $\langle J_n \rangle = 0$ and $\langle J_n^2 \rangle = \text{const}$ [51]. These states are not necessarily unique for a given J . Additionally, higher-order isotropic states exist [52,53], raising the question of whether increasing isotropy further reduces the $O(1/\nu^2)$ terms of the MSE.

Higher-order isotropic states naturally emerge when analyzing the fidelity between a quantum state and its rotated counterpart [54]:

$$\begin{aligned} |\langle \psi | e^{iJ_n\theta} | \psi \rangle|^2 &= 1 - \mathbf{m}_2\theta^2 + \frac{1}{12}(3\mathbf{m}_2^2 + \mathbf{m}_4)\theta^4 \\ &+ \frac{1}{360}(10\mathbf{m}_3^2 - 15\mathbf{m}_2\mathbf{m}_4 - \mathbf{m}_6)\theta^6 + O(\theta^7). \end{aligned} \quad (3.4)$$

Our goal is to identify the states that, when averaged over all rotation axes $\mathbf{n}$, minimize this fidelity; i.e., those that undergo the largest change. To achieve this, we analyze the average behavior of each order in θ separately.

(i) *Second-order.* This term is proportional to the QFI. The average value $\overline{\mathbf{m}_2}$ is maximized by states satisfying $\langle J_n \rangle = 0$; that is, first-order isotropic [32].

(ii) *Fourth-order.* Among first-order isotropic states, the optimal ones are those that further minimize the next-order term. Since $\langle J_n \rangle = 0$, we have $\mathbf{m}_k = \langle J_n^k \rangle$. However, it can be readily shown that $\overline{\langle J_n^k \rangle} = \text{const}$ and, consequently, $\overline{\mathbf{m}_4}$ is constant and independent of the state.

For the term $\overline{\mathbf{m}_2^2}$ we use Jensen's inequality to show that

$$\frac{1}{4\pi} \int_{S_2} \langle J_n^k \rangle^2 d\Omega \geq \left(\frac{1}{4\pi} \int_{S_2} \langle J_n^k \rangle d\Omega \right)^2 = \text{const}. \quad (3.5)$$

Therefore, this term is minimized by states with $\mathbf{m}_2 = \text{const}$. This condition is satisfied by second-order isotropic states, making them the optimal choice at this order.

(iii) *Sixth-order.* Applying the same reasoning, the optimal states at this order are those that minimize $\overline{\mathbf{m}_3^2}$. This requirement is satisfied by third-order isotropic states.

In Appendix A we provide a mathematical proof that this pattern continues to arbitrary orders, demonstrating that higher-order isotropic states are always optimal at minimizing fidelity under rotation.

B. Ambiguity in optimal measurements

For the CFI to match the QFI, a specific POVM must be used. This POVM is generally not unique. Despite its importance, this issue has received relatively little attention in the quantum metrology community. Research in this field typically focuses on the QFI itself, with the assumption that any measurement achieving it is sufficient.

The conditions that a POVM must satisfy to ensure that the CFI equals the QFI were recently expressed in a particularly simple form in Ref. [55]. Building on this formulation, Ref. [28] derived the most general form of such a POVM for pure states, which is the case we consider here.

Accordingly, let us define

$$|\pm\rangle = \frac{1}{\sqrt{2}} \left[|\psi_{\theta_0}\rangle \mp \frac{1}{\sqrt{\mathbf{m}_2}} (J_n - \langle J_n \rangle) |\psi_{\theta_0}\rangle \right], \quad (3.6)$$

where θ_0 is the value at which the measurement is evaluated. The most general quantum measurement that attains the QFI is given by a set of projectors $\Pi_i = \lambda_i |\Phi_i\rangle \langle \Phi_i|$ onto the states

$$|\Phi_i\rangle = \frac{1}{\sqrt{2}} (|+\rangle + e^{i\phi_i} |-\rangle), \quad (3.7)$$

such that $\sum_i \lambda_i = 2$ and $\sum_i \lambda_i e^{i\phi_i} = 0$.

We focus on the simplest case of two projectors, with $\lambda_1 = \lambda_2 = 1$. This yields a family of measurements parametrized by an angle ϕ_+ ,

$$\begin{aligned} |\Phi_{\pm}(\theta_0)\rangle &= \frac{1}{2} \left[(1 + e^{i\phi_{\pm}}) |\psi_{\theta_0}\rangle + \frac{(1 - e^{i\phi_{\pm}})}{\sqrt{\mathbf{m}_2}} (J_n - \langle J_n \rangle) |\psi_{\theta_0}\rangle \right], \end{aligned} \quad (3.8)$$

where $\phi_- = \phi_+ - \pi$.

A subtle but often overlooked point in quantum metrology is that when considering the probability distribution $p(x|\theta)$, the parameter θ_0 that influences the measurement must be treated as a fixed constant, even when $\theta_0 = \theta$. As a consequence, the resulting probabilities, $|\langle \psi_\theta | \Phi_{\pm}(\theta_0) \rangle|^2$, sum to unity only when $\theta = \theta_0$. Actually, a simple calculation shows

that

$$p(x_+|\theta) + p(x_-|\theta) = 1 + \frac{m_3^2 + m_3^2 - m_2 m_4}{4m_2}(\theta - \theta_0)^4 + O(\theta^5). \quad (3.9)$$

This implies that we must extend our measurement. However, modifying the POVM can affect the performance of real estimators, such as the maximum likelihood estimator (MLE), even when $\theta = \theta_0$. We will consider two ways of addressing this issue.

The first method is to replace one of the initial projectors, say, $|\Phi_-(\theta_0)\rangle\langle\Phi_-(\theta_0)|$, with $\Pi_2 = \mathbb{1} - |\Phi_+(\theta_0)\rangle\langle\Phi_+(\theta_0)|$. When $\theta = \theta_0$, this is equivalent to the $\Pi_\pm$ measurement, but when $\theta \neq \theta_0$, Π_2 absorbs the missing probability, ensuring a complete POVM.

The second approach involves extending the two initial measurements by introducing a set of POVM elements $\{\Pi_3, \Pi_4, \dots\}$, with associated probabilities $\{p(x_3|\theta), p(x_4|\theta), \dots\}$ that are zero when $\theta = \theta_0$. In principle, other types of extensions might be conceived, but they would essentially be combinations of these two. While both extensions achieve the QCRB, we will demonstrate that their second-order corrections differ.

IV. BEYOND THE ASYMPTOTIC LIMIT

A. Higher-order asymptotics

A wide variety of estimators have been formulated by proposing ad hoc functions [34,35]. Among these, one of the most widely adopted is the MLE, largely due to its turn-the-crank nature. Popularized by Fisher [56], the MLE provides a general-purpose method that can be readily applied across a broad range of problems without requiring problem-specific tuning. It is defined as

$$\hat{\theta}_{\text{MLE}}(\mathbf{x}) = \arg \max_{\theta} p(\mathbf{x}|\theta). \quad (4.1)$$

Here the likelihood $p(\mathbf{x}|\theta)$ must be interpreted as the probability of θ being the true value given the dataset $\mathbf{x}$. This estimator just picks the value of θ that makes $\mathbf{x}$ the most likely observation; in other words, *it bets on the winner*.

Under mild regularity conditions, the MLE enjoys desirable asymptotic properties: It is consistent, asymptotically efficient, and normally distributed as the sample size grows [57]. Thus, in the regime of abundant resources, one can be confident in achieving the minimum possible MSE. Nevertheless, many estimators are known to reach the CCRB asymptotically, raising the question of why one should be preferred over another.

In a series of influential papers, Rao addressed this question [58–60] (for additional perspectives, see also Refs. [61,62]). For any efficient estimator, the MSE can always be expanded as

$$\text{MSE}[\hat{\theta}(\mathbf{x})] = \frac{1}{v} \frac{1}{F_x(\theta)} + \frac{1}{v^2} \left[\Gamma(\theta) + b^2(\theta) + 2 \frac{b'(\theta)}{F_x(\theta)} \right] + O\left(\frac{1}{v^3}\right), \quad (4.2)$$

where $b(\theta)/v = \langle \hat{\theta}(\mathbf{x}) - \theta \rangle$ is the bias of order $O(1/v)$ and $\Gamma(\theta)$ is an estimator-dependent function. Specifically, for the MLE, we have

$$b(\theta) = -\frac{1}{2v} \frac{\gamma_{11}}{F_x^2(\theta)},$$

$$\Gamma(\theta) = -\frac{1}{F_x(\theta)} + \frac{\gamma_{02}(\theta) - 2\gamma_{21}(\theta) + \gamma_{40}(\theta)}{F_x^3(\theta)} + \frac{\frac{1}{2}\gamma_{11}^2(\theta) - [\gamma_{11}(\theta) - \gamma_{30}(\theta)]^2}{F_x^4(\theta)}, \quad (4.3)$$

with

$$\gamma_{rs}(\theta) = \sum_{i=1}^N p(x_i|\theta) \left[\frac{\partial_{\theta} p(x_i|\theta)}{p(x_i|\theta)} \right]^r \left[\frac{\partial_{\theta}^2 p(x_i|\theta)}{p(x_i|\theta)} \right]^s. \quad (4.4)$$

In general, no single estimator minimizes the $O(1/v^2)$ terms in Eq. (4.2). To establish a fundamental bound at this order, it is necessary to restrict the analysis to estimators that are unbiased up to $O(1/v)$. More precisely, while comparisons could be made between estimators with the same bias $b(\theta)$, the most natural approach is to consider the strictly unbiased case, where $b(\theta) = 0$ [27]. This restriction does not result in a significant loss of generality, as any estimator $\hat{\theta}(\mathbf{x})$ can be transformed into a bias-corrected version

$$\hat{\theta}^*(\mathbf{x}) = \hat{\theta}(\mathbf{x}) - \frac{b[\hat{\theta}(\mathbf{x})]}{v}, \quad (4.5)$$

which is unbiased to order $O(1/v)$ [60].

In this case, the MSE takes the form:

$$\text{MSE}[\hat{\theta}^*(\mathbf{x})] = \frac{1}{v} \frac{1}{F_x(\theta)} + \frac{\Gamma(\theta)}{v^2} + O\left(\frac{1}{v^3}\right). \quad (4.6)$$

In this way, $\Gamma(\theta)$ appears as a second-order correction to the CCRB. The key result established by Rao is that the MLE is the unique estimator that minimizes $\Gamma(\theta)$. This highlights the distinguished status of the MLE among all efficient estimators, not just in terms of first-order efficiency but also in minimizing second-order error contributions.

B. First POVM extension

Even once the MLE is chosen as the estimator, one must still choose among POVM strategies. First, we calculate $\Gamma(\theta)$ for the first POVM extension, $p(x_2|\theta) = 1 - p(x_+|\theta)$. We assume without loss of generality that $\theta_0 = 0$ from now on. In this case, one gets

$$\Gamma(0) = \frac{\cot^2 \phi_+}{8m_2}. \quad (4.7)$$

The minimum value of $\Gamma(0)$ is zero when $\phi_+ = -\pi/2$, identifying this as the optimal measurement setting. At this value, the probabilities satisfy $p(x_+|0) = p(x_2|0) = 1/2$, corresponding to a fair coin scenario. In this case, the CCRB is attained asymptotically because the MLE converges to a Gaussian distribution. This convergence occurs more rapidly when the underlying distribution resembles a fair coin, compared to a biased one. This difference is captured in the $O(1/v^2)$ correction terms, which are smaller for the fair coin case, indicating faster asymptotic efficiency.

If we do not fix the measurement at $\phi_+ = -\pi/2$, then one could in principle average over all rotation axes to identify the states that minimize $\Gamma(\theta)$, among those that are optimal according to the QCRB. However, for these states $\langle J_n \rangle = 0$ and $\langle J_n^2 \rangle = m_2 = \text{const}$, so this term is the same for all of them. Interestingly, the measurement with $\phi_+ = -\pi/2$ is also optimal according to the Fisher information measurement noise susceptibility, a recently introduced concept that quantifies a measurement's resilience to noise [28].

It is also interesting to analyze the standard MLE, without bias correction. The $O(1/\nu^2)$ correction is given by

$$\begin{aligned} \Gamma(0) + b^2(0) + 2 \frac{b'(0)}{F_x(0)} \\ = \frac{m_2^2 [2 \cos(2\phi_+) + 5] \csc^2(\phi_+) + m_4}{16m_2^3}. \end{aligned} \quad (4.8)$$

As before, the measurement that minimizes this expression is $\phi_+ = -\pi/2$. With this choice, a remaining $O(1/\nu^2)$ term from the bias derivative appears, $(3m_2^2 + m_4)/(16\nu^2 m_2^3)$, which is always positive.

1. Third-order corrections

Since we can set $\Gamma(0) = 0$ by choosing $\phi_+ = -\pi/2$, we then look at the next term $O(1/\nu^3)$ in the expansion. With the measurement fixed, we calculate these corrections following the procedure in Ref. [63].

The very same definition of the MLE; that is, $\max_{\theta} \{v_+ \ln[p(x_+|\theta)] + v_2 \ln[p(x_2|\theta)]\}$, implies

$$\frac{v_+}{p(x_+|\theta)} \frac{dp(x_+|\theta)}{d\theta} + \frac{v_2}{p(x_2|\theta)} \frac{dp(x_2|\theta)}{d\theta} = 0. \quad (4.9)$$

We substitute a Taylor series expansion for $p(x_+|\theta)$

$$p(x_+|\theta) = \frac{1}{2} - \sqrt{m_2} \theta + \frac{(3m_2^2 + m_4)\theta^3}{6\sqrt{m_2}} + O(\theta^4) \quad (4.10)$$

into the MLE equation (4.9):

$$0 = -4\delta_+ \sqrt{m_2} - 4m_2 \nu \theta - \frac{2\delta_+(5m_2^2 - m_4)}{\sqrt{m_2}} \theta^2 + O(\theta^3), \quad (4.11)$$

where we have used the variable δ_+ defined as $v_+ = \delta_+ + \nu/2$, since we will need to take the limit $\nu \rightarrow \infty$ and require the remaining terms to be small compared to ν . We invert this series to obtain the value of θ_{MLE} . This can be done using the Lagrange-Bürmann inversion formula [64,65]; the result reads

$$\begin{aligned} \hat{\theta}_{\text{MLE}}^2(\mathbf{x}) = \frac{\delta_+^2}{m_2 \nu^2} + \frac{\delta_+^4 (3m_2^2 + m_4)}{3m_2^3 \nu^4} \\ - \frac{\delta_+^5 (m_3^2 - m_2 m_4 + m_3^2)}{4m_2^4 \nu^5} + \dots \end{aligned} \quad (4.12)$$

This expression can be used to estimate the value of $\hat{\theta}_{\text{MLE}}^2(\mathbf{x})$ as a function of the experimental outcomes v_+ . Finally, we take the expectation value $\langle \theta_{\text{MLE}}^2(\mathbf{x}) \rangle$. The quantities $\langle \delta_+^k \rangle$ are the central moments of a binomial distribution with a fair coin [66], given by $\langle \delta_+^2 \rangle = \nu/4$, $\langle \delta_+^4 \rangle = \nu(3\nu - 2)/16$ and $\langle \delta_+^6 \rangle = [15(\nu - 2)\nu + 16]/64$. For odd k , they are 0. Substituting and performing an expansion in powers of $1/\nu$, we

arrive at

$$\begin{aligned} \text{MSE}[\hat{\theta}_{\text{MLE}}(\mathbf{x})]_{\theta=0} \\ = \frac{1}{4m_2 \nu} + \frac{3m_2^2 + m_4}{16m_2^3 \nu^2} \\ + \frac{219m_2^4 + 133m_2^2 m_4 + m_2(30m_3^2 - 3m_6) + 35m_4^2}{768m_2^5 \nu^3} \\ + O\left(\frac{1}{\nu^4}\right). \end{aligned} \quad (4.13)$$

We can refine our search for optimal states by considering the $O(1/\nu^3)$ terms, since the $O(1/\nu^2)$ term does not distinguish between different optimal states. When averaging over all possible rotation axes, the only state-dependent terms in $O(1/\nu^3)$ are $30m_2 m_3^2$ and $35m_4^2$. As we previously observed, both of these terms are minimized by fourth-order isotropic states. This reveals that not all states that were second-order isotropic are equivalent: The MLE favors higher-order isotropic states.

2. Third-order corrections with bias correction

We next consider the MSE of the bias-corrected MLE estimator. We expand $b(\hat{\theta}) = b(0) + a_1 \hat{\theta} + a_2 \hat{\theta}^2 + a_3 \hat{\theta}^3$ in $\langle [\hat{\theta}(\mathbf{x}) - b(\hat{\theta}(\mathbf{x}))]/\nu \rangle^2$. Note that $b(0) = 0$ for $\phi_+ = -\pi/2$. We obtain a lengthy expression in terms of $\langle \hat{\theta}^k(\mathbf{x}) \rangle$. The details can be seen in the Appendix B. The final result is

$$\begin{aligned} \text{MSE}[\hat{\theta}_{\text{MLE}}^*(\mathbf{x})]_{\theta=0} \\ = \frac{1}{4m_2 \nu} \\ + \frac{84m_2^4 - 11m_2^2 m_4 + 3m_2(-10m_3^2 + m_6) - 28m_4^2}{768m_2^5 \nu^3} \\ + O\left(\frac{1}{\nu^4}\right). \end{aligned} \quad (4.14)$$

When searching for the optimal states for detecting an arbitrary rotation, we find that the only nonconstant terms are m_3^2 and m_4^2 . Unlike the standard MLE case, where fourth-order isotropic states were optimal, here they are actually the worst choices, because these nontrivial terms contribute with negative coefficients. To our knowledge, states that are second-order isotropic but exhibit *anti-isotropic* behavior at higher orders have not been studied.

3. Preasymptotic performance

We turn to an important application: estimating performance before reaching the asymptotic limit; a regime where the QCRB provides limited guidance.

It is crucial to emphasize that discussing the MSE of the MLE (or any estimator) in the low-trial regime is meaningless without specifying prior information about the parameter. In particular, the admissible range of values for θ must be clearly defined by the experimental context.

To illustrate this, we analyze the behavior of the MLE using the previous POVM extension, with $\phi_+ = -\pi/2$, which is a binomial model. Similar qualitative conclusions would apply in more general multinomial scenarios. The key idea is that, given an observed outcome v_+ , the MLE estimates θ

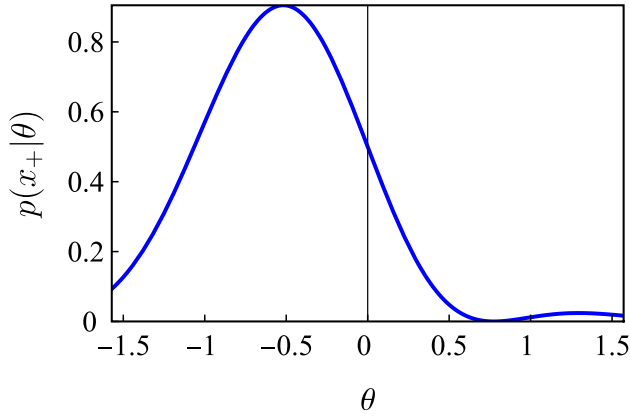


FIG. 1. Probability $p(x_+|\theta)$ for a coherent state (with $J = 3$) pointing along the z axis, after being rotated by an angle θ around the x axis: $|\psi_\theta\rangle = e^{-\theta J_x}|J, J\rangle$.

by setting $p(x_+|\theta) = v_+/v$, as it follows from Eq. (4.9), after substituting $v = v_+ + v_2$ and $p(x_2|\theta) = 1 - p(x_+|\theta)$. Alternatively, this result follows from the invariance property of the MLE and the fact that for a binomial distribution, the MLE for the coin's q value is simply $\hat{q} = v_+/v$. If this value does not exist, then it returns the closest value.

Figure 1 illustrates the behavior of $p(x_+|\theta)$ for a coherent state pointing in the z axis, rotated around the x axis $|\psi_\theta\rangle = e^{-\theta J_x}|J, J\rangle$ with $J = 3$. This example will serve as a guiding thread throughout the discussion. The maximum detectable range of θ is limited by the region where the function does not turn around (i.e., where the derivative remains nonzero). If the function were to turn around, then the MLE would yield multiple values of θ , making the problem unidentifiable.

The importance of prior information can be stressed by considering a single trial ($v = 1$ and $v_+ = 1$). The MLE equation simplifies to $\arg \max_\theta p(x_+|\theta)$. If $v_2 = 1$, then the MLE instead maximizes $1 - p(x_+|\theta)$, leading to the lowest possible θ . These extrema may coincide with the range beyond which the model becomes unidentifiable, but this is not always the case. If the system rotates slightly beyond this range, then it becomes impossible to detect, making the effective range smaller in practice.

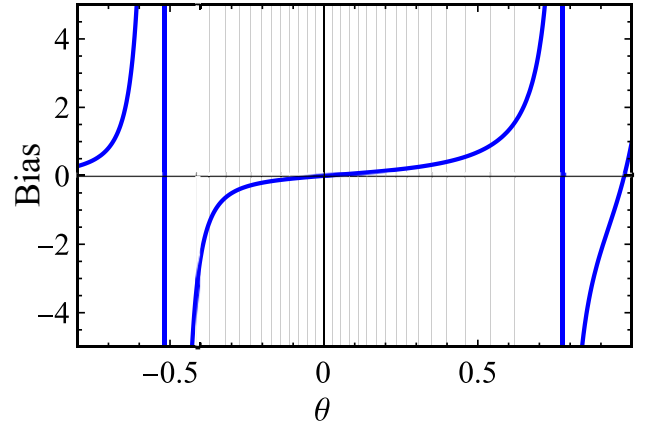
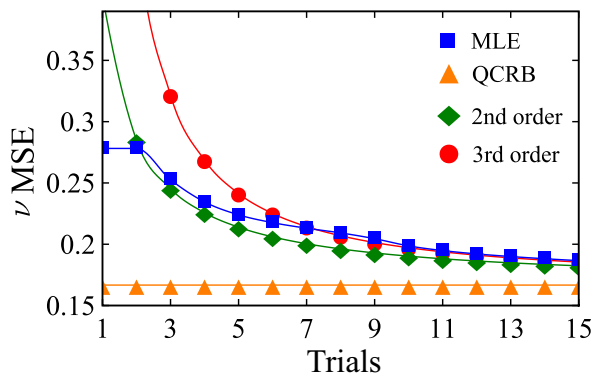


FIG. 3. Bias $b(\theta)$ for the same state as in Fig. 1. The vertical lines indicate the values of θ returned by the MLE for the different experimental outputs.

Figure 2 illustrates how the MSE varies with the choice of prior range for θ . Comparing the full range with a reduced 80% interval, the results reveal a strong dependence of the MSE on the assumed prior. Further narrowing the range to 50% would reduce the MSE even further.

Since we have to choose a range, and there is no single correct answer, the most natural choice is simply to take the maximum range. We have observed that for all states (not just for our example) the bias diverges near the edges, rendering bias correction unfeasible. A practical workaround is to either exclude these boundary regions or apply a cutoff beyond which bias correction is no longer applied.

A more natural approach, though, is simply to restrict the θ range to avoid the divergence. This is illustrated in Fig. 3, where we restrict the range to 80% to avoid falling into the diverging region. The figure shows the bias $b(\theta)$, and how the MLE assigns values of θ based on different experimental results ($v = 30$), with $p(x_+|\theta) = 0, 1/30, 2/30, \dots, 1$. Owing to the 80% threshold, some of the equations lack solutions, causing several solution lines to cluster at the boundaries. This behavior follows directly from the original definition of the MLE.

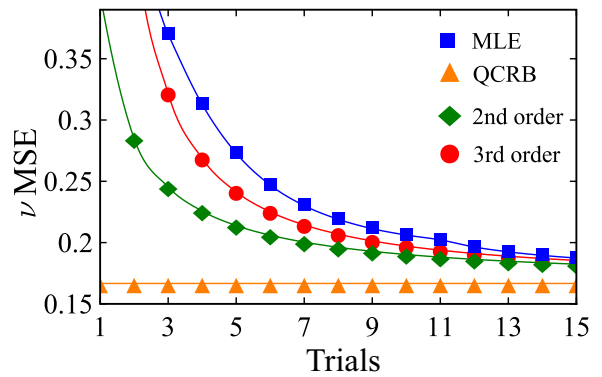


FIG. 2. MSE of the MLE as a function of the number of trials for the same state as in Fig. 1. In the left panel, the parameter θ is restricted to 80% of its maximum admissible range, while in right panel, the full range is considered. The plots also show the QCRB and the approximations given by equation (B6), including corrections up to orders up to $1/v^2$ and $1/v^3$.

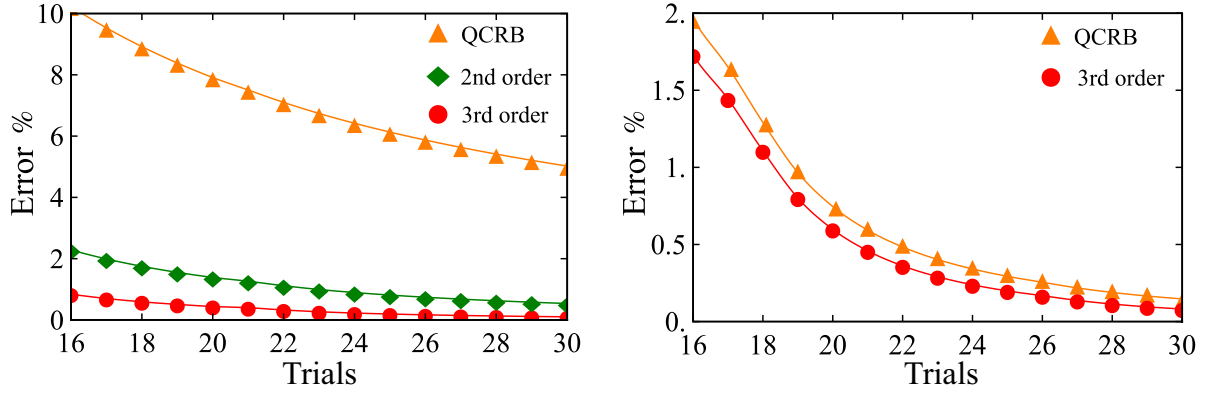


FIG. 4. Relative error in the MSE when using the MLE with different approximations. The left panel shows results for the standard MLE, while the right panel presents the bias-corrected MLE. Insets within each panel illustrate the specific approximation method used for the first POVM extension.

To assess the usefulness of our analytical corrections, Fig. 4 shows the percentage error when approximating the MSE using the QCRB alone versus including $O(1/\nu^2)$ and $O(1/\nu^3)$ corrections. This is calculated as $|\text{MSE}_{\text{approx}} - \text{MSE}|/\text{MSE}$. For the standard MLE, these corrections significantly improve accuracy over the QCRB. For the bias-corrected MLE, the corrections have little impact because eliminating the $O(1/\nu^2)$ term causes the MSE to converge to the QCRB much faster. In general, whether a $O(1/\nu^3)$ or even $O(1/\nu^2)$ correction is relevant or not depends on the specific context.

A final numerical validation of our expressions can be performed by computing $\nu^2[\langle \hat{\theta}^2(\mathbf{x}) \rangle - 1/m_2\nu]$ and similarly for third order. In Appendix B we show how the theoretical values of these corrections are in excellent agreement with the numerical results.

C. Second POVM extension

This case is more complex than the previous one. The key insight is that, according to Eq. (4.2), the corrections $O(1/\nu^2)$ depend on the derivative of the bias; specifically, on its behavior in a neighborhood of the point of interest.

The MLE equation for this extension $\{\Pi_3, \Pi_4, \dots\}$, with probabilities $\{p(x_3|\theta), p(x_4|\theta), \dots\}$ that are zero at $\theta = 0$, is $\max_{\theta} \sum_x \nu_x \ln[p(x|\theta)]$.

At $\theta = 0$, we have $\nu_3, \nu_4, \dots = 0$, as the corresponding outcomes have zero probability. As a result, the MLE depends solely on $p(x_+|\theta)$ and $p(x_-|\theta)$, assuming it is maximized at $\theta = 0$. This is somewhat counterintuitive, because the bias $b(\theta)$, for $\theta \neq 0$, clearly depends on the chosen extension, which in turn should influence its derivative and, consequently, the MSE at $\theta = 0$. This highlights how probabilities that approach zero can introduce subtle complications in the analysis.

In fact, attempting to compute second-order corrections using only $p(x_+|\theta)$ and $p(x_-|\theta)$ leads to incorrect results. One possible approach is to introduce a third outcome, defining $p(x_3|\theta) = 1 - p(x_+|\theta) - p(x_-|\theta)$, but this produces an inaccurate expression. More generally, this theoretical framework breaks down at these points and the fundamental quantity

$\Gamma(\theta)$ loses its significance. We discuss this in the next subsection.

We proceed by computing these corrections following the same approach used to determine the $O(1/\nu^3)$ terms in the previous case. The procedure is largely similar, with two key differences. First, we have $x_+ = \delta_+ + \nu \cos^2(\phi_+/2)$, as we do not fix the value of ϕ_+ . This also changes the form of the $\langle \delta_+^k \rangle$; specifically, we obtain $\langle \delta_+^2 \rangle = \frac{1}{4}\nu \sin^2 \phi_+$, $\langle \delta_+^3 \rangle = -\frac{1}{4}\nu \sin^2 \phi_+ \cos \phi_+$, and $\langle \delta_+^4 \rangle = \frac{1}{32}\nu \sin^2 \phi_+[3(2 - \nu)\cos(2\phi_+) + 3\nu + 2]$. The second change arises after substituting these terms into Eq. (4.12), where terms involving δ_+ are in the denominator. To eliminate them and subsequently compute the mean values of δ_+^k , we must first expand in powers of $1/\nu$ before proceeding further. The details can be found in Appendix B. This leads us to

$$\begin{aligned} \text{MSE}[\hat{\theta}_{\text{MLE}}(\mathbf{x})]_{\theta=0} &= \frac{1}{4m_2\nu} + \frac{m_2^3(14 \csc^2 \phi_+ - 5) - m_2m_4 + 3m_3^2}{32m_4^2\nu^2} \\ &\quad + O\left(\frac{1}{\nu^3}\right). \end{aligned} \quad (4.15)$$

As in the previous case, the measurement that minimizes the order $O(1/\nu^2)$ is $\phi_+ = -\pi/2$.

Regarding the optimal states, when taking the mean value, the only term that does not remain constant is $\overline{m_3^2}$, which is minimized by third-order isotropic states. This result aligns with the MLE without bias correction. However, upon calculating the next correction, we find that the fourth-order isotropic states are the worst, and the best states would be those that maximize the value of $\overline{m_4^2}$. The calculation is lengthy, we defer its full derivation to Appendix B.

As with the previous POVM extension, we test the corrections using the same rotated coherent state as before and also fix $\phi_+ = -\pi/2$. Here we take the maximum admissible range for θ , as bias correction is not meaningful in this regime. In Fig. 5 we compare the error predicted by the QCRB with the error obtained from (4.15). As before, the expressions provide significant insight. As previously noted, the relevance of corrections of order $O(1/\nu^2)$ or $O(1/\nu^3)$ ultimately depend on the particular context.

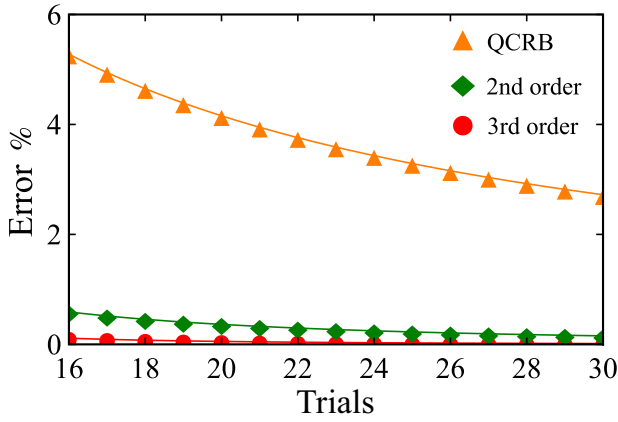


FIG. 5. Relative error in the MSE when using the MLE with the approximations indicated in the inset, computed using Eq. (4.15), for the second POVM extension.

D. The problem of vanishing probabilities

Before tackling the complications of vanishing probabilities, let us first consider a simpler case that clarifies the core issue. Returning to the first POVM extension with $\phi_+ = 0$, we find $p(x_2|0) = 0$, indicating a probability that vanishes at $\theta = 0$. Unlike the more complex scenario where the outcome space grows rapidly with ν , here the number of possible outcomes is limited to $\nu + 1$, and the MLE is determined by solving $p(x_+|\theta) = \nu_+/\nu$.

In Fig. 6, we show the MSE and the bias of the MLE for a NOON state (3.1), though this choice does not play a particularly special role in the analysis. On the right side of the

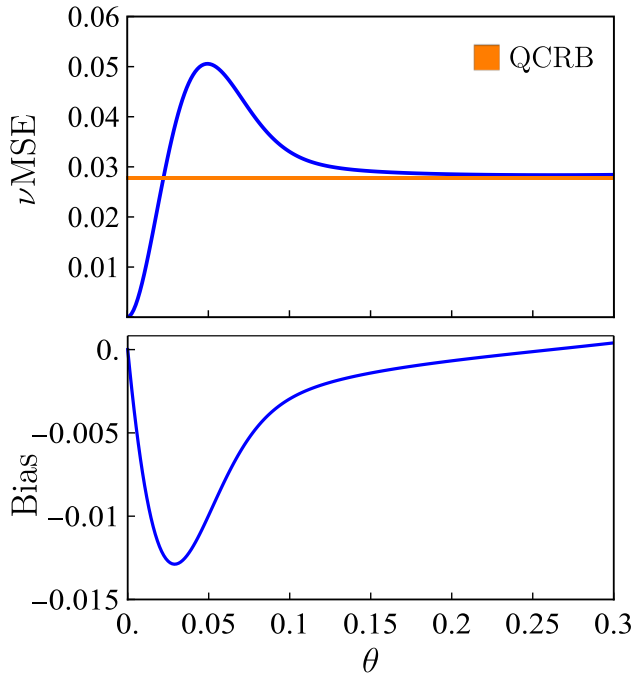


FIG. 6. MSE and bias of the MLE for the estimation of the rotation angle around the z axis with a NOON state (3.1). The number of trials is $\nu = 30$, although this behavior is observed for arbitrarily high values of ν .

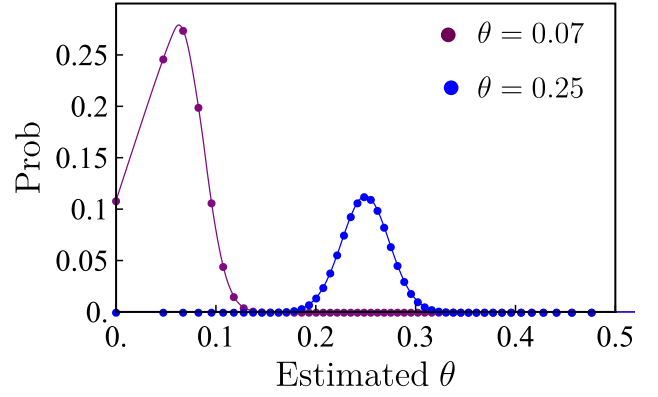


FIG. 7. Probability distribution of the MLE for the true values $\theta = 0.07$ and $\theta = 0.25$ with $\nu = 30$ trials for the same NOON state as in Fig. 6.

plot, the QCRB is attained and the MLE remains unbiased, as it happens in the asymptotic limit. However, as θ approaches zero, clear deviations emerge: The MSE diverges from the asymptotic trend, and the estimator becomes biased. This behavior persists even as ν increases: Larger ν merely shrinks the region of deviation but does not eliminate it.

To understand this behavior, Fig. 7 shows the probability distribution of the MLE when the true values of the parameter are $\theta = 0.07$ and $\theta = 0.25$. In the first case, the distribution is still far from a Gaussian, whereas in the second case it already closely approximates one. No matter how large we make ν , we can always find a sufficiently small θ for which the MLE has not yet converged to a Gaussian distribution. This explains the *bump* observed in Fig. 6: The system has simply not yet reached the asymptotic regime. The fact that the MSE becomes 0 at $\theta = 0$ is specific to this particular example. It arises because at $\theta = 0$ the experiment can only return one result. Since the MLE assigns $\theta = 0$ to this outcome, the MSE is necessarily zero, as the estimator is always correct in this specific case.

Now let us examine the case of $p(x_+|\theta)$ and $p(x_-|\theta)$. When $\theta = 0$, we attain the QCRB, as previously observed, and we do not encounter the same issues as in the previous example. To analyze this further, let us consider the simplest possible extension: $p(x_3|\theta) = 1 - p(x_+|\theta) - p(x_-|\theta)$. The choice of this particular extension is not crucial, as the argument we are about to make holds for any other extension as well.

The MSE of the MLE can be decomposed into two distinct components:

$$\begin{aligned} \text{MSE}[\hat{\theta}_{\text{MLE}}(\mathbf{x})] &= \sum_{\substack{v_3=0 \\ v_+, v_-}} [\hat{\theta}_{\text{MLE}}(v_+, v_-, v_3) - \theta]^2 p(v_+, v_-, 0) \\ &\quad + \sum_{\substack{v_3=1 \\ v_+, v_-}}^{\nu} [\hat{\theta}_{\text{MLE}}(v_+, v_-, v_3) - \theta]^2 p(v_+, v_-, v_3) \\ &\equiv \Delta_0 + \Delta. \end{aligned} \quad (4.16)$$

Consider a sufficiently large ν such that, at $\theta = 0$, we are already in the asymptotic regime (e.g., $\nu \sim 30$ is often sufficient). In this case, the MSE is given by Δ_0 , with $\Delta = 0$. Now, if we choose a small nonzero value for θ while keeping ν

fixed, then the probability that the outcome v_3 does not occur in the experiment is given by

$$\sum_{v_+, v_-} p(v_+, v_-, 0) = [1 - p(x_3|\theta)]^v \simeq (1 - \theta^4)^v. \quad (4.17)$$

For a fixed v , we can always take θ small enough that the probability of observing v_3 becomes negligible. As a result, the contribution from Δ can be made arbitrarily small, since the relevant probabilities vanish while the MLE remains bounded. By continuity of the MSE, this implies that the QCRB is saturated in the limit of small θ .

Now, keeping this small fixed θ , let us increase v until $[1 - p(x_3|\theta)]^v \simeq 0$. This marks the asymptotic regime for this value of θ , where the MSE converges to $1/F_x(\theta)$. In this limit, $\Delta_0 \simeq 0$, and the MSE is dominated entirely by the contribution from Δ . Between the initial regime and this asymptotic limit, there exists a transition region in which the CCRB is not attained.

For any fixed v , there always exists a range of θ values lying within the transition region between the asymptotic regime and the QCRB value at $\theta = 0$. In this transition region, the qualitative behavior of vanishing probabilities resembles that shown in Fig. 6, with one key difference: At $\theta = 0$, the MSE reaches the QCRB rather than dropping to zero.

It is noteworthy that this seemingly pathological phenomenon is, in fact, likely to be common in quantum metrology. Unless the density matrix is full rank, it is typical for its support to change as the parameter varies. While this behavior has been previously observed [67,68], to the best of our knowledge, its underlying origin has not been thoroughly investigated nor has its generality across quantum estimation problems been fully recognized.

V. BHATTACHARYYA BOUND

This section revisits the BB [22]. While it provides less detail than higher-order asymptotic theories, it still offers a rigorous lower bound. A key advantage is its quantum extension, allowing direct evaluation from the state without searching for the optimal POVM, making calculations much simpler.

A. Classical Bhattacharyya bound

We begin with the derivation of the classical Bhattacharyya bound (CBB), presenting a general framework that encompasses this bound and extends to the multiparameter case.

Consider a probability distribution $p(\mathbf{x}|\theta)$ that depends on a set of parameters $\theta = \{\theta_1, \dots, \theta_n\}$ and let $\hat{\theta}(\mathbf{x}) = \{\hat{\theta}_1(\mathbf{x}), \dots, \hat{\theta}_n(\mathbf{x})\}$ be the corresponding estimators. Let us set $\Delta\hat{\theta}_i = \hat{\theta}_i(\mathbf{x}) - \langle\hat{\theta}_i(\mathbf{x})\rangle$. Define the quantities

$$y_i = \Delta\hat{\theta}_i - A_{im}c_m, \quad (5.1)$$

where A_{im} is an arbitrary matrix with real entries and c_m is a vector that can depend on both $\mathbf{x}$ and θ . Throughout, we adopt the Einstein summation convention when no ambiguity arises. We construct a positive semidefinite matrix with matrix elements

$$y_i y_j = (\Delta\hat{\theta}_i - A_{im}c_m)(\Delta\hat{\theta}_j - A_{jp}c_p). \quad (5.2)$$

Averaging with respect to the probability distribution $p(\mathbf{x}|\theta)$ preserves the positive semidefiniteness,

$$\begin{aligned} \langle\Delta\hat{\theta}_i\Delta\hat{\theta}_j\rangle &\geq \langle\Delta\hat{\theta}_i c_p\rangle A_{jp} + A_{im}\langle\Delta\hat{\theta}_j c_m\rangle - A_{im}\langle c_m c_p\rangle A_{jp} \\ &= (TA^\top + AT^\top - AHA^\top)_{ij}, \end{aligned} \quad (5.3)$$

where the matrices T and C have elements

$$T_{ij} = \langle\Delta\hat{\theta}_i c_j\rangle, \quad C_{ij} = \langle c_i c_j\rangle. \quad (5.4)$$

The left-hand side in Eq. (5.3) is the covariance matrix. Our goal is to minimize the sum of its diagonal entries, which corresponds to minimizing the total variance. To achieve the tightest bound, we take the trace of both sides and differentiate with respect to the elements A_{ij} , yielding the optimal choice $A = TC^{-1}$. Substituting into (5.3), we obtain

$$\text{Cov}[\hat{\theta}(\mathbf{x})] \geq TC^{-1}T^\top. \quad (5.5)$$

Since for general estimators we have $\text{MSE}[\hat{\theta}(\mathbf{x})]_{ij} = \text{Cov}[\hat{\theta}(\mathbf{x})]_{ij} + \langle\delta\hat{\theta}_i\rangle\langle\delta\hat{\theta}_j\rangle$, where the deviation from the true parameter value is given $\delta\hat{\theta}_i = \hat{\theta}_i(\mathbf{x}) - \theta_i$ and $\delta\hat{\theta} = \{\delta\hat{\theta}_1, \delta\hat{\theta}_2, \dots, \delta\hat{\theta}_n\}$, we then have

$$\text{MSE}[\hat{\theta}(\mathbf{x})] \geq TC^{-1}T^\top + \langle\delta\hat{\theta}\rangle\langle\delta\hat{\theta}^\top\rangle, \quad (5.6)$$

which lets us derive different classical bounds. For unbiased estimators, the bias term disappears and T becomes independent of the estimator.

We now focus on the single-parameter case with unbiased estimators and choose the c_k to arrive at the CBB

$$c_k = \frac{1}{p(\mathbf{x}|\theta)} \frac{d^k p(\mathbf{x}|\theta)}{d\theta^k}. \quad (5.7)$$

The CBB of order k corresponds to the k th derivative. For instance, the second-order CBB takes the form

$$\text{MSE}[\hat{\theta}(\mathbf{x})] \geq \frac{1}{C_{11}} + \frac{C_{12}^2}{C_{11}(C_{11}C_{22} - C_{12}^2)}. \quad (5.8)$$

The term $1/C_{11}$ corresponds to the standard CCRB, while the extra term, which is always positive, is a correction that makes it tighter. In some cases, the CCRB is too optimistic because no unbiased estimator can achieve it. Nonetheless, an unbiased estimator might still exist that saturates the higher-order CBB. A classic example is estimating the square of the mean of a Gaussian distribution: Although the CCRB can be saturated for the mean itself, no unbiased estimator achieves it for the square. Yet, an unbiased estimator does exist that saturates the second-order CBB [57].

As formulated above, the theory applies to the full probability distribution; i.e., over v trials. For the term C_{11} , which is the CFI, we simply have that $C_{11} = v c_{11}$. The other quantities scale as [57]

$$\begin{aligned} C_{12} &= v c_{12}, \quad C_{13} = v c_{13}, \quad C_{22} = v[c_{22} + 2(v-1)c_{11}^2], \\ C_{23} &= v[c_{23} + 6(v-1)c_{12}c_{11}], \\ C_{33} &= v[c_{33} + 9(v-1)(c_{22}c_{11} + c_{12}^2) + 6(v-1)(v-2)c_{11}^3]. \end{aligned} \quad (5.9)$$

In this way, we get

$$\text{MSE}[\hat{\theta}(\mathbf{x})] \geq \frac{1}{\nu} \frac{1}{c_{11}} + \frac{1}{\nu^2} \frac{c_{12}^2}{2c_{11}^4} + \frac{1}{\nu^3} \frac{2c_{11}^2 c_{13}^2 + c_{12}^2 (6c_{11}^3 + 21c_{12}^2 - 3c_{11}c_{22} - 12c_{11}c_{13})}{12c_{11}^7} + O\left(\frac{1}{\nu^4}\right), \quad (5.10)$$

where we have included terms $O(1/\nu^3)$. Note that this bound, to order $O(1/\nu^i)$, is only applicable to estimators that are unbiased to the i th order.

The correction term of order $O(1/\nu^2)$ remains unchanged even when higher-order derivatives are added to the CBB (by introducing additional c_k). However, this is no longer true for terms of order $1/\nu^3$ and beyond.

Importantly, we note that the $O(1/\nu^2)$ term in the CBB is precisely the contribution $\gamma_{11}^2/[2\mathbf{F}_x^4(\theta)]$ appearing in $\Gamma(\theta)$. This term corresponds to the so-called *naming curvature* [69] and is one of the three positive terms that constitute $\Gamma(\theta)$. As a result, the CBB should be interpreted as a lower bound for the fundamental quantity $\Gamma(\theta)$.

Having understood the nature of the CBB, we can apply it to search for corrections to the QCRB. For the first POVM extension, the $O(1/\nu^2)$ coincides with $\Gamma(0)$; that is, it reduces to $\cot(\phi_+)^2/8m_2$. We fix $\phi_+ = -\pi/2$ and calculate the third correction; the result is $(3m_2^2 + m_4)^2/(384m_2^5)$.

For the second POVM extension, things are more complicated. First, we complete the POVM in the simplest possible way: adding a third output $p(x_3|\theta) = 1 - p(x_+|\theta) - p(x_-|\theta)$. The term of order $O(1/\nu^2)$ coincides again with $\Gamma(0)$. However, when calculating the correction of order $O(1/\nu^3)$, the expansion (5.10) is no longer useful. This is because it was derived by taking the limit $\nu \rightarrow \infty$, which implicitly assumes that no terms larger than ν exist. In this case we find that $c_{33} \rightarrow \infty$ as $\theta \rightarrow 0$. To address this, we substitute Eq. (5.9) in the general CBB, assuming the estimator is unbiased to order $O(1/\nu^3)$. We take the limit $\theta \rightarrow 0$ and then expand in powers of $1/\nu$, leading to

$$\text{MSE}[\hat{\theta}(\mathbf{x})]_{\theta=0} \geq \frac{1}{4m_2\nu} + \frac{\cot^2 \phi_+}{8m_2\nu^2} + \frac{\cot^2 \phi_+}{8m_2\nu^3} + O\left(\frac{1}{\nu^4}\right). \quad (5.11)$$

This indicates that by choosing the optimal measurement, all correction terms can be eliminated, making the bound tighter.

It is important to highlight that this case is problematic. The derivation of these expressions assumes an estimator unbiased to the order of the calculated correction. However, as shown in Fig. 6, this assumption does not hold (at least for the MLE). When θ is sufficiently small, the bias does not tend to zero, regardless of the value of ν . As we already discussed, dealing with probabilities that approach zero introduces significant difficulties.

B. Quantum Bhattacharyya bound

The quantum Bhattacharyya bound (QBB) was first introduced by Tsuda [70], who also extended other classical bounds useful for nondifferentiable models [71]. Later, Parthasarathy provided a rigorous proof that helped to derive quantum versions of several classical bounds [72].

Another approach, proposed by Brody and coworkers [73–75], used a geometric formulation of quantum

mechanics. In this framework, quantum states exist in a real Hilbert space with a complex structure. They derived a series of Bhattacharyya-style bounds for pure quantum states, which depend on the estimator used. To remove this dependence, special types of estimators are needed. More recently, these bounds were extended to general density matrices [76].

In Appendix C we provide an alternative proof that closely parallels the one for the CBB. We summarize here the results. We proceed as in the classical case, defining $Y_i = \Delta X_i - A_{im}L_m$, where $X_i = \sum_{\mathbf{x}} \hat{\theta}_i(\mathbf{x})\Pi_{\mathbf{x}}$ and $\Delta X_i = X_i - \langle \hat{\theta}_i(\mathbf{x}) \rangle = \sum_{\mathbf{x}} \Delta \hat{\theta}_i(\mathbf{x})\Pi_{\mathbf{x}}$ and L_m are operators and $\{\Pi_{\mathbf{x}}\}$ is the POVM representing the measurement. Operating much in the same way we arrive at

$$\text{Cov}[\hat{\theta}(\mathbf{x})] \geq TQ^{-1}T^{\top}, \quad (5.12)$$

where the matrices T and Q have elements

$$T_{ij} = \frac{1}{2} \text{Tr}[\varrho^{(v)}(\Delta X_i L_j^{\dagger} + L_j \Delta X_i)],$$

$$Q_{ij} = \frac{1}{2} \text{Tr}[\varrho^{(v)}(L_i L_j^{\dagger} + L_j L_i^{\dagger})], \quad (5.13)$$

where $\varrho^{(v)}$ is the density matrix corresponding to the ν trials. As in the classical case, we can add the bias to get the MSE.

Equation (5.12) provides a bound for the MSE, but how does it relate to the CBB and, more generally, to other classical bounds derived from Eq. (5.6)? The key idea is that T remains the same as in the classical case when L_k are chosen appropriately. For the QBB we choose them as the k th-order SLDs, ensuring they are Hermitian,

$$\frac{d^k \varrho_{\theta}^{(v)}}{d\theta^k} = \frac{1}{2} (L_k \varrho_{\theta}^{(v)} + \varrho_{\theta}^{(v)} L_k). \quad (5.14)$$

Also, one can demonstrate that the POVM consisting of eigenvectors of the operator $X = \langle \hat{\theta}(\mathbf{x}) \rangle + A_m L_m = \langle \hat{\theta}(\mathbf{x}) \rangle + T_k(Q^{-1})_{km} L_m$ saturate the QBB and the other single parameter bounds.

While we aim to apply this bound to pure states, a fundamental issue arises: These operators do not exist for pure states. This becomes evident when taking the second derivatives of the state ϱ_{θ} ,

$$\frac{d^2 \varrho_{\theta}}{d\theta^2} = -J_n^2 |\psi_{\theta}\rangle \langle \psi_{\theta}| - |\psi_{\theta}\rangle \langle \psi_{\theta}| J_n^2 + J_n |\psi_{\theta}\rangle \langle \psi_{\theta}| J_n. \quad (5.15)$$

Clearly, the last term cannot be produced by an operator of the form L_i as in Eq. (5.14), since it maps from outside the span of ϱ to outside the span of ϱ . But what does this imply? To gain insight, we introduce a small amount of white noise and then gradually let it approach zero,

$$\varrho_{\theta} = \frac{1}{1 + \varepsilon d} (|\psi_{\theta}\rangle \langle \psi_{\theta}| + \varepsilon \mathbb{1}), \quad (5.16)$$

where d is the dimension of the space. The analytical expressions for the L_k may be obtained by the use of the methods in Ref. [77]. It turns out that the nondiagonal terms are 0, so there

is no second-order correction. This is because $\text{Tr}(\varrho L_2 L_1)$ is an imaginary number. A similar thing happens for $i = 4, 6, \dots$. We continue and calculate the third-order corrections. In that case, we have

$$\frac{d^3 \varrho_\theta}{d\theta^3} = \frac{1}{1 + \epsilon d} (iJ_n^3 |\psi_\theta\rangle \langle \psi_\theta| + 3iJ_n |\psi_\theta\rangle \langle \psi_\theta| J_n^2 + \text{c. c.}), \quad (5.17)$$

which leads to the following values for q (we use again q instead of Q to denote the $\nu = 1$ case)

$$\begin{aligned} q_{11} &= \frac{4m_2}{(1 + 2\epsilon)(1 + d\epsilon)}, \\ q_{13} &= -\frac{4(3m_2^2 + m_4)}{(1 + 2\epsilon)(1 + d\epsilon)}, \\ q_{22} &= \frac{4m_2^2(3\epsilon^2 + 3\epsilon + 1) + m_4\epsilon(1 + \epsilon)}{\epsilon(1 + \epsilon)(1 + 2\epsilon)(1 + d\epsilon)}, \\ q_{33} &= \frac{-18m_2^3 + 6m_4m_2(10\epsilon + 3) - 2m_3^2(20\epsilon + 9) + 4m_6\epsilon}{\epsilon(1 + 2\epsilon)(1 + d\epsilon)}. \end{aligned} \quad (5.18)$$

The rest of the terms involving L_2 are zero. See Appendix D for the scaling of Q .

We calculate the third-order corrections,

$$(Q^{-1})_{11} = \frac{1}{Q_{11}} \left(1 + \frac{Q_{13}^2}{Q_{33} - Q_{13}^2} \right). \quad (5.19)$$

To analyze the asymptotic expansion in powers of ν , it is crucial to carefully consider the order in which limits are taken. If we first take the limit $\epsilon \rightarrow 0$, then the third-order corrections vanish. This behavior mirrors what we observed with the CBB in the second POVM extension, where choosing $\phi_+ = -\pi/2$ canceled these corrections.

On the other hand, if we assume ϵ small, but nonzero, and apply Eq. (5.10) before taking $\epsilon \rightarrow 0$, then we recover a correction that matches the result from the first POVM extension:

$$\text{MSE}[\hat{\theta}(\mathbf{x})]_{\theta=0} \geq \frac{1}{4m_2\nu} + \frac{(3m_2^2 + m_4)^2}{384m_2^5\nu^3} + O\left(\frac{1}{\nu^4}\right). \quad (5.20)$$

To interpret this, recall that the element Q_{11} corresponds to the QFI. The QBB must be attained by measurements that also achieve the QFI, while minimizing higher-order corrections. This is precisely what we did in the previous section.

When ϵ is small but finite, the scenario resembles the first POVM extension, where no probabilities vanish. This explains why the QBB corrections match the case with $\phi_+ = -\pi/2$, which cancels the corrections of order $O(1/\nu^2)$.

Conversely, in the second POVM extension, certain probabilities vanish as $\epsilon \rightarrow 0$, preventing higher-order corrections when $\phi_+ = -\pi/2$. This corresponds to $\epsilon \rightarrow 0$ before taking $\nu \rightarrow \infty$.

This interplay between limits offers an intuitive explanation of the observed behavior, although a more rigorous treatment is needed to fully characterize the underlying relationships.

This brings us to the practical significance of the CBB (or its quantum analog, the QBB). In the classical case, the

CBB naturally arises within the framework of second-order asymptotic theory. Specifically, the $1/\nu^2$ term in the CBB is one of the three components that define the full second-order asymptotic behavior.

The key advantage of the QBB is its ability to determine the minimum possible value of the $1/\nu^2$ term in the CBB. Moreover, it helps to identify the measurements that achieve this minimum within the class of measurements that saturate the QFI.

However, the QBB is not guaranteed to exist in all scenarios. In particular, problems arise when the span of the quantum state changes with the parameter. One possible workaround is to introduce a small amount of white noise and consider the limit as it vanishes. This regularization approach has been shown to recover the corrections associated with both POVM extensions, though a deeper analysis is required to fully understand its scope and limitations.

In summary, the QBB is especially useful for full-rank density matrices. In cases where the rank varies, its applicability is less clear and requires deeper investigation.

VI. CONCLUSIONS AND OUTLOOK

In this work, we have derived $O(1/\nu^2)$ and $O(1/\nu^3)$ corrections to the MSE in quantum metrology, where ν is the number of trials. These corrections improve the accuracy of MSE estimates in the preasymptotic regime.

Our results highlight that these corrections depend on the estimator used. While the QCRB sets a fundamental asymptotic limit, estimator-specific effects emerge in the preasymptotic regime. If we consider estimators unbiased up to order $O(1/\nu)$, then the second-order corrections simplify, revealing a fundamental quantity akin to the QCRB.

We also showed that standard higher-order correction methods may fail when the density matrix's span depends on the parameter. Despite this, we demonstrated that these corrections remain computable. Importantly, they allow us to distinguish measurement strategies that appear equivalent under the QCRB. For instance, measurements achieving the QFI may still exhibit different preasymptotic performance. Additionally, for optimal state families used in detecting unitary transformations, higher-order corrections help refine optimality beyond the asymptotic limit.

Furthermore, we explored the connection between these corrections and the CBB, a tighter bound than the CCRB. We demonstrated that the BB provides a lower limit for these $O(1/\nu^2)$ corrections. Notably, like the QCRB, it has a quantum extension that offers a POVM-independent lower bound on preasymptotic corrections, making it a powerful and easy-to-use tool.

Looking ahead, several research directions arise. A key next step is generalizing the $O(1/\nu^2)$ correction formulas to probability distributions that do not sum to one, a common scenario in practice. This could follow the approach in Ref. [63] but without assuming probability normalization, simplifying calculations by avoiding explicit series inversion.

Another significant direction is extending these results to arbitrary quantum processes and mixed states. Investigating

measurement parametrizations that saturate the QFI in broader scenarios and optimizing them using second-order corrections is also a natural extension.

The role of POVM extensions in parameter estimation accuracy deserves further study. A deeper understanding of their most general forms and implications would be valuable.

Additionally, while our work focuses on frequentist quantum metrology, these techniques could be adapted for Bayesian methods. Extending this approach to multiparameter quantum metrology is, in principle, feasible but potentially challenging, as the asymptotic case remains unresolved and is an active research area [78–81].

In conclusion, while we focused on unitary transformations and pure states, our methods provide a strong analytical foundation for quantum metrology beyond the asymptotic regime. By enabling finite-resource studies without relying on numerical simulations, this approach holds great potential for further theoretical advancements.

ACKNOWLEDGMENTS

We acknowledge discussions with A. Smerzi, L. Pezzè, A. Datta, M. Barbieri, I. Gianani, and M. Evans at an early stage of this work. This publication has received financial support from the QuantEra program (Project ApresSF), from the Agencia Estatal de Investigación (Grant No. PID2021-127781NB-I00), and from Palacký University (Grant No. IGA PRF 2025_005). A.Z.G. acknowledges that the NRC headquarters is located on the traditional unceded territory of the Algonquin Anishinaabe and Mohawk people, as well as support from the NRC Quantum Sensors Challenge program. L.L.S.-S. was supported in part by the Grant No. NSF PHY-1748958 to the Kavli Institute for Theoretical Physics (KITP).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: FROM FIDELITY TO HIGHER-ORDER AVERAGES

The fidelity between a state and a slightly rotated version of the state is given by $\mathcal{F}(\theta|\mathbf{n}) \equiv |\langle\psi|e^{i\theta J_{\mathbf{n}}}|\psi\rangle|^2 = 1 - \frac{1}{4}\theta^2 \mathbf{Q}_{\psi}(\mathbf{n}) + O(\theta^3)$. This means that the states that change most rapidly with small θ are those with maximal QFI and that the states that change the most on average for all rotation axes $\mathbf{n}$ are those with the maximal values of $\int_{S_2} \mathbf{Q}_{\psi}(\theta|\mathbf{n})d\Omega$, which are the states with isotropic first-order moments $\langle J_{\mathbf{n}} \rangle = \text{const} = 0$. Within this family of states that have such a maximal average QFI, which states have their fidelities decrease the most to next leading order in θ ?

We begin by writing

$$\mathcal{F}(\theta) \equiv \frac{1}{4\pi} \int_{S_2} \mathcal{F}(\theta|\mathbf{n})d\Omega = \frac{1}{4\pi} \int_{S_2} \left\{ \left[\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} \langle\psi|(\theta J_{\mathbf{n}})^{2k}|\psi\rangle \right]^2 + \left[\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \langle\psi|(\theta J_{\mathbf{n}})^{2k+1}|\psi\rangle \right]^2 \right\} d\Omega. \quad (\text{A1})$$

The integrals are all over terms of the form $\int_{S_2} n_1^j n_2^k n_3^l d\Omega$, which equals $\frac{2\Gamma(\frac{j+1}{2})\Gamma(\frac{k+1}{2})\Gamma(\frac{l+1}{2})}{\Gamma(\frac{1}{2}(j+k+l+3))}$ when each of j , k , and l are even and vanishes otherwise. We can write $\mathcal{F}(\theta) = \sum_{n=0}^{\infty} f_n \theta^{2n}$ and note the first few terms:

$$\begin{aligned} f_0 &= 1, \\ f_1 &= \frac{1}{4\pi} \int_{S_2} (\langle J_{\mathbf{n}} \rangle^2 - \langle J_{\mathbf{n}}^2 \rangle) d\Omega = \frac{1}{3} \sum_{i=1}^3 (\langle J_i \rangle^2 - \langle J_i^2 \rangle) = -\frac{1}{16\pi} \int_{S_2} \mathbf{Q}_{\psi}(\mathbf{n}) d\Omega, \\ f_2 &= \frac{1}{4\pi} \int_{S_2} \frac{1}{12} (\langle J_{\mathbf{n}}^4 \rangle - 4\langle J_{\mathbf{n}}^3 \rangle \langle J_{\mathbf{n}} \rangle + 3\langle J_{\mathbf{n}}^2 \rangle^2) d\Omega, \end{aligned} \quad (\text{A2})$$

and so on. We see that f_0 is the same for any state. The sum $\sum_{i=1}^3 \langle J_i^2 \rangle$ is constant for any state because $\sum_{i=1}^3 J_i^2$ is proportional to the identity operator, so f_1 is minimized by states with $\langle J_1 \rangle = \langle J_2 \rangle = \langle J_3 \rangle = 0$, which are first-order isotropic ($\langle J_{\mathbf{n}} \rangle = \text{const}$). Among these states, since $\int_{S_2} J_{\mathbf{n}}^4 d\Omega$ is proportional to the identity operator [see Eq. (A8) onwards], f_2 simplifies to

$$f_2 = \text{const} + \frac{1}{16\pi} \int_{S_2} \langle J_{\mathbf{n}}^2 \rangle^2 d\Omega \quad (\text{A3})$$

and

$$\frac{1}{4\pi} \int_{S_2} \langle J_{\mathbf{n}}^2 \rangle^2 d\Omega \geq \left(\frac{1}{4\pi} \int_{S_2} \langle J_{\mathbf{n}}^2 \rangle d\Omega \right)^2 = \text{const} \quad (\text{A4})$$

has the same lower bound regardless of the state due to $\int_{S_2} J_{\mathbf{n}}^2 d\Omega$ being proportional to identity [see Eq. (A8) onwards]; this lower bound is achieved if and only if the state obeys $\langle J_{\mathbf{n}}^2 \rangle = \text{const}$. This was found using Jensen's inequality that the average of a square is greater than the square of the average, with equality if and only if the quantity being averaged is constant, as above.

To summarize, all states have the same f_0 , states that are isotropic to first order have the smallest f_1 , states that are isotropic to first and second order have smallest f_2 , and so on.

We proceed to prove the general case by induction. Consider that the subleading order in the expansion of $\mathcal{F}(\theta)$ to order $\theta^{2(n-1)}$ is minimized by states that are isotropic to order $n-1$; i.e., by states $|\psi\rangle$ with $\langle J_n^k \rangle = \text{const}$ for all $k \in \{0, 1, \dots, n-1\}$. Then, the terms of order θ^{2n} in the integrand are

$$I_n = \sum_{k=0}^n \frac{(-1)^k}{(2k)!} \frac{(-1)^{n-k}}{(2n-2k)!} \langle \psi | J_n^{2k} | \psi \rangle \langle \psi | J_n^{2n-2k} | \psi \rangle + \sum_{k=0}^{n-1} \frac{(-1)^k}{(2k+1)!} \frac{(-1)^{n-k-1}}{(2n-2k-1)!} \langle \psi | J_n^{2k+1} | \psi \rangle \langle \psi | J_n^{2n-2k-1} | \psi \rangle. \quad (\text{A5})$$

Since all the expectation values $\langle J_n^k \rangle \equiv c_k$ to order k are independent from $\mathbf{n}$ for $k < n$, all of the c_k with odd values of $k < n$ will vanish. The second sum only retains terms with $2k+1 = 2n-2k-1 = n$ and so only remains if n is odd. The first sum is always a constant times a single expectation value, if n is odd, yielding another constant after integration. We thus have the possibilities:

$$I_n = \begin{cases} \frac{(-1)^{(n-1)/2}}{n!} \langle J_n^n \rangle^2 + 2 \sum_{k=0}^{(n-1)/2} \frac{(-1)^k}{(2k)!} \frac{(-1)^{n-k}}{(2n-2k)!} c_{2k} \langle J_n^{2n-2k} \rangle, & n \text{ odd}, \\ \frac{(-1)^{n/2}}{n!} \langle J_n^n \rangle^2 + 2 \sum_{k=0}^{n/2-1} \frac{(-1)^k}{(2k)!} \frac{(-1)^{n-k}}{(2n-2k)!} c_{2k} \langle J_n^{2n-2k} \rangle, & n \text{ even}. \end{cases} \quad (\text{A6})$$

Performing the integral over the angular coordinates $f_n = \frac{1}{4\pi} \int_{S_2} I_n d\Omega$ again uses that $\int_{S_2} (\mathbf{J} \cdot \mathbf{n})^{2n-2k} d\Omega$ is proportional to the identity operator, such that, for all states isotropic to order $n-1$,

$$f_n = \frac{1}{n!^2} \frac{1}{4\pi} \int_{S_2} \langle J_n^n \rangle^2 d\Omega + \text{const}. \quad (\text{A7})$$

Such a quantity is uniquely minimized by states with $\langle J_n^n \rangle = \text{const}$. This proves that a state minimizing the axis-averaged fidelity between itself and a rotated version of itself to every order in rotation angle θ up to and including θ^{2n} is a state that is isotropic to order n . None of this proof relies on a particular group structure and so these results apply to any unitary transformation.

Finally, we prove that all averages over the unit sphere S_2 of the type

$$\overline{\langle J_n^k \rangle} = \frac{1}{4\pi} \int_{S_2} \langle J_n^k \rangle d\Omega, \quad (\text{A8})$$

are independent of the state. Consider the operator

$$O = \int_{S_2} J_n^k d\Omega. \quad (\text{A9})$$

Applying an arbitrary rotation $R \in \text{SU}(2)$, we have

$$ROR^\dagger = \int_{S_2} RJ_n^k R^\dagger d\Omega = \int_{S_2} (RJ_n R^\dagger)(RJ_n R^\dagger) \cdots (R^\dagger J_n R) d\Omega = \int_{S_2} J_{n'}^k d\Omega, \quad (\text{A10})$$

where $J_{n'}$ is simply the rotated angular momentum operator. Because we are integrating over all directions $\mathbf{n}$, $ROR^\dagger = O$. But the only operator that remains invariant under all rotations is an operator proportional to the identity [82], thus

$$\overline{\langle J_n^k \rangle} = \text{const}. \quad (\text{A11})$$

APPENDIX B: SERIES EXPANSION OF THE MAXIMUM LIKELIHOOD ESTIMATOR

We expand the two outcomes probabilities around $\theta = 0$, collecting terms up to $O(\theta^6)$; we obtain

$$\begin{aligned} p(x_+|\theta) &= \cos^2\left(\frac{\phi_+}{2}\right) + \theta\sqrt{m_2} \sin \phi_+ - \theta^2 m_2 \cos \phi_+ - \frac{\theta^3(3m_2^2 + m_4) \sin \phi_+}{6\sqrt{m_2}} \\ &\quad + \frac{\theta^4[3m_2^3(1 + \cos \phi_+) - m_2 m_4(3 - 5 \cos \phi_+) + 3m_3^2(1 - \cos \phi_+)]}{24m_2} + \frac{\theta^5(15m_2 m_4 - 10m_3^2 + m_6) \sin \phi_+}{120\sqrt{m_2}} \\ &\quad + O(\theta^6), \\ p(x_-|\theta) &= \sin^2\left(\frac{\phi_+}{2}\right) - \theta\sqrt{m_2} \sin \phi_+ + \theta^2 m_2 \cos \phi_+ + \frac{\theta^3(3m_2^2 + m_4) \sin \phi_+}{6\sqrt{m_2}} \\ &\quad + \frac{\theta^4[3m_2^3(1 - \cos \phi_+) - m_2 m_4(3 + 5 \cos \phi_+) + 3m_3^2(1 + \cos \phi_+)]}{24m_2} - \frac{\theta^5(15m_2 m_4 - 10m_3^2 + m_6) \sin \phi_+}{120\sqrt{m_2}} \\ &\quad + O(\theta^6). \end{aligned} \quad (\text{B1})$$

1. First POVM extension

We now focus on the first POVM extension, in which only two outcomes are possible. Their probabilities are $p(x_+|\theta)$ and $p(x_-|\theta) = 1 - p(x_+|\theta)$. The MLE is obtained by using (4.9). We introduce the shorthand $\delta_+ \equiv \nu_+ - \nu/2$ and note that $\nu_- = \nu - \nu_+$. Fixing $\phi_+ = -\pi/2$, as determined earlier by the minimization of $\Gamma(0)$, we then obtain:

$$\begin{aligned} & \frac{\nu_+}{p(x_+|\theta)} \frac{dp(x_+|\theta)}{d\theta} + \frac{\nu_-}{p(x_-|\theta)} \frac{dp(x_-|\theta)}{d\theta} \\ &= -4\delta_+ \sqrt{m_2} - 4m_2\nu\theta + \frac{2\delta_+(m_4 - 5m_2^2)}{\sqrt{m_2}} \theta^2 + \left[2m_2^2(\delta_+ - 4\nu) + \frac{2m_3^2\delta_+}{m_2} + \frac{2}{3}m_4(4\nu - 3\delta_+) \right] \theta^3 \\ &+ \frac{[5m_2m_4(13\delta_+ - 3\nu) + 3m_3^2(5\nu - 48\delta_+) + 5m_3^2(2\delta_+ + 3\nu) - m_6\delta_+]}{6\sqrt{m_2}} \theta^4 \\ &+ \frac{\{m_2^4(360\delta_+ - 570\nu) + 15m_2^2m_4(22\nu - 25\delta_+) + m_2[10m_3^2(37\delta_+ + 6\nu) - 15m_3m_5\delta_+ + 10m_4^2(\delta_+ - \nu) + m_6(5\delta_+ - 6\nu)]\}}{30m_2} \theta^5 \\ &+ O(\theta^6). \end{aligned} \quad (B2)$$

Formally, this can be written as

$$0 = a_0 + a_1\theta_{\text{MLE}} + a_2\theta_{\text{MLE}}^2 + a_3\theta_{\text{MLE}}^3 + a_4\theta_{\text{MLE}}^4 + a_5\theta_{\text{MLE}}^5 + O(\theta_{\text{MLE}}^6), \quad (B3)$$

and applying the Lagrange-Bürmann inversion formula [64,65] provides an explicit power-series expansion for θ_{MLE} . Squaring this series yields θ_{MLE}^2 ,

$$\theta_{\text{MLE}}^2 = \frac{a_0^2}{a_1^2} + \frac{2a_0^3a_2}{a_1^4} + \frac{a_0^4(5a_2^2 - 2a_1a_3)}{a_1^6} + \frac{2a_0^5(7a_2^3 - 6a_1a_2a_3 + a_1^2a_4)}{a_1^8} + \frac{a_0^6(42a_2^4 - 56a_1a_2^2a_3 + 14a_1^2a_2a_4 + 7a_1^2a_3^2 - 2a_1^3a_5)}{a_1^{10}}. \quad (B4)$$

Identifying terms this yields

$$\theta_{\text{MLE}}^2(\mathbf{x}) = \frac{\delta_+^2}{m_2\nu^2} + \frac{\delta_+^4(3m_2^2 + m_4)}{3m_2^3\nu^4} - \frac{\delta_+^5(m_3^2 - m_2m_4 + m_3^2)}{4m_2^4\nu^5} + \frac{\delta_+^6[315m_2^4 + 165m_2^2m_4 + m_2(30m_3^2 - 3m_6) + 35m_4^2]}{180m_2^5\nu^6}. \quad (B5)$$

Because δ_+ follows a centered binomial distribution (i.e., a fair-coin distribution), its even moments are $\langle \delta_+^2 \rangle = \nu/4$, $\langle \delta_+^4 \rangle = \nu(3\nu - 2)/16$, and $\langle \delta_+^6 \rangle = [15(\nu - 2)\nu + 16]/64$, while all odd moments vanish. Substituting these averages and retaining terms up to $O(1/\nu^3)$ we obtain

$$\langle \hat{\theta}_{\text{MLE}}^2(\mathbf{x}) \rangle = \frac{1}{4m_2\nu} + \frac{3m_2^2 + m_4}{16m_2^3\nu^2} + \frac{219m_2^4 + 133m_2^2m_4 + m_2(30m_3^2 - 3m_6) + 35m_4^2}{768m_2^5\nu^3} + O(1/\nu^4). \quad (B6)$$

To reach this expression we require the series for θ_{MLE} up to fifth order and the $1/\nu$ expansion up to sixth order; higher-order terms do not contribute. Because the estimator is still biased, a positive $1/\nu^2$ term arising from the derivative of the bias survives, highlighting the advantage of employing the bias-corrected estimator.

To obtain the MSE of the bias-corrected estimator we expand $b(\hat{\theta}) = b(0) + a_1\hat{\theta} + a_2\hat{\theta}^2 + a_3\hat{\theta}^3$ in $\langle (\hat{\theta}(\mathbf{x}) - b(\hat{\theta}))/\nu \rangle^2$. Since $b(0) = 0$ for $\phi_+ = -\pi/2$ we obtain

$$\langle \hat{\theta}^2(\mathbf{x}) \rangle \left(1 + \frac{a_1^2}{\nu^2} - \frac{2a_1}{\nu} \right) + \langle \hat{\theta}^3(\mathbf{x}) \rangle \left(\frac{2a_1a_2}{\nu^2} - \frac{2a_2}{\nu} \right) + \langle \hat{\theta}^4(\mathbf{x}) \rangle \left(\frac{a_2^2}{\nu^2} + \frac{2a_1a_3}{\nu^2} - \frac{2a_3}{\nu} \right) + \langle \hat{\theta}^5(\mathbf{x}) \rangle \frac{2a_2a_3}{\nu^2} + \langle \hat{\theta}^6(\mathbf{x}) \rangle \frac{a_3^2}{\nu^2}. \quad (B7)$$

While the intermediate algebra is omitted here, the same expansion technique used for $\langle \hat{\theta}^2(\mathbf{x}) \rangle$ can be repeated to obtain $\langle \hat{\theta}^3(\mathbf{x}) \rangle$, $\langle \hat{\theta}^4(\mathbf{x}) \rangle$, and higher moments,

$$\langle \hat{\theta}_{\text{MLE}}^3(\mathbf{x}) \rangle = \frac{45(m_2^3 + m_3^2 - m_2m_4)}{512m_2^{9/2}\nu^3} + O(1/\nu^4); \quad \langle \hat{\theta}_{\text{MLE}}^4(\mathbf{x}) \rangle = \frac{3}{16m_2^2\nu^2} + O(1/\nu^3). \quad (B8)$$

Thus, the only contributions that remain at $O(1/\nu^3)$ are

$$\langle (\hat{\theta}_{\text{MLE}}^*(\mathbf{x}))^2 \rangle = \langle \hat{\theta}_{\text{MLE}}^2(\mathbf{x}) \rangle \left(1 - \frac{2a_1}{\nu} + \frac{a_1^2}{\nu^2} \right) - \langle \hat{\theta}_{\text{MLE}}^4(\mathbf{x}) \rangle \frac{2a_3}{\nu}. \quad (B9)$$

The $2a_1/\nu$ term cancels the contribution from the bias derivative at $O(1/\nu^2)$. The bias is given by

$$b(\theta) = \frac{1}{8} \left(3 + \frac{m_4}{m_2^2} \right) \theta + \frac{3(m_2^3 + m_3^2 - m_2m_4)}{16m_2^{5/2}} \theta^2 + \frac{(9m_2^4 + 15m_2^2m_4 + m_2(10m_3^2 - m_6) + 9m_4^2)}{48m_2^3} \theta^3 + O(\theta^4). \quad (B10)$$

Inserting these results reproduces Eq. (4.14).

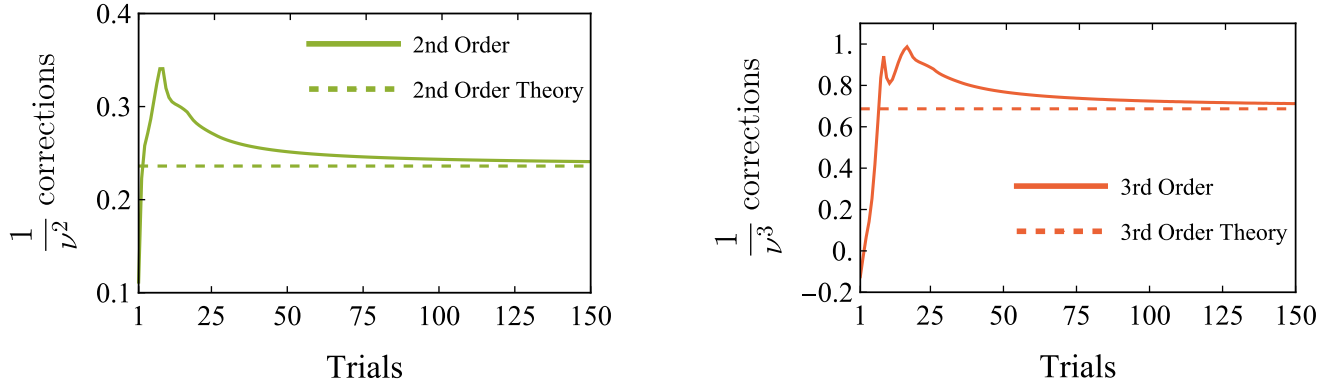


FIG. 8. Numerical validation of the asymptotic coefficients for the first POVM extension, using the uncorrected MLE. The quantum state is that employed in the main text.

For completeness, we present the general ϕ_+ -dependent expressions, valid at finite θ , for all quantities entering Eq. (4.2),

$$\begin{aligned}\Gamma(\theta) &= \frac{\cot^2 \phi_+}{8m_2} + \theta \frac{\cot \phi_+ [\mathbf{m}_2^2 \csc^2 \phi_+ (\cos 2\phi_+ + 7) + 2\mathbf{m}_4]}{16\mathbf{m}_2^{5/2}} + O(\theta^2), \\ b(\theta) &= \frac{\cot \phi_+}{4\sqrt{m_2}} + \frac{\theta}{8} \left(\frac{\mathbf{m}_4}{m_2^2} + 4 \csc^2 \phi_+ - 1 \right) + O(\theta^2), \\ \frac{1}{F_x(\theta)} &= \frac{1}{4m_2} + \frac{\theta^2}{4} \left(\frac{\mathbf{m}_4}{m_2^2} - 1 \right) + O(\theta^3).\end{aligned}\quad (\text{B11})$$

2. Second POVM extension

For the second POVM extension, the use of Eq. (4.2) produces incorrect second-order terms, so we derive them explicitly. Because ϕ_+ is no longer fixed, we redefine $\delta_+ \equiv \nu_+ - \nu \cos^2(\phi_+/2)$. Following the same steps as before, but after inserting the series into Eq. (B4) we expand in powers of $1/\nu$ to remove δ_+ from the denominators. This procedure yields

$$\theta_{\text{MLE}}^2(\mathbf{x}) = \frac{\delta_+^2 \csc^2(\phi_+)}{m_2 \nu^2} + \frac{2\delta_+^3 \cot(\phi_+) \csc^3(\phi_+)}{m_2 \nu^3} + \frac{\delta_+^4 \csc^4(\phi_+) \{3\mathbf{m}_2^3 [7 \cos(2\phi_+) + 13] \csc^2(\phi_+) - 2\mathbf{m}_4 m_2 + 6\mathbf{m}_3^2\}}{12\mathbf{m}_2^4 \nu^4}. \quad (\text{B12})$$

To evaluate the MSE we require the series for θ_{MLE} and the $1/\nu$ expansion only up to fourth order; higher orders leave the result unchanged. The central moments are $\langle \delta_+^2 \rangle = \frac{\nu}{4} \sin^2 \phi_+$, $\langle \delta_+^3 \rangle = -\frac{\nu}{4} \sin^2 \phi_+ \cos \phi_+$, and $\langle \delta_+^4 \rangle = \frac{\nu}{32} \sin^2 \phi_+ [3(\nu - 2) \cos(2\phi_+) - 3\nu - 2]$. Substituting these moments and keeping terms through $O(1/\nu^2)$ reproduces Eq. (4.15).

For completeness, Eq. (4.15) at the optimal setting $\phi_+ = -\pi/2$ reads

$$\text{MSE}[\hat{\theta}_{\text{MLE}}(\mathbf{x})]_{\theta=0} = \frac{1}{4m_2 \nu} + \frac{9\mathbf{m}_2^3 - m_2 \mathbf{m}_4 + 3\mathbf{m}_3^2}{32\mathbf{m}_2^4 \nu^2} + O(1/\nu^3). \quad (\text{B13})$$

Proceeding in the same fashion one can derive the $O(1/\nu^3)$ corrections, although the resulting expression is rather cumbersome. Their evaluation requires carrying the series for both θ_{MLE} and the $1/\nu$ expansion to sixth order:

$$\begin{aligned}\frac{1}{4m_2 \nu} &+ \frac{9\mathbf{m}_2^3 - m_2 \mathbf{m}_4 + 3\mathbf{m}_3^2}{32\mathbf{m}_2^4 \nu^2} + \frac{1719\mathbf{m}_2^6 - 116\mathbf{m}_2^4 \mathbf{m}_4 + 6\mathbf{m}_2^3 (223\mathbf{m}_3^2 + 3\mathbf{m}_6) - 5\mathbf{m}_2^2 (17\mathbf{m}_4^2 + 18\mathbf{m}_3 \mathbf{m}_5) - 30\mathbf{m}_2 \mathbf{m}_3^2 \mathbf{m}_4 + 315\mathbf{m}_3^3}{3072\mathbf{m}_2^7 \nu^3} \\ &+ O\left(\frac{1}{\nu^4}\right).\end{aligned}\quad (\text{B14})$$

Finally, we reproduce the erroneous second-order expression obtained by applying Eq. (4.2),

$$\Gamma(0) + \frac{2b'(0)}{F_x(0)} + b(0)^2 = \frac{\mathbf{m}_2^2 \csc^2(\phi_+) (2 \cos(2\phi_+) + 5) + \mathbf{m}_4}{16\mathbf{m}_2^3}. \quad (\text{B15})$$

3. Numerically testing the coefficients of the asymptotic expansions

To validate the expansion coefficients for the first POVM extension, Fig. 8 shows the numerical evaluation of the quantities,

$$\nu^2 \left(\langle \hat{\theta}^2(\mathbf{x}) \rangle - \frac{1}{m_2 \nu} \right); \quad \nu \left[\nu^2 \left(\langle \hat{\theta}^2(\mathbf{x}) \rangle - \frac{1}{m_2 \nu} \right) - \frac{3\mathbf{m}_2^2 + \mathbf{m}_4}{16\mathbf{m}_2^3} \right]. \quad (\text{B16})$$

The numerics converge to the analytical predictions, confirming the correctness of the series coefficients.

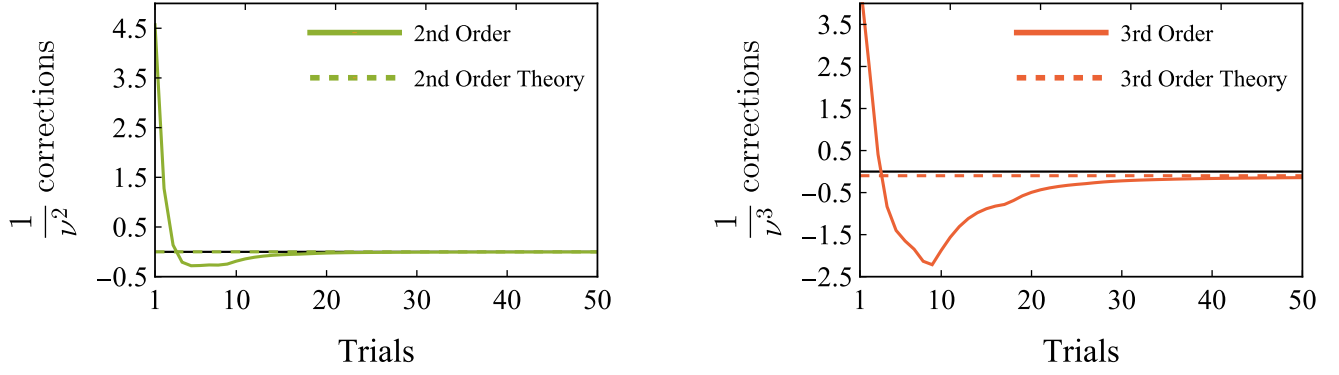


FIG. 9. Numerical validation of the asymptotic coefficients for the bias-corrected MLE in the first POVM extension. The quantum state is that employed in the main text.

Figure 9 repeats the numerical test for the bias-corrected estimator. As predicted, the second-order term vanishes, leaving the third-order correction as the leading contribution. The same verification for the second POVM extension is presented in Fig. 10.

APPENDIX C: PROOF OF THE QUANTUM BHATTACHARYYA BOUND

Let $\{|\omega_{\mathbf{x}}\rangle\}$ be a list of arbitrary states, $\{a_{\mathbf{x}}\}$ a list of arbitrary constants, and $\{\Pi, \hat{\theta}(\mathbf{x})\} = \{\Pi_{\mathbf{x}}, \hat{\theta}_i(\mathbf{x})\}$ a POVM and a set of estimators for the different parameters. By expanding

$$\sum_{\mathbf{x}} \left[\langle \omega_{\mathbf{x}} | a_{\mathbf{x}}^* - \left(\sum_{\mathbf{x}'} \langle \omega_{\mathbf{x}'} | a_{\mathbf{x}'}^* \Pi_{\mathbf{x}'} \right) \right] \Pi_{\mathbf{x}} \left[a_{\mathbf{x}} | \omega_{\mathbf{x}} \rangle - \left(\sum_{\mathbf{x}'} a_{\mathbf{x}'} \Pi_{\mathbf{x}'} | \omega_{\mathbf{x}'} \rangle \right) \right] \geq 0 \quad (\text{C1})$$

and using $\sum_{\mathbf{x}} \Pi_{\mathbf{x}} = \mathbb{1}$ we arrive at

$$\Pi_{\mathbf{x}} \delta_{\mathbf{x}\mathbf{x}'} - \Pi_{\mathbf{x}} \Pi_{\mathbf{x}'} \geq 0, \quad (\text{C2})$$

with no summation over the index $\mathbf{x}$ implied. This matrix of operators should be understood as an operator acting on the tensor product space $\mathcal{H} \otimes \mathbb{C}^n$, where $\mathcal{H}$ is the Hilbert space in which the states exist and n is the number of elements in the POVM [78], that is, it acts on vectors of kets. More precisely, it would be written as

$$\sum_{\mathbf{x}\mathbf{x}'} \Pi_{\mathbf{x}} \delta_{\mathbf{x}\mathbf{x}'} \otimes |\mathbf{x}\rangle \langle \mathbf{x}'| - \Pi_{\mathbf{x}} \Pi_{\mathbf{x}'} \otimes |\mathbf{x}\rangle \langle \mathbf{x}'| \geq 0, \quad (\text{C3})$$

where $|\mathbf{x}\rangle$ is an orthonormal basis of $\mathbb{C}^n$. We now act on the right with $\mathbb{1} \otimes \Delta \hat{\theta}_i(\mathbf{x})$, and on the left with its conjugate transpose. We do the same with $\sqrt{\varrho^{(v)}} \otimes \mathbb{1}$. This preserves the positive-definiteness of the operators and the inequality. With that we

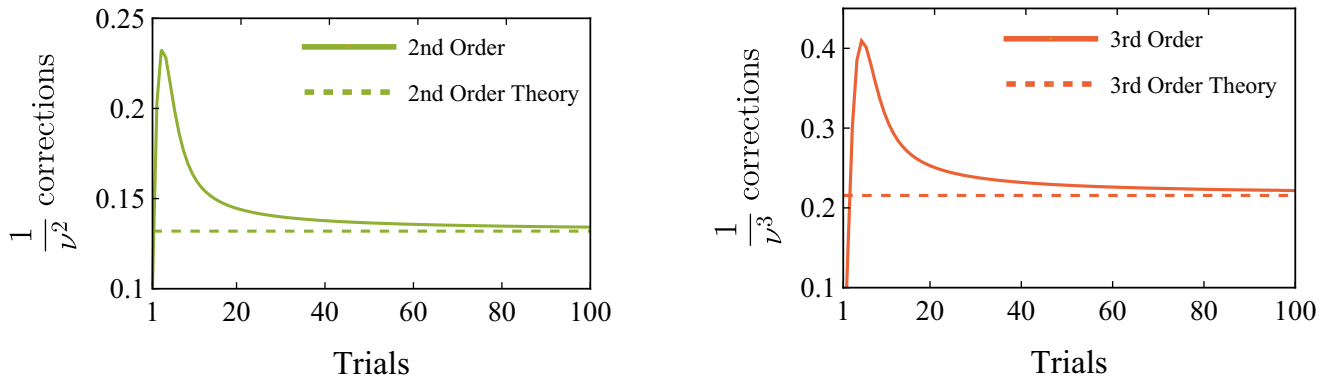


FIG. 10. Numerical validation of the asymptotic coefficients for the second POVM extension. The quantum state is that employed in the main text.

arrive at

$$\sum_{\mathbf{x}} \Delta \hat{\theta}_i(\mathbf{x}) \Delta \hat{\theta}_j(\mathbf{x}) \sqrt{\varrho^{(v)}} \Pi_{\mathbf{x}} \sqrt{\varrho^{(v)}} \geq \sum_{\mathbf{x}, \mathbf{x}'} \Delta \hat{\theta}_i(\mathbf{x}) \Delta \hat{\theta}_j(\mathbf{x}') \sqrt{\varrho^{(v)}} \Pi_{\mathbf{x}} \Pi_{\mathbf{x}'} \sqrt{\varrho^{(v)}}. \quad (\text{C4})$$

We now take the partial trace with the respect to $\mathcal{H}$ and get

$$\text{Cov}[\hat{\theta}(\mathbf{x}), \Pi]_{ij} = \sum_{\mathbf{x}} \Delta \hat{\theta}_i(\mathbf{x}) \Delta \hat{\theta}_j(\mathbf{x}) \text{Tr}(\Pi_{\mathbf{x}} \varrho^{(v)}) \geq \sum_{\mathbf{x}, \mathbf{x}'} \Delta \hat{\theta}_i(\mathbf{x}) \Delta \hat{\theta}_j(\mathbf{x}') \text{Tr}(\varrho^{(v)} \Pi_{\mathbf{x}} \Pi_{\mathbf{x}'})) = \text{Tr}(\varrho^{(v)} \Delta X_i \Delta X_j), \quad (\text{C5})$$

where we define the estimator operators $X_i = \sum_{\mathbf{x}} \hat{\theta}_i(\mathbf{x}) \Pi_{\mathbf{x}}$ and $\Delta X_i = X_i - \langle \hat{\theta}_i(\mathbf{x}) \rangle = \sum_{\mathbf{x}} \Delta \hat{\theta}_i(\mathbf{x}) \Pi_{\mathbf{x}}$.

At this point, we may proceed in two ways, depending on whether we want to obtain the symmetric version or the nonsymmetric version of the quantum bounds. First, we show the symmetric version. For that we notice that if we have two matrices such that $A \geq B$, then it implies that $\bar{A} \geq \bar{B}$, where the bar denotes the complex conjugate. This allows us to write

$$\text{Cov}[\hat{\theta}(\mathbf{x}), \Pi]_{ij} \geq \frac{1}{2} \text{Tr}[\varrho^{(v)} (\Delta X_i \Delta X_j + \Delta X_j \Delta X_i)]. \quad (\text{C6})$$

We proceed as in the classical case, defining the operators $Y_i = \Delta X_i - A_{im} L_m$, where L_m are now operators and A_{im} is a real matrix as before. We can build the following positive semidefinite matrix:

$$\text{Tr}[\varrho^{(v)} (Y_i Y_j^\dagger + Y_j Y_i^\dagger)] \geq 0, \quad (\text{C7})$$

and substituting

$$\begin{aligned} \frac{1}{2} \text{Tr}[\varrho^{(v)} (\Delta X_i \Delta X_j + \Delta X_j \Delta X_i)] &\geq \frac{A_{im}}{2} \text{Tr}[\varrho^{(v)} (L_m \Delta X_j + \Delta X_j L_m^\dagger)] + \frac{A_{jp}}{2} \text{Tr}[\varrho^{(v)} (\Delta X_i L_p^\dagger + L_p \Delta X_i)] \\ &\quad - \frac{A_{im}}{2} \text{Tr}[\varrho^{(v)} (L_m L_p^\dagger + L_p L_m^\dagger)] = (AT^\top + TA^\top - QAQ^\top)_{ij}, \end{aligned} \quad (\text{C8})$$

where

$$T_{ij} = \frac{1}{2} \text{Tr}[\varrho^{(v)} (\Delta X_i L_j^\dagger + L_j \Delta X_i)], \quad Q_{ij} = \frac{1}{2} \text{Tr}[\varrho^{(v)} (L_i L_j^\dagger + L_j L_i^\dagger)]. \quad (\text{C9})$$

Thanks to symmetrization, everything is real. We find the optimum A as in the classical case by differentiating the trace with respect to A to find the value of A that minimizes it; the result is

$$\text{Cov}[\hat{\theta}(\mathbf{x}), \Pi] \geq T Q^{-1} T^\top. \quad (\text{C10})$$

We can add the bias to get the MSE,

$$\text{MSE}[\hat{\theta}(\mathbf{x}), \Pi] \geq T Q^{-1} T^\top + \langle \delta \hat{\theta} \rangle \langle \delta \hat{\theta} \rangle^\top. \quad (\text{C11})$$

To derive the nonsymmetric version, we proceed without symmetrizing and do not assume A to be real, arriving at

$$\text{Tr}(\varrho^{(v)} \Delta X_i \Delta X_j) \geq A_{im} \text{Tr}(\varrho^{(v)} L_m \Delta X_j) + \text{Tr}(\varrho^{(v)} \Delta X_i L_m^\dagger) A_{jm}^* - A_{il} \text{Tr}(\varrho^{(v)} L_l L_m^\dagger) A_{jm}^* = (AT^\dagger + TA^\dagger - AGA^\dagger)_{ij}, \quad (\text{C12})$$

where now

$$T_{ij} = \text{Tr}(\varrho^{(v)} \Delta X_i L_j^\dagger), \quad Q_{ij} = \text{Tr}(\varrho^{(v)} L_i L_j^\dagger). \quad (\text{C13})$$

Now, we simply set $A = T Q^{-1}$. We can also assume T to be real. This will in general be the case, as T will have the same value as in the classical version by choosing the L_i appropriately. As in the symmetric case, we get

$$\text{Cov}[\hat{\theta}(x), \Pi] \geq T Q^{-1} T^\top. \quad (\text{C14})$$

Notice that, in general, Q is a complex matrix.

Next, we focus on the symmetric version. Equation (C11) provides a bound for the MSE, but what is its relationship to the CBB or, more generally, to other classical bounds derived from Eq. (5.6)? The key observation is that T takes the same values as in the classical case, provided that the L_i are chosen appropriately. For the QBB, the L_i we select are the higher-order SLDs (5.14). This yields:

$$T_j \equiv T_{1j} = \text{Tr} \left[\left(\frac{d^j}{d\theta^j} \varrho_\theta^{(v)} \right) \Delta X \right] = \frac{d^j}{d\theta^j} \langle \hat{\theta}(\mathbf{x}) \rangle - \langle \hat{\theta}(\mathbf{x}) \rangle \frac{d^j}{d\theta^j} \text{Tr}(\varrho_\theta^{(v)}) = \frac{d^j}{d\theta^j} \langle \hat{\theta}(\mathbf{x}) \rangle, \quad (\text{C15})$$

which coincides with the T of the CBB. We have omitted the subscript in ΔX_1 since we only have one parameter. Knowing this, since the QBB applies to any POVM, it must be true that

$$\text{Cov}[\hat{\theta}(\mathbf{x}), \Pi] \geq T C^{-1} T^\top \geq T Q^{-1} T^\top. \quad (\text{C16})$$

Adding the bias, we obtain the version for the MSE

$$\text{MSE}[\hat{\theta}(\mathbf{x}), \Pi] \geq TC^{-1}T^\top + \langle \delta\hat{\theta} \rangle \langle \delta\hat{\theta} \rangle^\top \geq TQ^{-1}T^\top + \langle \delta\hat{\theta} \rangle \langle \delta\hat{\theta} \rangle^\top. \quad (\text{C17})$$

This can be generalized to different classical bounds, for example, choosing $L_i = (\varrho_\theta^{(v)})^{-1} \varrho_{\theta_i}^{(v)} - 1$ gives the quantum version of the Barankin bound [25]. In the same way, choosing L_i as the SLD for the parameter θ_i we obtain the multiparameter QCRB [72].

The first inequality, Eq. (C2) can be saturated using projective measurements, since in that case $\Pi_{\mathbf{x}}\Pi_{\mathbf{x}'} = \Pi_{\mathbf{x}}\delta_{\mathbf{x}\mathbf{x}'}$. The second inequality, Eq. (C7), is saturated when $Y = 0$; that is, $X = \langle \hat{\theta}(\mathbf{x}) \rangle + A_m L_m = \langle \hat{\theta}(\mathbf{x}) \rangle + T_k(Q^{-1})_{km} L_m$ (we define $A_{1m} \equiv A_m$ as it is a single parameter). What we do then is to define the POVM Π as the eigenvectors of this operator, and also define a new estimator $\hat{\theta}(x)$ as its eigenvalues. In this way, X saturates the inequality.

APPENDIX D: SCALING OF Q

The derivation of the asymptotic CBB in Eq. (5.10) assumed the scaling relations in Eq. (5.9). We seek to verify whether the scaling of the matrix elements Q_{ij} , matches that of their classical counterparts, C_{ij} .

We begin by analyzing the element Q_{11} , which is well known to scale linearly with the number of trials v ; i.e., $Q_{11} = vq_{11}$. This linear dependence arises because the SLD for v trials, $L_1^{(v)}$, acting on the state $\varrho_\theta^{(v)}$, is simply the sum of the individual SLDs, L_1 . For example, in the case of $v = 2$, the derivative of the composite state is given by

$$\partial_\theta \varrho^{(2)} = \partial_\theta \varrho \otimes \mathbb{1} + \mathbb{1} \otimes \partial_\theta \varrho, \quad (\text{D1})$$

and the SLD is given by $L_1^{(2)} = L_1 \otimes \mathbb{1} + \mathbb{1} \otimes L_1$. This linear scaling also extends to the off-diagonal elements Q_{1i} . To prove this, let us examine the element Q_{12} for the $v = 2$ case, which is given by

$$Q_{12} = \text{Tr} [L_1^{(2)} \partial_\theta^2 \varrho] = \text{Tr}(L_1 \otimes \mathbb{1} + \mathbb{1} \otimes L_1)(\partial_\theta^2 \varrho \otimes \varrho + 2\partial_\theta \varrho \otimes \partial_\theta \varrho + \varrho \otimes \partial_\theta^2 \varrho). \quad (\text{D2})$$

By invoking the identity $\text{Tr}(\partial_\theta^i \varrho) = 0$ for any integer $i \geq 1$, we find that the only terms contributing to the trace are

$$\text{Tr}(L_1 \partial_\theta^2 \varrho \otimes \varrho + \varrho \otimes L_1 \partial_\theta^2 \varrho) = 2q_{12}. \quad (\text{D3})$$

This can be extended to the general case of v trials, as each of the v terms composing $L_1^{(v)}$ contributes a single q_{1i} to the final trace, confirming the linear scaling $Q_{1i} = vq_{1i}$.

The analysis for the Q_{22} term is more involved. Its classical counterpart exhibits a more complex scaling, given by $C_{22} = vC_{22} + 2v(v-1)c_{11}^2$. To determine if Q_{22} follows a similar pattern, we first consider the $v = 2$ case, for which we must construct the operator $L_2^{(2)}$ satisfying

$$\frac{1}{2}[L_2^{(2)} \varrho \otimes \varrho + \varrho \otimes L_2^{(2)}] = (\partial_\theta^2 \varrho \otimes \varrho + 2\partial_\theta \varrho \otimes \partial_\theta \varrho + \varrho \otimes \partial_\theta^2 \varrho). \quad (\text{D4})$$

The terms $\partial_\theta^2 \varrho \otimes \varrho$ and $\varrho \otimes \partial_\theta^2 \varrho$ are generated by including the operators $L_2 \otimes \mathbb{1}$ and $\mathbb{1} \otimes L_2$ in the construction of $L_2^{(2)}$. The contribution of these components to the QFI element Q_{22} is thus given by

$$\text{Tr}(L_2 \otimes \mathbb{1} + \mathbb{1} \otimes L_2)(\partial_\theta^2 \varrho \otimes \varrho + 2\partial_\theta \varrho \otimes \partial_\theta \varrho + \varrho \otimes \partial_\theta^2 \varrho). \quad (\text{D5})$$

The only nonzero terms are as follows:

$$\text{Tr}(L_2 \partial_\theta^2 \varrho \otimes \varrho + \varrho \otimes L_2 \partial_\theta^2 \varrho) = 2q_{22}. \quad (\text{D6})$$

This contribution provides the linear scaling component, analogous to the vc_{22} term in the classical case. The quadratic scaling component must therefore arise from the remaining term, $2\partial_\theta \varrho \otimes \partial_\theta \varrho$. By expressing $\partial_\theta \varrho$ in terms of the SLD L_1 , this term expands as follows:

$$2\partial_\theta \varrho \otimes \partial_\theta \varrho = \frac{1}{2}(L_1 \varrho \otimes L_1 \varrho + L_1 \varrho \otimes \varrho L_1 + \varrho L_1 \otimes L_1 \varrho + \varrho L_1 \otimes \varrho L_1 + \varrho L_1 \otimes L_1 \varrho). \quad (\text{D7})$$

The terms $\frac{1}{2}(L_1 \varrho \otimes L_1 \varrho + \varrho L_1 \otimes \varrho L_1)$ can be generated by an $L_1 \otimes L_1$ component in the SLD $L_2^{(2)}$. The contribution of this component to Q_{22} is then given by

$$\text{Tr}(L_1 \otimes L_1)(\partial_\theta^2 \varrho \otimes \varrho + 2\partial_\theta \varrho \otimes \partial_\theta \varrho + \varrho \otimes \partial_\theta^2 \varrho). \quad (\text{D8})$$

Recalling that $\text{Tr}(L_i \varrho) = 0$ for any i , the only nonvanishing contribution to the trace is as follows:

$$2 \text{Tr}(L_1 \partial_\theta \varrho \otimes L_1 \partial_\theta \varrho) = 2q_{11}^2. \quad (\text{D9})$$

However, generating the remaining cross-terms, $L_1 \varrho \otimes \varrho L_1 + \varrho L_1 \otimes L_1 \varrho$, is not straightforward, as they cannot be produced by any simple tensor product of the form $L_i \otimes L_j$. Accounting for these terms would, in principle, require introducing a new operator X into the structure of $L_2^{(2)}$ that satisfies

$$\frac{1}{2}(X \varrho \otimes \varrho + \varrho \otimes X) = \frac{1}{2}(L_1 \varrho \otimes \varrho L_1 + \varrho L_1 \otimes L_1 \varrho). \quad (\text{D10})$$

To fully replicate the classical scaling, this new operator would need to yield a contribution of $2q_{11}^2$. However, an explicit demonstration of this result has proven elusive. Given the difficulty of this direct approach, we now turn to an alternative method to validate the asymptotic expressions.

We pursue an alternative approach to show that Eq. (5.10) remains valid to second order under the replacement of c_{ij} with q_{ij} . This method begins by establishing a general inequality between the two quantities. For any list of real constants $\{a_i\}$, it has been shown that [25] that

$$\begin{aligned} a_i c_{ij} a_j &= \left\langle \left(\frac{a_i}{p(x|\theta)} \frac{d^i p(x|\theta)}{d\theta^i} \right)^2 \right\rangle = \sum_k \frac{1}{p(x_k|\theta)} \left[a_i \frac{d^i p(x_k|\theta)}{d\theta^i} \right]^2 = \sum_k \frac{1}{\text{Tr}(\varrho \Pi_k)} \left\{ a_i \text{Tr} \left[\frac{1}{2} (L_i \varrho + \varrho L_i) \Pi_k \right] \right\}^2 \\ &\leq \sum_k \frac{|\text{Tr}[\varrho(a_i L_i) \Pi_k]|^2}{\text{Tr}(\varrho \Pi_k)} \leq a_i \text{Tr} \left[\frac{1}{2} (L_i L_j + L_j L_i) \varrho \right] a_j = a_i Q_{ij} a_j; \end{aligned} \quad (\text{D11})$$

that is, $C \preceq Q$. Notably, this inequality does not guarantee the existence of a single POVM for which equality holds, as the optimal measurement projectors—eigenstates of the operator $a_i L_i$ —depend on the specific choice of coefficients a_i . Despite this, the relationship provides a powerful tool for constraining individual matrix elements. By isolating the second-order terms (i.e., setting $a_i = \delta_{i2}$), we find that $C_{22} \leq Q_{22}$. Now, in the asymptotic limit $\nu \rightarrow \infty$, a measurement that saturates the QFI yields $c_{11} = q_{11}$. For such a measurement, the classical element scales as follows:

$$C_{22} = 2\nu^2 c_{11}^2 = 2\nu^2 q_{11}^2 \leq Q_{22}. \quad (\text{D12})$$

Therefore, in the asymptotic limit $\nu \rightarrow \infty$, it must hold that $Q_{22} \geq 2\nu^2 q_{11}^2$. This lower bound is sufficient to validate the asymptotic expression in Eq. (5.10), since the dominant contribution from the classical term C_{22} to that expression is precisely its leading-order term, $2\nu^2 c_{11}^2$. While this analysis confirms the result to order $O(1/\nu^2)$, a complete determination of higher-order corrections would necessitate a more precise characterization of the scaling for all Q_{ij} elements.

-
- [1] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, *Phys. Rev. Lett.* **96**, 010401 (2006).
 - [2] M. G. A. Paris, Quantum estimation for quantum technology, *Int. J. Quantum Inf.* **07**, 125 (2009).
 - [3] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, *Nat. Photon.* **5**, 222 (2011).
 - [4] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kolodyński, Quantum limits in optical interferometry, *Prog. Opt.* **60**, 345 (2015).
 - [5] J. S. Sidhu and P. Kok, Geometric perspective on quantum parameter estimation, *AVS Quantum Sci.* **2**, 014701 (2020).
 - [6] F. Albarelli, M. Barbieri, M. G. Genoni, and I. Gianani, A perspective on multiparameter quantum metrology: From theoretical tools to applications in quantum imaging, *Phys. Lett. A* **384**, 126311 (2020).
 - [7] E. Polino, M. Valeri, N. Spagnolo, and F. Sciarrino, Photonic quantum metrology, *AVS Quantum Sci.* **2**, 024703 (2020).
 - [8] M. Barbieri, Optical quantum metrology, *PRX Quantum* **3**, 010202 (2022).
 - [9] A. P. Alodjants, D. V. Tsarev, D. A. Kuts, S. A. Podoshvedov, and S. P. Kulik, Quantum optical metrology, *Phys. Usp.* **67**, 668 (2024).
 - [10] G. Casella and R. Berger, *Statistical Inference*, 2nd ed. (Duxbury Thomson Learning, Albany, NY, 2002).
 - [11] D. Petz and C. Ghinea, Introduction to quantum fisher information, in *Quantum Probability and Related Topics*, Vol. 27 (World Scientific, Singapore, 2011), pp. 261–281.
 - [12] J. Rubio, P. Knott, and J. Dunningham, Non-asymptotic analysis of quantum metrology protocols beyond the Cramér–Rao bound, *J. Phys. Commun.* **2**, 015027 (2018).
 - [13] J. Rubio and J. Dunningham, Quantum metrology in the presence of limited data, *New J. Phys.* **21**, 043037 (2019).
 - [14] J. J. Meyer, S. Khatri, D. S. França, J. Eisert, and P. Faist, Quantum metrology in the finite-sample regime, [arXiv:2307.06370](https://arxiv.org/abs/2307.06370) [PRX Quantum (to be published)].
 - [15] W. Salmon, S. Strelchuk, and D. Arvidsson-Shukur, Only classical parameterised states have optimal measurements under least squares loss, *Quantum* **7**, 998 (2023).
 - [16] M. Hayashi, Comparison between the Cramér–Rao and the mini-max approaches in quantum channel estimation, *Commun. Math. Phys.* **304**, 689 (2011).
 - [17] M. J. W. Hall and H. M. Wiseman, Heisenberg-style bounds for arbitrary estimates of shift parameters including prior information, *New J. Phys.* **14**, 033040 (2012).
 - [18] G. Chesi, A. Riccardi, R. Rubboli, L. Maccone, and C. Macchiavello, Protocol for global multiphase estimation, *Phys. Rev. A* **108**, 012613 (2023).
 - [19] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry, π -corrected Heisenberg limit, *Phys. Rev. Lett.* **124**, 030501 (2020).
 - [20] W. Górecki and R. Demkowicz-Dobrzański, Multiparameter quantum metrology in the Heisenberg limit regime: Many-repetition scenario versus full optimization, *Phys. Rev. A* **106**, 012424 (2022).
 - [21] D. G. Chapman and H. Robbins, Minimum variance estimation without regularity assumptions, *Ann. Math. Stat.* **22**, 581 (1951).
 - [22] A. Bhattacharyya, On some analogues of the amount of information and their use in statistical estimation, *Sankhya: Ind. J. Stat.* **8**, 1 (1946).
 - [23] E. W. Barankin, Locally best unbiased estimates, *Ann. Math. Stat.* **20**, 477 (1949).
 - [24] J. S. Abel, A bound on mean-square-estimate error, *IEEE Trans. Inf. Theory* **39**, 1675 (1993).

- [25] M. Gessner and A. Smerzi, Hierarchies of frequentist bounds for quantum metrology: From Cramér-Rao to Barankin, *Phys. Rev. Lett.* **130**, 260801 (2023).
- [26] J. R. Hervas, A. Z. Goldberg, A. S. Sanz, Z. Hradil, J. Řeháček, and L. L. Sánchez-Soto, Beyond the quantum Cramér-Rao bound, *Phys. Rev. Lett.* **134**, 010804 (2025).
- [27] J. K. Ghosh, Higher order asymptotics, in *NSF-CBMS Regional Conference Series in Probability and Statistics* (Institute of Mathematical Statistics, Alexandria, VA, 1994), p. 4153181.
- [28] S. Kurdziałek and R. Demkowicz-Dobrzanski, Measurement noise susceptibility in quantum estimation, *Phys. Rev. Lett.* **130**, 160802 (2023).
- [29] P. Kolenderski and R. Demkowicz-Dobrzanski, Optimal state for keeping reference frames aligned and the platonic solids, *Phys. Rev. A* **78**, 052333 (2008).
- [30] G. Björk, A. B. Klimov, P. De La Hoz, M. Grassl, G. Leuchs, and L. L. Sánchez-Soto, Extremal quantum states and their Majorana constellations, *Phys. Rev. A* **92**, 031801(R) (2015).
- [31] D. Baguette, F. Damanet, O. Giraud, and J. Martin, Anticoherence of spin states with point-group symmetries, *Phys. Rev. A* **92**, 052333 (2015).
- [32] A. Z. Goldberg, A. B. Klimov, G. Leuchs, and L. L. Sánchez-Soto, Rotation sensing at the ultimate limit, *J. Phys. Photon.* **3**, 022008 (2021).
- [33] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [34] S. M. Kay, *Fundamentals of Statistical Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
- [35] H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 2001).
- [36] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. (Springer, New York, 1998).
- [37] R. A. Fisher, Theory of statistical estimation, *Math. Proc. Cambr. Philos. Soc.* **22**, 700 (1925).
- [38] J. Shang, H. K. Ng, and B.-G. Englert, Quantum state tomography: Mean squared error matters, bias does not, [arXiv:1405.5350](https://arxiv.org/abs/1405.5350) [quant-ph].
- [39] I. Chuang and M. Nielsen, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
- [40] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, Quaderni Monographs No. 1 (Ed. della Normale, Pisa, 2011).
- [41] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, New York, 2002).
- [42] W. K. Wootters, Statistical distance and Hilbert space, *Phys. Rev. D* **23**, 357 (1981).
- [43] D. Bures, An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^* -algebras, *Trans. Am. Math. Soc.* **135**, 199 (1969).
- [44] A. Uhlmann, The "transition probability" in the state space of a $*$ -algebra, *Rep. Math. Phys.* **9**, 273 (1976).
- [45] D. Šafránek, Discontinuities of the quantum Fisher information and the Bures metric, *Phys. Rev. A* **95**, 052320 (2017).
- [46] D. Šafránek, Simple expression for the quantum Fisher information matrix, *Phys. Rev. A* **97**, 042322 (2018).
- [47] J. Liu, M. Zhang, H. Chen, L. Wang, and H. Yuan, Optimal scheme for quantum metrology, *Adv. Quantum Tech.* **5**, 2100080 (2022).
- [48] L. J. Fiderer, J. M. E. Fraisse, and D. Braun, Maximal quantum Fisher information for mixed states, *Phys. Rev. Lett.* **123**, 250502 (2019).
- [49] J. P. Dowling, Quantum optical metrology—The lowdown on high-N00N states, *Contemp. Phys.* **49**, 125 (2008).
- [50] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, Bell's theorem without inequalities, *Am. J. Phys.* **58**, 1131 (1990).
- [51] E. Serrano-Ensástiga, C. Chryssomalakos, and J. Martin, Quantum metrology of rotations with mixed spin states, *Phys. Rev. A* **111**, 022435 (2025).
- [52] J. Martin, S. Weigert, and O. Giraud, Optimal detection of rotations about unknown axes by coherent and anticonherent states, *Quantum* **4**, 285 (2020).
- [53] J. L. Romero, A. B. Klimov, A. Z. Goldberg, G. Leuchs, and L. L. Sánchez-Soto, Multipoles from Majorana constellations, *Phys. Rev. A* **109**, 012214 (2024).
- [54] C. Chryssomalakos and H. Hernández-Coronado, Optimal quantum roto-sensors, *Phys. Rev. A* **95**, 052125 (2017).
- [55] S. Zhou, C.-L. Zou, and L. Jiang, Saturating the quantum Cramér-Rao bound using LOCC, *Quantum Sci. Technol.* **5**, 025005 (2020).
- [56] R. H. Norden, A survey of Maximum Likelihood Estimation, *Int. Stat. Rev.* **40**, 329 (1972).
- [57] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. (McGraw-Hill, Boston, 2002).
- [58] C. R. Rao, Asymptotic efficiency and limiting information, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Vol. 4 (University of California Press, Los Angeles, 1961), pp. 531–546.
- [59] C. R. Rao, Apparent anomalies and irregularities in maximum likelihood estimation (with discussion), *Sankhya Ser. A* **24**, 73 (1962).
- [60] C. R. Rao, Criteria of estimation in large samples, *Sankhya Ser. A* **25**, 189 (1963).
- [61] J. K. Ghosh and K. Subramanyam, Second order efficiency of maximum likelihood estimators, *Sankhya Ser. A* **36**, 325 (1974).
- [62] J. Berkson, Minimum Chi-Square, not maximum likelihood!, *Ann. Stat.* **8**, 457 (1980).
- [63] J. B. S. Haldane and S. M. Smith, The sampling distribution of a maximum-likelihood estimate, *Biometrika* **43**, 96 (1956).
- [64] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs and Mathematical Tables* (Dover, New York, 1972).
- [65] E. W. Weisstein, Series reversion, <https://mathworld.wolfram.com/SeriesReversion.html>.
- [66] F. Ouimet, General formulas for the central and non-central moments of the multinomial distribution, *Stats* **4**, 18 (2021).
- [67] M. Tsang, Conservative classical and quantum resolution limits for incoherent imaging, *J. Mod. Opt.* **65**, 1385 (2018).
- [68] Y. Ye and X.-M. Lu, Quantum Cramér-Rao bound for quantum statistical models with parameter-dependent rank, *Phys. Rev. A* **106**, 022429 (2022).
- [69] B. Efron, Defining the curvature of a statistical problem (with applications to second order efficiency), *Ann. Statist.* **3**, 1189 (1975).
- [70] Y. Tsuda, Bhattacharyya inequality for quantum state estimation, *J. Phys. A: Math. Theor.* **40**, 793 (2007).

- [71] Y. Tsuda and K. Matsumoto, Quantum estimation for non-differentiable models, *J. Phys. A: Math. Gen.* **38**, 1593 (2005).
- [72] K. R. Parthasarathy, On the philosophy of Cramér-Rao-Bhattacharya Inequalities in quantum statistics, *Sankhya A* **83**, 521 (2021).
- [73] D. C. Brody and L. P. Hughston, Geometry of quantum statistical inference, *Phys. Rev. Lett.* **77**, 2851 (1996).
- [74] D. C. Brody and L. P. Hughston, Generalised Heisenberg relations for quantum statistical estimation, *Phys. Lett. A* **236**, 257 (1997).
- [75] D. C. Brody and L. P. Hughston, Statistical geometry in quantum mechanics, *Proc. R. Soc. Lond. A* **454**, 2445 (1998).
- [76] A. J. Belfield and D. C. Brody, Higher-order uncertainty bounds for mixed states, *J. Phys. A: Math. Theor.* **54**, 435302 (2021).
- [77] L. J. Fiderer, T. Tufarelli, S. Piano, and G. Adesso, General expressions for the quantum Fisher information matrix with applications to discrete quantum imaging, *PRX Quantum* **2**, 020308 (2021).
- [78] L. O. Conlon, J. Suzuki, P. K. Lam, and S. M. Assad, Efficient computation of the Nagaoka-Hayashi bound for multiparameter estimation with separable measurements, *npj Quantum Inf.* **7**, 110 (2021).
- [79] M. Hayashi and Y. Ouyang, Tight Cramér-Rao type bounds for multiparameter quantum metrology through conic programming, *Quantum* **7**, 1094 (2023).
- [80] M. Hayashi and Y. Ouyang, Finding the optimal probe state for multiparameter quantum metrology using conic programming, *npj Quantum Inf.* **10**, 111 (2024).
- [81] L. O. Conlon, J. Suzuki, P. K. Lam, and S. M. Assad, Role of the extended Hilbert space in the attainability of the quantum Cramér-Rao bound for multiparameter estimation, *Phys. Lett. A* **542**, 130445 (2025).
- [82] W.-K. Tung, *Group Theory in Physics* (World Scientific, Singapore, 1985).