

Efficient qubit phase estimation using adaptive measurements

Marco A. Rodríguez-García¹, Isaac Pérez Castillo², and P. Barberis-Blostein¹

¹Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Ciudad Universitaria, Ciudad de México 04510, Mexico

²Departamento de Física, Universidad Autónoma Metropolitana-Iztapalapa, San Rafael Atlixco 186, Ciudad de México 09340, Mexico

Estimating correctly the quantum phase of a physical system is a central problem in quantum parameter estimation theory due to its wide range of applications from quantum metrology to cryptography. Ideally, the optimal quantum estimator is given by the so-called quantum Cramér-Rao bound, so any measurement strategy aims to obtain estimations as close as possible to it. However, more often than not, the current state-of-the-art methods to estimate quantum phases fail to reach this bound as they rely on maximum likelihood estimators of non-identifiable likelihood functions. In this work we thoroughly review various schemes for estimating the phase of a qubit, identifying the underlying problem which prohibits these methods to reach the quantum Cramér-Rao bound, and propose a new adaptive scheme based on covariant measurements to circumvent this problem. Our findings are carefully checked by Monte Carlo simulations, showing that the method we propose is both mathematically and experimentally more realistic and more efficient than the methods currently available.

1 Introduction

The aim of statistical estimation theory in classical systems is to estimate a probability distribution based on a series of observations. More precisely, given a statistical model $\mathcal{P} = \{p(x | \theta) | \theta \in \Theta \subset \mathbb{R}^n\}$ and a series of observations $\{X_1, \dots, X_N\}$ generated by a particular distribution density $p(x | \theta)$, the aim is to design an estimator $\hat{\theta}(X_1, \dots, X_N)$ which is as close as possible to the parameter θ , measured in terms of a cost function, usually the Mean Square Error (MSE), between $\hat{\theta}$ and θ . A fairly remarkable result, the celebrated Cramér-Rao bound (CRB), states that the variance of an unbiased estimator, is at least as high as the inverse of the Fisher information of $p(x | \theta)$. The estimators that saturate this bound are called efficient. An example of an efficient estimator is the maximum likelihood, that saturates the CRB when the number of measurements tends to infinity.

This classical result has, however, some subtleties, namely: for the maximum likelihood estimator to be consistent (i.e. the estimator converges to the parameter when the

Marco A. Rodríguez-García: marg@ciencias.unam.mx

Isaac Pérez Castillo: iperez@izt.uam.mx

P. Barberis-Blostein: pbb@iimas.unam.mx

number of outcomes tends to infinity), the likelihood function must be identifiable (i.e. the likelihood has a unique global maximum), Θ must be a compact set, and all probability distributions in the statistical model must have the same support. Statistical models satisfying the previous conditions are called regular models.

While the actual role that the process of measurement plays in extracting information from a classically system is not particularly relevant, the situation is completely different in a quantum system, making the process of parameter estimation mathematically more involved. Suppose ρ is an unknown density operator acting on a Hilbert space \mathcal{H} , that represents the state of the system. Let us assume that ρ belongs to a particular subset of parametrized states

$$\mathcal{S} = \{\rho_\theta; \theta \in \Theta \subset \mathbb{R}^n\} . \quad (1)$$

The parametrized states ρ_θ are the result of the evolution of an initial probe state ρ by a trace-preserving dynamical process dependent upon a parameter θ . Note that different initial conditions give rise to different subsets of parametrized states.

Quantum parameter estimation aims to produce the best estimation for θ under the premise that the system is prepared in a state of the parametric family \mathcal{S} . Notice that, unlike its classical counterpart, quantum estimation consists of two parts: the choice of the measurements to perform on the system, and then the processing of their outcomes through an estimator $\hat{\theta}$. Note that measured data is a set of random outcomes with a probability distribution depending on the parameter, and an estimator is a random variable whose outcomes estimates the parameter. Thus, the pair measurement-estimator forms a strategy of estimation in the quantum case one can play around with seeking for an optimal estimation strategy. Indeed, each choice of measurement defines a classical statistical model, whose lowest possible CRB maximises the Fisher information. Thus the maximum of all possible Fisher informations, over the space of measurements allowed by the postulates of Quantum Mechanics, is called the Quantum Fisher Information (QFI), and its inverse, the Quantum Cramér-Rao Bound (QCRB). In other words, the QCRB is the minimum over all possible MSE of any possible estimation strategy allowed by Quantum Mechanics. If the Fisher information for a particular choice of measurement M is equal to the QFI, this measurement is called optimal.

One way to find an optimal measurement is to use the set of operators $\{M(j)\}$ to represent the measurement, where each $M(j)$ corresponds to a projector onto each of the eigenspaces of the symmetric logarithmic derivative (SLD) $\mathcal{L}(\theta)$ [5], defined by

$$\frac{\partial \rho_\theta}{\partial \theta} = \frac{1}{2} (\mathcal{L}(\theta)\rho_\theta + \rho_\theta\mathcal{L}(\theta)) . \quad (2)$$

It turns out that, in general, this measurement is so-called locally optimal. This means that the optimal measurement depends on the actual value of the parameter we want to estimate and the classical Fisher information given by that measurement is a local maximum in the parameter space that characterizes the measurement. Thus estimating the parameter using a locally optimal measurement is rather impractical. Two approaches have been developed to overcome this problem. The first one is based on adaptive estimation schemes which updates a guess for a locally optimal measurement [1, 13]. The second method seeks a set of initial conditions that do not depend on the unknown parameter [7, 33]. Both methods are based on the maximum likelihood estimator (MLE) which implies that in order to obtain it, the aforementioned set of regularity conditions must be met [6, 11, 20]. We will see that the MLE derived from optimal measurements may fail to satisfy these conditions and thus these two approaches fail to attain the QCRB.

The problem of non-identifiable likelihood functions often appears when estimating the quantum phase of a system, a particular problem with a wide range of applications from

quantum metrology to cryptography, as many problems can be recast into a quantum phase estimation [10, 24, 27]. In this framework, when trying to use locally optimal measurements [12, 14, 19, 31, 33], it may happen that adaptive schemes fail to reach the QCRB [28] due to the fact that regularity conditions are not met [13]. A paradigmatic example of non-identifiability is provided when trying to perform qubit phase estimation which are, in general, unable to reach the QCRB [1, 7].

The main goal of the present work is to introduce a new adaptive scheme for quantum phase estimation which saturates the QCRB independently of the initial condition that generates \mathcal{S} . We will first show that the Adaptive Quantum State Estimation method (AQSE), a general method for parameter estimation [35], does not achieve the QCRB when applied to a qubit phase estimation. The reason why AQSE does not converge is that the statistical model, built with the measurement that maximizes the Fisher information (the locally optimal measurement), is not identifiable. We will further show that the non-identifiability problem is solved by taking a sample of measures from a covariant measurement. The covariant measurement is identifiable and is chosen to minimize the MSE for one measure. With the covariant sample, we can then construct a confidence interval where the underlying statistical model is now regular. Then we will apply the AQSE method inside the confidence interval.

The paper is organized as follows. We start by giving a brief review of quantum estimation theory with special emphasis in covariant estimation techniques for periodic parameter estimation (section 2). Then, we review different estimation strategies for the phase of a qubit and discuss their weakness and strengths, in particular we discuss the effects of non-identifiable probabilistic models in the estimation error (section 3). Once we understand the problems with different estimation strategies, we present a two-step estimation scheme. Numerical simulations suggest that this scheme is able to reach the QCRB (section 4). We end the paper with a summary of the main results and our conclusions (section 5).

2 Quantum parameter estimation

Recall that, given a set of independent observations $\{x_1, \dots, x_N\}$ from a random variable X with probability density $p(x | \theta)$, the likelihood function is defined as

$$L(\theta | x_1, \dots, x_N) = \prod_{i=1}^N p(x_i | \theta). \quad (3)$$

From here, we can derive the maximum likelihood estimator for θ , whose mean square error obeys the CRB [20]

$$\text{MSE}(\hat{\theta}) \geq \frac{1}{NF(\theta; X)}, \quad (4)$$

where the MSE is defined by

$$\text{MSE}(\hat{\theta}) = E_{\theta} \left[\left(\hat{\theta} - \theta \right)^2 \right], \quad (5)$$

and $F(\theta; X)$ is the Fisher information:

$$F(\theta; X) = E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log p(x | \theta) \right)^2 \right]. \quad (6)$$

Intuitively, the Fisher information quantifies how much information carries a sample about the unknown parameter. The ultimate aim in classical parameter estimation is to find the

estimator that achieves the Cramér-Rao bound. If the statistical model $\{p(x | \theta) | \theta \in \Theta\}$ for a random variable X is regular the maximum likelihood estimator produced by a sequence of N independent and identically distributed of X can attain the Cramér-Rao bound asymptotically for large N [11, 20].

In the quantum case, we must first describe the process of measuring the system. This is better achieved by using the concept of Positive Operator-Valued Measures (POVMs). A POVM with outcomes on a set \mathcal{X} is a family of bounded positive operators

$$M = \{M(B); B \text{ is a Borel set in } \mathcal{X}\}$$

acting over the Hilbert space of the system \mathcal{H} such that $M(\mathcal{X}) = I$. When the state of the system is given by the density operator ρ_θ , a POVM M specifies the conditional probability of obtaining the event B by Born's rule

$$P(B | \theta) = \text{Tr} [M(B)\rho_\theta]. \quad (7)$$

Thus, in quantum parameter estimation, the Fisher information, given by Eq. (6), is a function of the POVMs and, henceforth, we will denote it as $F(\theta; M)$. This implies that, according to Eq. (4), maximising $F(\theta; M)$ over the set of POVMs gives the lowest CRB. This results into the so-called Quantum Fisher Information (QFI), which we will denote as $F_Q(\theta)$, and its lowest bound is so-called the Quantum Cramér-Rao Bound (QCRB) [15, 16, 18].

For the particular case of a scalar parameter estimation, the QFI has the following form [5, 15, 16]

$$F_Q(\theta) = \text{Tr} \left[\rho_\theta \mathcal{L}(\theta)^2 \right], \quad (8)$$

where $\mathcal{L}(\theta)$ is the symmetric logarithmic derivative, also called quantum score, defined by Eq. (2). When the system is in a pure state $\rho_\theta = |\psi_\theta\rangle \langle \psi_\theta|$, the quantum score is easy to calculate [29], obtaining $\mathcal{L}(\theta) = 2 \frac{\partial \rho_\theta}{\partial \theta}$, so that the QFI becomes:

$$F_Q(\theta) = 4 \text{Tr} \left[\left(\frac{\partial}{\partial \theta} (|\psi_\theta\rangle \langle \psi_\theta|) \right)^2 |\psi_\theta\rangle \langle \psi_\theta| \right]. \quad (9)$$

Notice that the set of POVMs $\{M_\theta(j)\}$ constructed as the projections of the quantum score $\mathcal{L}(\theta)$ do depend on the parameter θ one is trying to infer [5]. A natural way around this is to introduce adaptive schemes to estimate θ . One approach relies on adaptive quantum estimation schemes based on locally optimal POVMs that could, in principle, asymptotically construct the optimal POVM without knowing the parameter beforehand [2, 13, 19, 25]. Nevertheless, a set of precise mild regularity conditions for each statistical model involved in the method are required. For instance, in the adaptive quantum state estimation (AQSE) method [13, 25], it is necessary to assume regular statistical models for every measurement to guarantee a consistent and efficient estimator. However, for estimation problems on which the quantum parameter is periodic, as it is the case for phase estimation, the likelihood functions produced by locally optimal POVMs are not identifiable [3, 7]. Consequently, there is no mathematical reason that ensures the saturation of the CRB. The second approach searches specific initial conditions, for which the optimal POVM does not depend on the unknown value of θ [7, 33]. Hence, in principle, performing an extensive independent sequence of this POVM, along with the maximum likelihood estimator, it is possible to achieve the quantum Cramér-Rao bound. Nonetheless, for periodic parameter estimation, this method often produces non-identifiable likelihood functions, and as a consequence, the CRB is not attained.

We will show now that, for the case of quantum phase estimation, we can solve the problem of non-identifiable likelihood functions by using the formalism of covariant estimation, which considers the symmetries of the system and can saturate the QCRB under specific initial conditions [16].

2.1 Covariant estimation

The following discussion is based on [16]. In quantum parameter estimation, each $\rho_\theta \in \mathcal{S}$ obtains its parametric dependency through a physical transformation U_θ . Usually, the set of $\{U_\theta\}_{\theta \in \Theta}$ forms a group, which induces an action over \mathcal{S} . When the family \mathcal{S} is invariant under the conjugation of U_θ , for any $\theta \in \Theta$, we say that the problem of estimation involves a symmetry. The problem of quantum parameter estimation involving symmetries is called covariant. In the following, we summarise the approach of covariant estimation.

Let G be a locally compact Lie group acting transitively over the parametric set Θ . Thus for any base point $\theta_0 \in \Theta$, we have

$$\Theta = \theta_0 \cdot G \cong G/K, \quad (10)$$

where K is the stabilizer subgroup of θ_0 . According to Wigner's theorem [34], the Hilbert space \mathcal{H} of the system has a unitary G -representation $U = \{U_g\}_{g \in G}$. Hence, the probe states can be transformed by the conjugation

$$\rho \mapsto U_g \rho U_g^\dagger, \quad (11)$$

where U_g is an element of the G -unitary representation of \mathcal{H} . Specifically, we say that we have a quantum covariant estimation problem whenever

$$\{U_g \rho_\theta U_g^\dagger \mid \theta \in \Theta\} = \mathcal{S} \quad \forall g \in G. \quad (12)$$

In this framework, a POVM M on the system \mathcal{H} taking values in Θ is called *covariant* with respect the G -unitary representation $\{U_g\}$, if for every event B and $g \in G$, it satisfies that

$$M(Bg^{-1}) = U_g^\dagger M(B) U_g, \quad (13)$$

where $Bg^{-1} = \{\theta \mid \theta = g\theta', \theta' \in B\}$. Note that this condition is equivalent to

$$P(B \mid \rho) = P(gB \mid U_g \rho U_g^\dagger). \quad (14)$$

We will denote as $\mathcal{M}(\Theta)$ the set of POVMs with outcomes in Θ . Similarly, we will denote as $\mathcal{M}(\Theta, U)$ the set of the covariant POVMs with respect the G -unitary representation. The class of covariant POVMs takes advantage of group symmetries. Specifically, when the outcomes of the measurement are considered as the estimates, and the cost function $c(\hat{\theta}, \theta)$ under consideration is G -invariant, that is $c(\hat{\theta}, \theta) = c(g\hat{\theta}, \theta) \forall g \in G$, one has that

$$\begin{aligned} E_{g\theta, M} [c(\hat{\theta}, g\theta)] &= \int_{\Theta} c(\hat{\theta}, g\theta) \text{Tr} [M(d\hat{\theta}) \rho_{g\theta}] = \int_{\Theta} c(g\hat{\theta}', g\theta) \text{Tr} [M(d\hat{\theta}') \rho_\theta] \\ &= E_{\theta, M} [c(\hat{\theta}', \theta)], \quad \forall M \in \mathcal{M}(\Theta, U) \quad \text{and} \quad \hat{\theta}' := g^{-1}\hat{\theta}. \end{aligned} \quad (15)$$

As a result, we can find optimal measurements on $\mathcal{M}(\Theta, U)$ in the Bayesian approach, by taking the average of $c(\theta, \hat{\theta})$ over a prior probability measure. The following theorem ensures this assertion.

Theorem 1 ([18]). *When Θ is compact,*

$$\min_{M \in \mathcal{M}(\Theta)} \int_{\Theta} E_{\theta, M} [c(\hat{\theta}, \theta)] \mu(d\theta) = \min_{M \in \mathcal{M}(\Theta, U)} \int_{\Theta} E_{\theta, M} [c(\hat{\theta}, \theta)] \mu_{\Theta}(d\theta). \quad (16)$$

where μ_{Θ} is the Haar measure over Θ .

The previous result restricts the search for optimal measurements to the class $\mathcal{M}(\Theta, U)$. This fact is particularly useful as, according to the following theorem, the covariant POVMs can be characterised as follows.

Theorem 2 ([16]). *Let P_0 be a Hermitian positive operator on \mathcal{H} , commuting with the operators $\{U_g\}_{g \in K}$ and satisfying*

$$\int_{\Theta} U_{g(\theta)} P_0 U_{g(\theta)}^{\dagger} \mu_{\Theta}(d\theta) = I, \quad (17)$$

where $g(\theta) \in G$ is any representative element of the equivalence class $\theta \in \Theta$. Then, a POVM $M(d\theta)$ with outcomes in $\Theta \cong G/K$ is covariant if and only if it has the form

$$M(d\theta) = U_{g(\theta)} P_0 U_{g(\theta)}^{\dagger} \mu_{\Theta}(d\theta). \quad (18)$$

A particular case, which is relevant for quantum phase estimation, is when the stability group is the trivial one, that is when $K = \{e\}$ with e the identity element. Here, we have that $\Theta \cong G$, that is, the parametric space Θ is in one-to-one correspondence to the elements of the group G . This implies, in turn, that $E_{\theta, M} [c(\hat{\theta}, \theta)]$ is constant for all $\theta \in \Theta$ and, as a consequence of theorem (1), the optimization of the expectation of the invariant cost function can be restricted to the set of covariant POVMs. Notice that for quantum phase estimation, there is a little caveat: here the one-parameter symmetry group G is isomorphic to $\Theta = [0, 2\pi)$ but this is not a compact set. In practice, we can consider it to be compact, assuming that the unknown parameter is an interior point of Θ . Thus, we can analyze the CRB for any covariant POVM.

We proceed to discuss in more detail how these ideas apply to quantum phase estimation for qubit states.

3 Phase estimation strategies in a qubit

Let $\mathcal{S} = \{\rho_{\theta}; \theta \in \Theta = [0, 2\pi)\}$ be a parametric family of density operators, on a 2-dimensional Hilbert space \mathcal{H} . In this case, each state $\rho_{\theta} \in \mathcal{S}$ represents the state of a qubit. Here, the states ρ_{θ} obtain their parametric dependency by applying an arbitrary unitary transformation U_{θ} over a probe state ρ on \mathcal{H} , that is

$$\rho_{\theta} = U_{\theta} \rho U_{\theta}^{\dagger}. \quad (19)$$

An arbitrary unitary transformation (up to a global phase) on a 2-dimensional Hilbert space can be written using the generators of the Lie algebra $\mathfrak{su}(2)$ and has the following form [23, 26]

$$U_{\theta} = e^{-i\theta \vec{n} \cdot \frac{\vec{\sigma}}{2}}, \quad (20)$$

where $\vec{n} = n_x \hat{x} + n_y \hat{y} + n_z \hat{z} \in \mathbb{R}^3$ is a unit vector and $\vec{\sigma} = \sigma_1 \hat{x} + \sigma_2 \hat{y} + \sigma_3 \hat{z}$, with σ_i , for $i = 1, 2, 3$, the Pauli matrices. The estimation problem in this case consists in finding the

best strategy to estimate the phase $\theta \in \Theta$ of the exponential operator appearing in Eq. (20).

To analyse the QCRB in this case it is better to work using the Bloch sphere representation for a qubit state. The probe state can be written as

$$\rho = \frac{1}{2} (I + \vec{a} \cdot \vec{\sigma}), \quad \|\vec{a}\| \leq 1, \quad (21)$$

and the transformation U_θ can be seen as the rotation of the probe Bloch vector \vec{a} by the angle θ around the axis \vec{n} . We will denote the transformed vector as $\vec{a}(\theta)$, so the transformed qubit state ρ_θ is solely determined by $\vec{a}(\theta)$. One can show that, for this problem, the quantum score takes the following form [7]:

$$\mathcal{L}(\theta) = (\vec{n} \times \vec{a}(\theta)) \cdot \vec{\sigma}, \quad (22)$$

where

$$\vec{a}(\theta) = \cos(\theta)\vec{a} + (\sin(\theta)(\vec{n} \times \vec{a})) + \left(2(\vec{n} \cdot \vec{a}) \sin^2\left(\frac{\theta}{2}\right)\right) \vec{n}. \quad (23)$$

This implies that, according to Eq. (8), the quantum Fisher information becomes

$$F_Q = \|\vec{a}\|^2 - (\vec{a} \cdot \vec{n})^2. \quad (24)$$

Thus, if the probe qubit is a pure state (i.e., $\|\vec{a}\| = 1$) and the Bloch vector \vec{a} is orthogonal to the rotation axis \vec{n} , then Eq. (24) achieves their maximal value $F_Q^{\max} = 1$. We will refer to this as the optimal initial conditions. Let us proceed to discuss the different estimation strategies for the phase θ .

3.1 Locally Optimal POVMs

From the expression of the quantum score, given by Eq. (22), it is straightforward obtain that its projection operators are given by:

$$M_g(0) = \frac{1}{2} \left(I + \frac{(\vec{n} \times \vec{a}(g))}{\|\vec{n} \times \vec{a}(g)\|} \cdot \vec{\sigma} \right), \quad M_g(1) = \frac{1}{2} \left(I - \frac{(\vec{n} \times \vec{a}(g))}{\|\vec{n} \times \vec{a}(g)\|} \cdot \vec{\sigma} \right), \quad g \in \Theta. \quad (25)$$

Hence, the family of POVMs is $M_g = \{M_g(0), M_g(1)\}$, with outcomes in the set $\mathcal{X} = \{0, 1\}$, yield the following classical Fisher information

$$F(\theta; g) = \frac{F_Q \cos^2(\theta - g)}{1 - F_Q \sin^2(\theta - g)}. \quad (26)$$

From Eq. (26) it is fairly obvious to realize that $F(\theta, g) = F_Q(\theta)$ whenever $g = \theta$ for any $\theta \in \Theta$. That is, the POVM $M_\theta = \{M_\theta(0), M_\theta(1)\}$ is locally optimal. Note, however, that if $\vec{a} \perp \vec{n}$ then

$$F(\theta; g) = \frac{\cos^2(\theta - g)}{1 - \sin^2(\theta - g)} = 1 = F_Q^{\max}. \quad (27)$$

In other words, when the probe state is prepared in the optimal initial conditions, the POVM M_g is optimal, independent of $\theta \in \Theta$. Therefore, any POVM M_g is optimal.

At this point, one may think that a POVM M_g produces an *optimal estimation strategy* in the sense that if we were to perform a sequence of N independent measurements using the maximum likelihood estimator, it should be possible to saturate the QCRB in the asymptotic regime. This is, however, false: the POVM M yields a non-identifiable likelihood function, and thus the maximum likelihood estimator is not consistent. Indeed,

when a sequence of N identically quantum systems with Hilbert space \mathcal{H} is prepared in the same state ρ_θ , the composite system is described by the N -tensor product $\underbrace{\mathcal{H} \otimes \cdots \otimes \mathcal{H}}_{N\text{-times}}$ and its state is described by an N -fold tensor product state $\underbrace{\rho_\theta \otimes \cdots \otimes \rho_\theta}_{N\text{-times}}$. Then, the likelihood function produced by the outcomes $\vec{x} = (x_1, \dots, x_N) \in \mathcal{X}^N$ from a sequence of N independent POVMs $M(g)$ is

$$\begin{aligned} L(\theta | \vec{x}; g) &= \text{Tr} [(M_g(x_1) \otimes \cdots \otimes M_g(x_N)) \rho_\theta \otimes \cdots \otimes \rho_\theta] = \prod_{i=1}^N \text{Tr} [M_g(x_i) \rho_\theta] = \prod_{i=1}^N p(x_i | \theta; g) \\ &= p(0 | \theta; g)^m p(1 | \theta; g)^{N-m}, \end{aligned} \quad (28)$$

where m is the number of 0's in \vec{x} , and

$$p(x | \theta; g) = \begin{cases} \text{Tr} [M_g(1) \rho_\theta] = \frac{1}{2} [1 + \sin(\theta - g) \sqrt{F_Q}] , & x = 1, \\ \text{Tr} [M_g(0) \rho_\theta] = \frac{1}{2} [1 - \sin(\theta - g) \sqrt{F_Q}] , & x = 0. \end{cases} \quad (29)$$

By looking for the MLE of θ , denoted $\hat{\theta}_{\text{MLE}}$, we obtain:

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta \in \Theta} L(\theta | \vec{x}; g) = \arcsin \left(\frac{1}{\sqrt{F_Q}} \left[1 - \frac{2m}{N} \right] \right) + g, \quad (30)$$

which returns two values in $\Theta = [0, 2\pi)$, indicating the the likelihood is non-identifiable. Thus, the Fisher information alone is not enough to characterize the error of this estimation strategy. The previous discussion also can be found in [7].

Now we consider the case of POVMs with $k \geq 3$ outcomes. Any element of a POVM $M = \{M(i)\}$ with outcomes in a set $\mathcal{X} = \{1, 2, \dots, k\}$ can be written as $M(i) = f_0^{(i)} I + \vec{f}^{(i)} \cdot \vec{\sigma}$, where $\sum_{i=1}^k f_0^{(i)} = 1$ and $\sum_{i=1}^k \vec{f}^{(i)} = \vec{0}$. When a qubit is in a state ρ_θ , the probability of measuring the outcome i is

$$p(i | \theta) = \text{Tr} [\rho_\theta M(i)] = f_0^{(i)} + \vec{a}(\theta) \cdot \vec{f}^{(i)}. \quad (31)$$

A necessary and sufficient condition for identifiability is that, for any parameters $\theta_1, \theta_2 \in \Theta$, the set of equations $p(i | \theta_1) = p(i | \theta_2)$, $i = 1, \dots, k$ admits only one solution in the parametric space Θ [21].

The problem of identifiability could be avoided by considering POVMs with several outcomes because it is easier to satisfy the requirements for identifiability given above. However, a POVM that produces an identifiable likelihood is not necessarily optimal. Outside the optimal initial conditions, there is no measurement M that does not depend on θ and that saturates the QCRB [1]. In [7] it is shown a family of identifiable POVMs with more than 2 outcomes, however, except for the optimal initial condition, this family is not locally optimal. The results in [1, 7] do not forbid the existence of an identifiable and locally optimal POVM with more than 2 outcomes, but we could not find one. Note also that for a small number of measurements, the locally optimal POVM does not necessarily saturate the QCRB. Nevertheless it is possible to built a POVM that, for one measurement, produces an identifiable model that minimizes the mean square error. We will review now how to build this POVM by using the covariant approach [16]. Once we show how this POVM is built, we proceed with our proposal to achieve the QCRB for qubit phase estimation.

3.2 Covariant phase estimation for qubits

The following discussion is based on [17], where M_* is deduced for any shift parameter.

As we have previously mentioned, phase estimation estimation is covariant with trivial stabilizer group since the group $G = [0, 2\pi)$ equipped with addition modulo 2π acts over itself. Then, Θ is isomorphic to G . Moreover, $U = \{U_\theta\}_{\theta \in [0, 2\pi)}$ is a G -unitary representation of \mathcal{H} . Using Eq. (18) the covariant POVM M to estimate θ has the form

$$M(d\hat{\theta}) = e^{-i\hat{\theta}\vec{n}\cdot\vec{\sigma}/2} P_o e^{i\hat{\theta}\vec{n}\cdot\vec{\sigma}/2} \frac{d\hat{\theta}}{2\pi}, \quad (32)$$

where P_o is a positive operator satisfying Eq. (17).

The generator of the unitary representation is the operator $J := \vec{n}\cdot\vec{\sigma}/2$, which has a spectral decomposition

$$J = \frac{1}{2} \left| \frac{1}{2} \right\rangle \left\langle \frac{1}{2} \right| - \frac{1}{2} \left| -\frac{1}{2} \right\rangle \left\langle -\frac{1}{2} \right|, \quad (33)$$

where we have used that $\left| \pm \frac{1}{2} \right\rangle \left\langle \pm \frac{1}{2} \right| = \frac{1}{2} (I \pm \vec{n}\cdot\vec{\sigma})$. In the basis $\left\{ \left| \frac{1}{2} \right\rangle, \left| -\frac{1}{2} \right\rangle \right\}$, one can express $M(d\hat{\theta})$ as

$$M(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \sum_{m,n \in \{-\frac{1}{2}, \frac{1}{2}\}} e^{i\hat{\theta}(n-m)} p_{nm} |m\rangle \langle n|, \quad (34)$$

where $p_{mn} = \langle m|P_o|n\rangle$. Thus, any covariant POVM is characterized by the real numbers $0 \leq p_{nm} \leq 1$. Then, by Eq. (7)

$$P(\hat{\theta} \in B|\theta) = \int_B \frac{d\hat{\theta}}{2\pi} \sum_{m,n \in \{-\frac{1}{2}, \frac{1}{2}\}} e^{i(n-m)(\hat{\theta}-\theta)} p_{mn} \langle n|\rho|m\rangle. \quad (35)$$

Note that the measure (35) is a 2π -periodic probability distribution, so it is natural to consider the moments of the random variable $e^{i\hat{\theta}}$ instead. The first moment for a circular distribution $p(\phi)$ is defined as

$$E \left[e^{i\phi} \right] = \int_0^{2\pi} e^{i\phi} p(\phi) d\phi, \quad (36)$$

so from here we can estimate the phase as $\bar{\phi} = \text{Arg} \left(E \left[e^{i\hat{\theta}} \right] \right)$.

To quantify the correct dispersion of the estimates we use the Holevo variance [16]

$$V_M^H(\hat{\theta}) = \mu^{-2} - 1, \quad (37)$$

where $M \in \mathcal{M}(\Theta, U)$ and $\mu = \left| E \left[e^{i\hat{\theta}} \right] \right|$. If one has a biased estimator, then $\mu = E \left[\cos(\hat{\theta} - \theta) \right]$ [3]. A circular estimator $\hat{\theta}$ is unbiased if $e^{i\hat{\theta}} \propto E \left[e^{i\hat{\theta}} \right]$. For narrowly peaked and symmetric distributions around θ , $V_M^H(\hat{\theta}) \sim \text{MSE}_M(\hat{\theta})$. As a result, Holevo's variance is lower bounded by

$$V_M^H(\hat{\theta}) \geq \frac{1}{F_Q(\theta)}.$$

As the quantum Fisher information reaches its largest values for the families of pure states, let us restrict our attention only to pure quantum probes. Moreover, when the probe is a pure state, one can find the covariant POVM that minimizes the Holevo variance using the spectrum of J for any shift parameter [16, 17]. Here, we adapt the proof of [16] to the problem of phase estimation.

Theorem 3. Let $\rho_\theta = U_\theta \rho U_\theta^\dagger$, with $\rho = |\psi\rangle\langle\psi|$ and

$$M_*(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \sum_{m,n \in \text{Spec}(J)} e^{i\hat{\theta}(n-m)} \frac{P_m \rho P_n}{\sqrt{\text{Tr}[P_m \rho] \text{Tr}[P_n \rho]}}, \quad (38)$$

where $P_m = |m\rangle\langle m|$ is the associated projector to the eigenvalue m of J . Then, for any $M \in \mathcal{M}(\Theta, U)$,

$$V_M^H(\hat{\theta}) \geq V_{M_*}^H(\hat{\theta}). \quad (39)$$

Proof. In Dirac notation, $\rho = |\psi\rangle\langle\psi|$ so that Eq. (38) takes the following form

$$M_*(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \sum_{m,n \in \text{Spec}(J)} e^{i\hat{\theta}(n-m)} p_{mn}^* |m\rangle\langle n|, \quad (40)$$

with $p_{mn}^* = \frac{\langle m|\psi\rangle\langle\psi|n\rangle}{|\langle m|\psi\rangle||\langle\psi|n\rangle|}$. On other hand, from Eq. (34), we have

$$E_{M,\theta}[e^{ik\hat{\theta}}] = \sum_{\substack{m,n: \\ |m-n|=k}} p_{mn} \langle\psi|m\rangle\langle n|\psi\rangle e^{-i(n-m)\theta}. \quad (41)$$

Since $|p_{mn}| \leq 1$,

$$E_{M,\theta}[e^{ik\hat{\theta}}] \leq \sum_{\substack{m,n: \\ |m-n|=k}} \langle\psi|m\rangle\langle n|\psi\rangle e^{-i(n-m)\theta} \leq \sum_{\substack{m,n: \\ |m-n|=k}} |\langle\psi|m\rangle| |\langle n|\psi\rangle| e^{-i(n-m)\theta}. \quad (42)$$

where the equality is achieved if and only if $p_{mn} = p_{mn}^*$. Setting $k = 1$ yields the assertion \square .

Besides, if $\langle m|\psi\rangle$ is a constant for all $m \in \text{Spec}(J)$, the measurement M_* has information about θ equal to the quantum Fisher information [17]. For qubits, the previous condition is equivalent to have an initial condition $\vec{a} \cdot \vec{n} = 0$.

In this context, we now prove that the measurement M_* maximizes the Fisher information over the set of covariant POVMs.

Theorem 4. Let $\rho_\theta = e^{-i\theta\vec{n} \cdot \frac{\vec{\sigma}}{2}} \rho e^{-i\theta\vec{n} \cdot \frac{\vec{\sigma}}{2}}$, with

$$\rho = |\psi\rangle\langle\psi| = \frac{1}{2} [I + \vec{a} \cdot \hat{n}], \quad \|\vec{a}\| = 1.$$

Then, for any $M \in \mathcal{M}(\Theta, U)$,

$$F(\theta; M) \leq F(\theta; M_*), \quad (43)$$

where $M_*(d\hat{\theta})$ is defined according to Eq. (38).

Proof. Writing explicitly Eq. (35) we obtain

$$p(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \left[1 + 2\text{Re} \left(e^{-(\hat{\theta}-\theta)} \cdot C \right) \right], \quad (44)$$

where $C = p_{\frac{1}{2}-\frac{1}{2}} \left\langle -\frac{1}{2}|\psi\rangle \left\langle \psi\right| \frac{1}{2} \right\rangle$ is a complex number with module $|C|$ and phase φ . Thus,

$$p(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \left[1 + 2|C| \cos \left((\hat{\theta} - \theta) + \varphi \right) \right]. \quad (45)$$

Therefore the Fisher information,

$$F(M) = \frac{1}{2\pi} \int_{\Theta} d\hat{\theta} \frac{4|C|^2 \sin^2((\hat{\theta} - \theta) + \varphi)}{1 + 2|C| \cos((\hat{\theta} - \theta) + \varphi)} = \frac{4|C|^2}{2\pi} \int_{\Theta} dx \frac{\sin^2(x)}{1 + 2|C| \cos(x)}. \quad (46)$$

The value of the expression given by Eq. (46) increases monotonically as $|C|$ increases. As $|\langle \pm \frac{1}{2} | \psi \rangle| = \sqrt{\frac{1}{2}(1 \pm (\vec{a} \cdot \vec{n}))}$, then $|C| = |\leq \frac{1}{2} \sqrt{F_Q}$, and the maximum value for $F(M)$ is attained by M_* . \square

To find an explicit expression of $p(d\hat{\theta}|\theta)$, we first notice that

$$\left| \pm \frac{1}{2} \right\rangle \left\langle \pm \frac{1}{2} | \psi \right\rangle \left\langle \psi | \mp \frac{1}{2} \right\rangle \left\langle \mp \frac{1}{2} | \right\rangle = \frac{1}{4} [2(\vec{a} \cdot \vec{\sigma}) + i(\vec{a} \times \vec{n}) \cdot \vec{\sigma} - (\vec{n} \cdot \vec{\sigma})(\vec{n} \cdot \vec{a})], \quad (47)$$

so that Eq. (38) can be rewritten as follows

$$M_*(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \left[I + \frac{1}{\sqrt{F_Q}} [(\vec{a} \cdot \vec{\sigma} - (\vec{n} \cdot \vec{\sigma})(\vec{n} \cdot \vec{a})) \cos(\hat{\theta}) - \sin(\hat{\theta})(\vec{a} \times \vec{n}) \cdot \vec{\sigma}] \right]. \quad (48)$$

Then, the measurement M_* yields the probability density function

$$p(d\hat{\theta} | \theta) = \frac{d\hat{\theta}}{2\pi} [1 + \sqrt{F_Q} \cos(\hat{\theta} - \theta)]. \quad (49)$$

As M_* is a POVM with an infinity number of outcomes, it is reasonable to get an identifiable statistical model independent of θ . To illustrate this fact, we perform a numerical simulation, generating random numbers with distribution (49). A particular likelihood function produced by (49) is shown in Fig. 1.

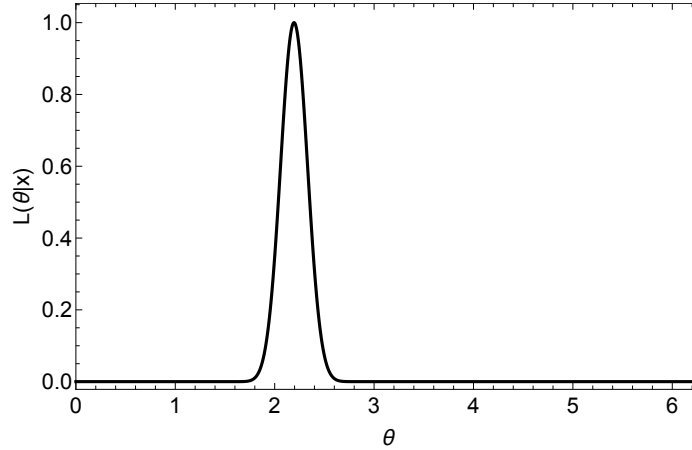


Figure 1: Plot of the likelihood function corresponding to the probability density Eq. (49) for the measure M_* and $N = 64$ probes. We have considered the value of the actual parameter to be $\theta = 2$ and used initial conditions $\vec{a} \cdot \vec{n} = 0$. For clarity we have rescaled the figure so that the maximum equals to one.

Moreover, according to Eq. (49), the Fisher information given M_* reads:

$$F(\theta; M_*) = \frac{1}{2\pi} \int_0^{2\pi} \frac{F_Q \sin^2(\hat{\theta} - \theta)}{1 + \sqrt{F_Q} \cos(\hat{\theta} - \theta)} d\hat{\theta} = 1 - \sqrt{1 - F_Q}. \quad (50)$$

When $\vec{a} \cdot \vec{n} = 0$, $F_Q = 1$ and therefore $F(\theta; M_*) = F_Q^{\max} = 1$. This is not surprising since, under this optimal initial condition, $|\langle \frac{1}{2}|\psi\rangle| = |\langle -\frac{1}{2}|\psi\rangle|$. Consequently, assuming this optimal setup, if the maximum likelihood is used for N independent copies of the system, one can attain the QCRB asymptotically for large N . Nevertheless, under the optimal initial condition, there are other covariant POVMs that can attain the QCRB. To see this, let us characterize all covariant measurements for quantum phase estimation of qubits.

Every positive bounded operator P_o can be expressed in the Bloch representation, that is, $P_o = d_0 + \vec{d} \cdot \vec{\sigma}$, where (d_0, \vec{d}) are real. When evaluating Eq. (17), it turns out that P_o forms a valid covariant POVM if $d_0 = 1$ and \vec{d} is orthogonal to \vec{n} . Therefore any covariant POVM has the form

$$M(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \left[I + \vec{d} \cdot \vec{\sigma} \cos(\hat{\theta}) - \sin(\hat{\theta})(\vec{d} \times \hat{n}) \cdot \vec{\sigma} \right]. \quad (51)$$

When a qubit is in the state ρ_θ , the outcomes for the POVM $M(d\hat{\theta})$ follows the probability density

$$p(d\hat{\theta}|\theta) = \frac{d\hat{\theta}}{2\pi} \left[1 + (\vec{a} \cdot \vec{d}) \cos(\hat{\theta} - \theta) - \vec{a} \cdot (\vec{d} \times \vec{n}) \sin(\hat{\theta} - \theta) \right]. \quad (52)$$

From this result, one can calculate the Fisher information associated with the covariant POVM M . In particular, the case $\vec{d} = \pm(\vec{n} \times \vec{a})$ is of interest, since the classical Fisher information reads

$$F(\theta; M|_{\vec{d}=\pm(\vec{n} \times \vec{a})}) = \frac{1}{2\pi} \int_0^{2\pi} \frac{F_Q^2 \cos^2(\hat{\theta} - \theta)}{1 \pm F_Q \sin(\hat{\theta} - \theta)} d\hat{\theta}. \quad (53)$$

As far as we are aware of, equations (51)-(53) are new results. Note that, in general, $F(\theta, M_*) \geq F(\theta; M|_{\vec{d}=\pm(\vec{n} \times \vec{a})})$, but for the particular case $\vec{a} \cdot \vec{n} = 0$, the POVM $M|_{\vec{d}=\pm(\vec{n} \times \vec{a})}$ is able to reach the maximum of the Fisher information, $F_Q^{\max} = 1$. Hence, the POVM given by Eq. (51) is equivalent to the measurement M_* .

Let us briefly summarise the results we have found thus far. Firstly, we have shown that estimation strategies based on locally optimal POVM with two outcomes cannot achieve the QCRB since the corresponding likelihood is non-identifiable. POVMs with more than two outcomes may be identifiable but we could not find the POVMs that are locally optimal (globally optimal POVMs do not exist [1]). Secondly, we have seen that for continuous and periodic outcomes the covariant POVM solves the identifiability problem and minimizes the MSE for one measurement, however, it can only achieve the QCRB (in the limit of a large number of measurements) with optimal initial conditions $\vec{a} \cdot \vec{n} = 0$. Hence, to the best of our knowledge, it does not exist a set of independent and identical measurements that achieve the QCRB.

It turns out that, under the optimal set of initial conditions, the covariant POVM M_* has been previously investigated and it is usually called *canonical phase measurement* [3, 17, 22] which, according to Eq. (40), takes the following form:

$$M_*^{\text{can}}(d\hat{\theta}) := \frac{d\hat{\theta}}{2\pi} \sum_{m,n \in \text{Spec}(J)} e^{i\hat{\theta}(n-m)} |m\rangle \langle n|. \quad (54)$$

Although the optimal covariant POVM is hard to realize experimentally, it can be well approximated by POVMs with large number of elements or adaptive measurements [7, 22,

30]. In particular, the canonical phase measurement can be implemented using adaptive measurements with quantum feedback [22], and the POVM M_* can be well approximated using a POVM with $k \geq 8$ that discretizes Eq. (49).

3.3 Entangled measurement

It is actually possible to achieve the QCRB for a probe not orthogonal to the rotation axis if we relax the condition of identical and independent set of measures. Let us indeed see it is possible to attain the QCRB for entangled measurements for any initial condition. Following [17], we estimate $\theta \in [0, 2\pi)$ for the family of states $\rho_\theta^{\otimes N}$ in the Hilbert space $\mathcal{H}^{\otimes N}$ with $\rho_\theta = U_\theta \rho U_\theta^\dagger$ and ρ being a probe pure qubit. Note that $\rho_\theta^{\otimes N}$ can be written as $\rho_\theta^{\otimes N} = e^{-i\theta J^{(N)}} \rho^{\otimes N} e^{i\theta J^{(N)}}$, where

$$J^{(N)} = \left(\hat{n} \cdot \frac{\vec{\sigma}}{2} \right) \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes \left(\hat{n} \cdot \frac{\vec{\sigma}}{2} \right), \quad (55)$$

that is, the family of states $\rho_\theta^{\otimes N}$ in $\mathcal{H}^{\otimes N}$ is covariant with respect to the unitary representation $e^{-i\theta J^{(N)}}$. Therefore, as in the case of one probe, the POVM that minimizes the Holevo variance is expressed in terms of the spectrum of $J^{(N)}$. Therefore, the measurement $M_*^{(N)}$ that minimizes the Holevo variance has the expression

$$M_*^{(N)}(d\hat{\theta}) = \frac{d\hat{\theta}}{2\pi} \left[\sum_{\substack{\lambda, \lambda' \in \\ \text{Spec}(J^{(N)})}} e^{i\hat{\theta}(\lambda' - \lambda)} \frac{P_\lambda \rho^{\otimes N} P_{\lambda'}}{\sqrt{\text{Tr}[P_\lambda \rho^{\otimes N}] \text{Tr}[P_{\lambda'} \rho^{\otimes N}]}} \right], \quad (56)$$

where, P_λ is the projection operator of $J^{(N)}$ associated to the eigenvalue λ .

The difference here, in contrast to the single measurement case, is that here the spectrum of $J^{(N)}$ is degenerate. Thus

$$P_* = \int_{\Theta} M(d\hat{\theta}) \neq I,$$

and therefore $M_*^{(n)}$ is not a resolution of identity. However, this is not a real problem because it can be extended to a POVM by adding $I - P_*$ in the orthogonal complement to the subspace $\mathcal{H}_* := P_*(\mathcal{H}^{\otimes N})$. So, the optimal POVM is

$$M_*^{(N)} := P_* \oplus I - P_*.$$

In particular, if $N = 1$, $M_*^{(N)} = M_*$.

As stated in [17], $M_*^{(N)}$ can achieve the QCRB for any probe state in the asymptotic limit. Besides, the scaling of the error is proportional to the square in the number of probes (Heisenberg scaling). Unfortunately, $M_*^{(N)}$ is an entangled measurement. Hence, if one wants to construct an entangled POVM, it is necessary to have N copies of the probe and perform quantum operations over all of them, making the implementation of this type of POVMs an experimental challenge.

3.4 Adaptive state quantum estimation (AQSE)

As we mentioned in the introduction, unlike its classical counterpart, quantum parameter estimation yields estimators which depend on the parameter one wants to estimate. A way to tackle this problem is using an adaptive quantum estimations scheme [13, 25, 35]. It

works as follows: suppose one has a set of optimal POVMs $\{M_g\}_{g \in \Theta}$. One begins with an arbitrary initial guess g_0 . Then one applies the optimal measurement at g_0 , M_{g_0} . Assuming the data x_1 is observed, one applies the MLE to the likelihood function $L_1(\theta | x_1; g_0) = p(x_1 | \theta; g_0)$ to obtain an estimate $\hat{\theta}_1(x_1) = g_1$. This process is then repeated iterative. For $n \geq 2$, one applies the POVM $M_{g_{n-1}}$, with $g_{n-1} = \hat{\theta}_{n-1}(x_1, \dots, x_{n-1})$, obtaining the outcome x_n . Here $\hat{\theta}_{n-1}$ is the estimation from the previous step using the outcomes x_1, \dots, x_{n-1} . The likelihood function for the n th step is

$$L_n(\theta | x_1, \dots, x_{n-1}; g_{n-1}) = \prod_{i=1}^n p(x_i | \theta; g_{i-1}), \quad (57)$$

where x_i is the data observed at step i . Applying the MLE one obtains the n th guess $g_n = \hat{\theta}_n(x_1, \dots, x_n)$. Even though this method relaxes the condition of identical measurements, this does not guarantee that the resulting classical statistical model is regular thus impacting its performance [8].

To illustrate this let us revisit the problem of non-identifiability for the POVM M_g previously discussed. To fix ideas we take $\theta = 2$ and consider $N = 64$ measurements. In panel A of Fig 2 we show the result of the likelihood function for $g = 1.5$ using Eq. (29). The likelihood is indeed non-identifiable and presents two global maxima in the interval $[0, 2\pi)$. Panel B in the same figure shows the results of using AQSE starting with the initial value $g_0 = 1.5$. In this case we can see that there is only one global maxima but at the incorrect value of θ . This is the result of having an original non-identifiable model, whose wrong maximum has been enhanced due the particular random trajectory in the parameter space generated by AQSE. Thus, even though AQSE is able to lift a possible degeneracy of the global maxima there is no a priori way to control on how to enhance the correct maximum.

With a modest amount of foresight, it seems fairly natural to expect that if we knew

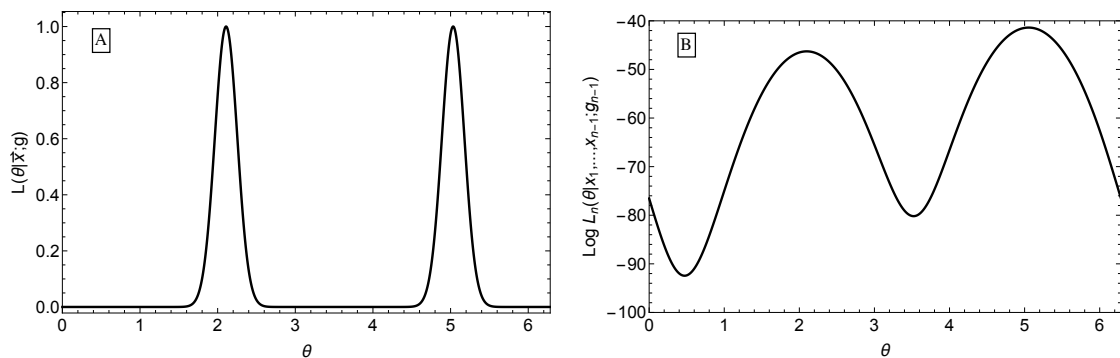


Figure 2: Plot of the likelihood functions for the two-outcome POVM M_g given by Eq. (25) and considering $\theta = 2$ and $N = 64$ probes. Panel (A) shows the likelihood function given by formula (28). Here we have considered $g = 1.5$ and the number of ones was 32 out of the possible $N = 64$. For clarity the figure has been rescaled so that the maxima equal to unity. Panel (B) shows the log-likelihood function for the AQSE scheme given by Eq. (57) using 64 adaptive steps. The existence of trajectories in the parameter space enhancing the incorrect maximum explains why the AQSE does not reach the QCRB. In both cases, the initial condition has been taken to be $\vec{a} \cdot \vec{n} = 0.5$.

in which interval the parameter lies and restricted the likelihood over that interval -so as to ensure that the restricted likelihood is identifiable- then the adaptive scheme will converge to the QCRB fast. To analyse this we have compared the behaviour of AQSE in two different scenarios: unrestricted and restricted likelihoods. In both cases we have taken $\theta = \pi$. For the unrestricted scheme the parameter space is $\Theta = [0, 2\pi)$, while for the restricted one we take $\theta \in \Theta = [\frac{\pi}{2}, \frac{3\pi}{2})$, so that the likelihood is identifiable. In Fig. 3 we

show the result of these two cases for two extremal initial conditions $\vec{a} \cdot \vec{n} = 0.5$ (panel A) and $\vec{a} \cdot \vec{n} = 0$ (panel B), corresponding to the lowest and largest value of F_Q , respectively. In both cases the numerical method was done using a bootstrap simulations with 10000 repetitions.

From these results we can observe that in the unrestricted case the lack of identifiability directly affects the scaling of the Holevo variance as a function of the number of adaptive steps. Furthermore, numerical results show that AQSE does not saturate the QCRB. On the other hand, in the restricted case, we find that the MLE saturates the QCRB around 128 measurements for both initial conditions. Thus we conclude that a possible deficient performance of AQSE is due to the non-identifiability for θ in the likelihood functions.

Finally, in Fig. 3, we have also compared the performance of AQSE with the optimal

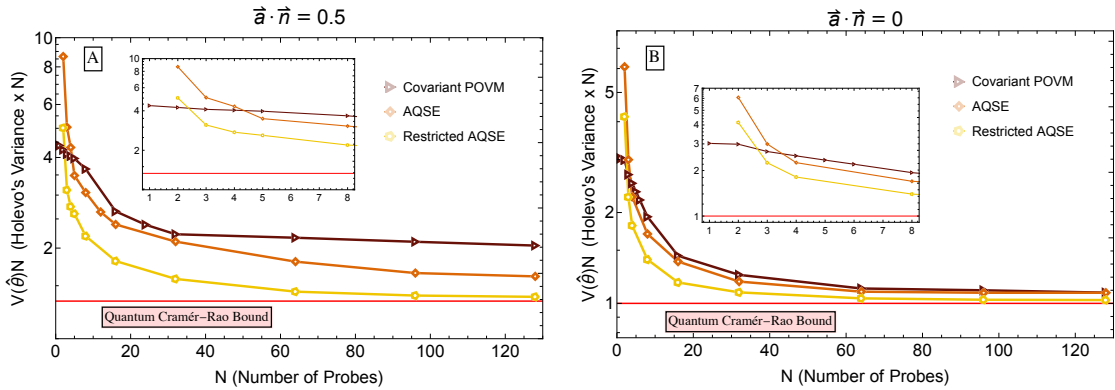


Figure 3: Result of Holevo's variance as a function of the number of probes for different estimation strategies and for two different initial conditions: $\vec{a} \cdot \vec{n} = 0.5$, corresponding to the smallest quantum Fisher information (shown in panel A) and $\vec{a} \cdot \vec{n} = 0$ (shown in panel B). In all cases we used 10000 sequences for each number of probes. As we can see the restricted AQSE performs much better than the AQSE. We have also included the strategy based on the optimal covariant POVM M_* as a benchmark. Notice that for both initial conditions the AQSE does not reach the QCRB, while for optimal initial condition the sequence of covariant POVMs is able to attain the QCRB (see panel B).

covariant inference. In this case, the sequence of M_* measurements has a better scaling for a small number of measurements (less than 4). Then, we can also conclude that the optimal covariant measurement is the best independent strategy in a regime of small N . This is expected because the covariant measurement is built to minimize the MSE for one measurement.

In the next section, we propose an estimation scheme that uses the covariant estimation and the AQSE method to avoid the problem of non-identifiable likelihood functions and thus is able to saturate the QCRB for any initial condition.

4 Adaptive estimation scheme with confidence intervals

From the previous discussion it seems clear that, conditioned of knowing in which interval the actual phase lies in and restricting the likelihood to that interval so it becomes identifiable, the AQSE method converges efficiently to the QCRB. With this in mind, we propose the following two-step scheme which allows us to initially guess the interval, thus making the iterative scheme a regular problem.

Our scheme consists of two main steps. In the first step one uses a sequence of independent optimal covariant measurements to produce a confidence interval (CI) using the maximum likelihood estimate $\hat{\theta}_{\text{MLE}}$. The CI gives a range of plausible values where the unknown

parameter most likely lies in. In the second step, one applies the AQSE method restricted to the confidence interval. Furthermore, one is able to update the center of the CI by the MLE of each previous step. The idea of using CIs to improve the error in estimations is exemplified, in a different context, in [4].

Let us assume that we have N copies of the state system ρ , so that each measurement is performed consecutively over each copy. Let us denote the result of each measurement as x_i . In the first part of our scheme we apply N_1 covariant measurements \mathcal{M}_1 given by Eq. (40), that is

$$\mathcal{M}_1 := M_*(dx_k), \quad (58)$$

and then we construct the MLE

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta \in \Theta} \prod_{i=1}^{N_1} p(dx_i | \theta) = \arg \max_{\theta \in \Theta} \prod_{i=1}^{N_1} \frac{dx_i}{2\pi} \left(1 + \sqrt{F_Q} \cos(x_i - \theta)\right) \quad (59)$$

to obtain the estimation $\hat{\theta}_{\text{MLE}}(x_1, \dots, x_{N_1})$. The CI is then given by [9]:

$$\text{CI}(\hat{\theta}_{\text{MLE}}(\vec{x})) = \left(\hat{\theta}_{\text{MLE}}(\vec{x}) - c \cdot F^{-\frac{1}{2}}, \hat{\theta}_{\text{MLE}}(\vec{x}) + c \cdot F^{-\frac{1}{2}}\right), \quad (60)$$

where c is the appropriate critical value in the standard normal distribution (e.g. 1.96 for 95% of confidence or 2.58 for 99% of confidence) and F is the Fisher information of M_* given by Eq. (50). To determine the minimum sample size N_1 in order to get a confidence interval of size $2E$, we use the CRB to write the lower bound

$$N_1 \geq \frac{c^2}{FE^2}. \quad (61)$$

In the second step of our scheme we then use the remainder of the copies, that is $N_2 = N - N_1$, to perform measurements using the POVM

$$\mathcal{M}_2 := M_{\hat{\theta}_{\text{MLE}}(x_1, \dots, x_k)}. \quad (62)$$

All in all, our two-step method is given by:

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta \in \Theta} \prod_{i=1}^{N_1} p(dx_i | \theta) \prod_{k=N_1}^{N_2} p(x_{k+1} | \theta; \hat{\theta}_{\text{MLE}}(x_1, \dots, x_k)), \quad (63)$$

with $\Theta = \text{CI}(\hat{\theta}_{\text{MLE}}(x_1, \dots, x_k))$. Eq. (63) is the main result of this work.

Before discussing some numerical results of our newly introduced method it is important to discuss the possible sources of errors associated to it. Suppose that we set a confidence level $0 \leq C_l \leq 1$ for a given marginal error E . Then we can write that

$$V^H(\hat{\theta}_{\text{MLE}}) = C_l \Delta_1 + (1 - C_l) \Delta_2, \quad (64)$$

where the two types of errors Δ_1 and Δ_2 correspond to the error associated to the AQSE method when the CI correctly includes the value of the actual parameters, and when it does not, respectively. Let us call the latter intervals bad CIs. Further, let us take $E < \frac{\pi}{2}$, so that likelihood functions used in the AQSE method are identifiable. Thus, the error Δ_1 when the CI includes the value of the parameter is lowered bounded by

$$\Delta_1 \geq \frac{1}{F_Q N_2 + N_1 F(\theta; M_*)}. \quad (65)$$

To characterize Δ_2 , first we note that there are two types of bad CIs. To see this, let θ_A be the parameter's value and $\theta_B = \theta_A + \pi$. Then, the likelihood function produced by AQSE has two local maxima, one around θ_A , and the other around θ_B . The first type of bad CIs corresponds to those that include the second maximum around θ_B . If one applies AQSE restricted to these CIs then θ_{MLE} tends to θ_B . The second type of bad CIs appear when the interval does not include θ_B . In this case, the maximum likelihood estimation produced by AQSE can tend to the point in the interval closest to either θ_A or to θ_B . When the estimation is closest to θ_A , by updating the interval's center with the posterior estimates, one can, in principle, obtain a confidence interval that captures the parameter's value. This explain why we obtain better results updating the center of the confidence interval.

To obtain a bad interval of type one, it is necessary that most of the data obtained from the covariant inference be numbers close to θ_B (the likelihood function has a global maximum around θ_B). From Eq. (49), for a covariant sample of size N , it follows that the probability of obtaining a likelihood function with the global maximum around θ_B is

$$\left[\int_{\theta_B - \epsilon}^{\theta_B + \epsilon} p(d\hat{\theta} | \theta) \right]^N = \left[\int_{\theta_B - \epsilon}^{\theta_B + \epsilon} \frac{d\hat{\theta}}{2\pi} \left(1 + \sqrt{F_Q} \cos(\hat{\theta} - \theta) \right) \right]^N. \quad (66)$$

Thus, for sufficiently large N , the above probability can be neglected. In this way, for a confidence level close to 1, and a sufficient small marginal error E , one expects that the number of bad confidence intervals disappear as the number of adaptive steps increases. To illustrate this fact, in Table 1, we show the number of bad intervals as a function of the adaptive steps in the AQSE part produced by a bootstrap simulation of 10000 repetitions. Here we have chosen a marginal error of $\frac{\pi}{4}$, with a confidence level of 0.99. In this case, we need to perform 11 and 22 covariant measurements for the initial conditions $\vec{a} \cdot \vec{n} = 0$ and $\vec{a} \cdot \vec{n} = 0.5$, respectively, to get the desired CI. For these values, the number of type one bad confidence intervals is negligible. Setting $\epsilon = E$ in (66), the probabilities of obtaining them are 2.3×10^{-10} and 2.3×10^{-18} for these two initial conditions, respectively.

Note that, the distribution for the estimates can be approximated as a mixture of 3 normal

AQSE Steps	# bad CIs ($\vec{a} \cdot \vec{n} = 0$)	# bad CIs ($\vec{a} \cdot \vec{n} = 0.5$)
0	111	125
4	50	56
8	19	20
16	8	1
32	6	1
48	2	0

Table 1: Number of bad confidence intervals as a function of adaptive steps for the AQSE part where the real value of θ was π .

distributions, $\mathcal{N}(\theta, \frac{1}{F_Q N_2 + N_1 F(\theta; M_*)})$, $\mathcal{N}(\theta + E, \frac{1}{F(\theta, M_*, M)})$, and $\mathcal{N}(\theta - E, \frac{1}{F(\theta, M_*, M)})$, where M is the sequence of AQSE and E is the marginal error. This implies that the lower bound for the Holevo variance, when the center of the CI is not updated, is given by

$$V^H(\hat{\theta}_{\text{MLE}}) \geq \frac{1}{F_Q N_2 + N_1 F(\theta; M_*)} + (1 - C_l) E^2. \quad (67)$$

The second term on the right-hand side of Eq. (67) dominates when $N_2 \rightarrow \infty$. This has a very simple interpretation: if the parameter is outside the confidence interval, making

more steps in the AQSE method does not diminish the error. Unfortunately, we are not able to provide a lower bound to the Holevo variance in the case for which the center of the CI is also updated. Nevertheless, as we will see, by setting a confidence level close to 1, a sufficiently small marginal error E , and by updating the center of the CIs, the Holevo variance approximates the QCRB in a small number of steps.

To assess the performance of our method we have performed a Monte Carlo simulations by setting $E = \frac{\pi}{4}$ and $C_l = .99$. Moreover, we have compared our numerical results with the ideal case ($C_l = 1$) and an estimation strategy for which the center of the CIs is not updated. The results are summarised in Figures 4 and 5 for initial conditions $\vec{a} \cdot \vec{n} = 0$ and $\vec{a} \cdot \vec{n} = 0.5$.

In Fig. 4 we have considered the initial condition $\vec{a} \cdot \vec{n} = 0$, so that $F(\theta; M_*) = 1$, and fixed $N_1 = 11$. The vertical dashed line at $N_1 = 11$ separates the point between covariant and AQSE inferences. Notice that in the covariant part, for $N \leq 11$, both strategies corresponding to either updating or not the CIs centers yields the same result. However, for $N > 11$, if the center of the CI is not updated, so that according to Eq. (67) its second term dominates, the estimation error grows quickly as the number of measurements increases. If, however, the center of the CIs is updated then the estimation error tends to the QCRB. Similarly in Fig. 5, for which we have chosen $N_1 = 22$, $\vec{a} \cdot \vec{n} = 0.5$ and $F(\theta; M_*) = 0.75$, we can observe the same qualitative behaviour, showing that our estimation strategy tends to the QCRB as the number of steps increases for any set of initial conditions.

We conclude by showing in Fig. 6 the performance of all the methods we have discussed

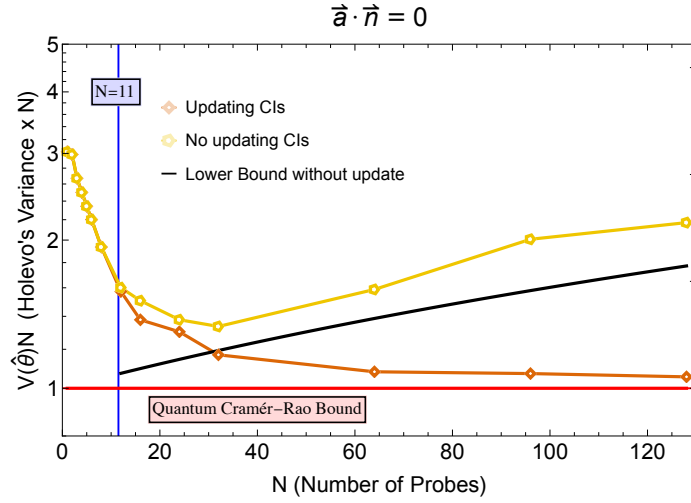


Figure 4: Plot showing Holevo's variance as a number of probes for our scheme. The optimal strategy (orange rhomboid markers), in which we update the centers of the CIs, tends to the QCRB as the number of probes increases. The latter bound corresponds to take a confidence level of 100%. We also show the alternative strategy for which the centers of the CIs are not updated (yellow pentagon markers). This result is lowered bounded by the solid black line, whose formula is given by Eq. (67). We have chosen the initial condition $\vec{a} \cdot \vec{n} = 0$. Here the vertical blue line at $N = 11$ marks the point separating the covariant inference performance from the AQSE inference to our two-step scheme.

throughout this work considering the initial conditions $\vec{a} \cdot \vec{n} = 0.5$ and $\vec{a} \cdot \vec{n} = 0$, which correspond to the smallest and largest quantum Fisher information, respectively. Obviously the optimal performance for the different estimation strategies is independent of θ , since the Holevo variance is invariant under translation modulo 2π over the parametric space. For all the strategies, the behavior is qualitatively the same for any initial condition. As expected, the entangled strategy is the best one and reaches the QCRB. The second best strategy

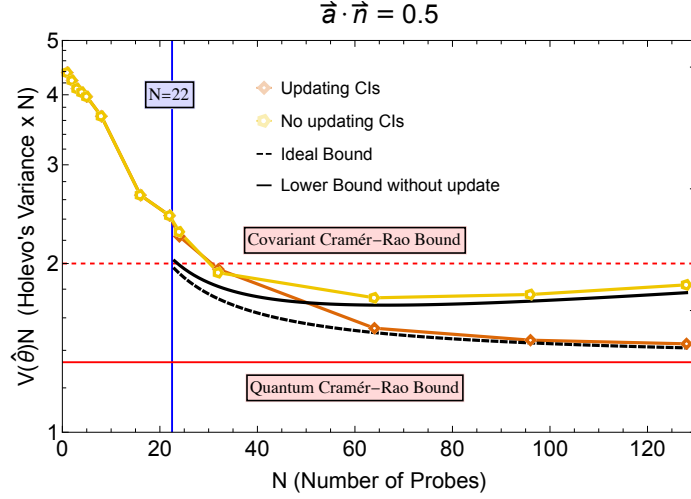


Figure 5: Plot showing Holevo's variance as a function of the number of probes N for our two-step adaptive scheme. The optimal strategy (orange rhomboid markers), in which we update the centers of the CIs, tends to the QCRB as the number of probes increases. This result is lowered bounded by the dashed black line, which is the QCRB corrected using Eq. (67) for a confidence level of 100%. We also show the alternative strategy for which the centers of the CIs are not updated (yellow pentagon markers). This result is lowered bounded by the solid black line, whose formula is given by Eq. (67). In this case, we have chosen the initial condition $\vec{a} \cdot \vec{n} = 0.5$. Here the vertical blue line at $N = 22$ marks the point separating the covariant inference performance from the AQSE inference to our two-step scheme. Finally, the horizontal dashed red line corresponds to the Cramér-Rao bound when only covariant measurements are used instead.

is the restricted ASQE, but this is unrealistic as it assumes that one knows beforehand an interval where the MLE is regular and includes the parameter. Our method, which we emphasize is the more realistic both mathematically and experimentally, is the third best strategy, as it reduces the error of the estimates improving the performance of AQSE and approximating the estimate to the QCRB.

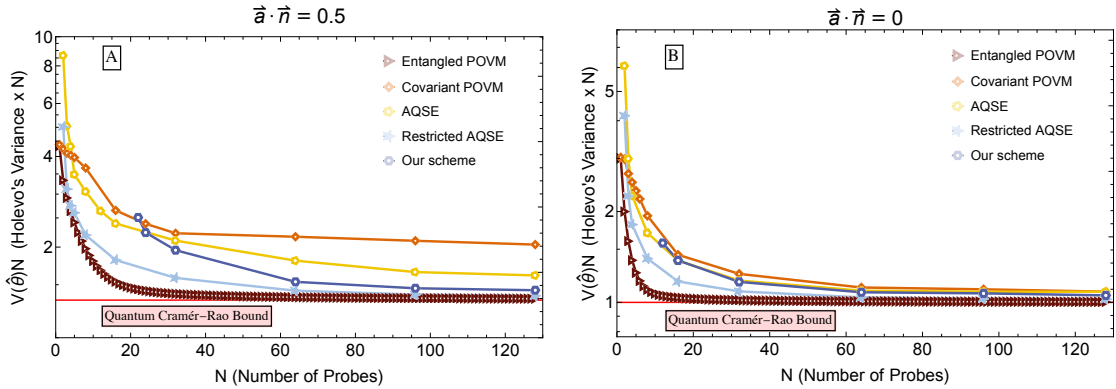


Figure 6: Holevo variance vs the number of probes for different strategies of estimation. In order to calculate the Holevo variance for the schemes based on the maximum likelihood estimator. A sequence of 10000 measurements were simulated in each point. The curve for the entangled measurement was analytically calculated. The curve for the proposed scheme shows the performance with a 99%-confidence intervals. The restricted AQSE method assume that the parametric space is an interval of length π that includes the real value of θ . The best strategy is the entangled measurement, followed by the restricted AQSE. These two strategies are unrealistic to implement. The third best strategy is our proposal. The y-axis is in log scale.

5 Summary and Conclusions

In this work we have thoroughly examined several strategies for quantum phase estimation. Their common denominator relies on maximizing likelihood density functions which are generally non-identifiable and/or non-optimal. This implies that the QCRB cannot be attained. We have developed a two-step strategy that circumvents this problem. Our method relies first on covariant measurements to identify a confidence interval within which the actual parameter is most likely to be, and then to apply an adaptive technique restricted to that interval. When compared with the current existing methods, ours is mathematically more robust, as it reaches the QCRB for any initial condition, and experimentally is more realistic, since neither an entangled measurement nor a priori information of the parameter's value are needed. Finally, we believe that our scheme can be generalized to any system so long as one can use a set of measurements to construct confidence intervals.

However, based on the results presented here and the current body of work, a number of questions remain open. First of all, it is not clear to us whether there exists a better strategy for which the subset of covariant measurements is intertwined with the subset of adaptive ones so as to identify the CI much faster. Secondly, the generalization to this work to multiparameter estimation does not seem straightforward, since although it is indeed possible to construct the covariant POVMs [16] for this case, little is known about non-identifiability for multiparameter likelihood functions. These, and other issues, will be addressed in forthcoming works.

6 Acknowledgments

We thank Laboratorio Universitario de Cómputo de Alto Rendimiento (LUCAR) of IIMAS-UNAM for their service on information processing. This research was supported by the Grant No. UNAM-DGAPA-PAPIIT IG100518 and IG101421, and Doctoral scholarship CONACYT 334231. I.P.C. acknowledges funding support from the London Mathematical Laboratory where he is an External Fellow.

A Code implementation

Our numerical results have been implemented using R language. An R library with the various methods discussed here can be publicly found in the repository [32]. To reproduce the numerical points for covariant strategy, AQSE and restricted-AQSE presented in Fig. 6, use the function:

```
Hvar_Scheme(theta_real, par_space, n_boost, num_prob, n,a, strategy),
```

set the value of θ , `theta_real` from 0 to 2π , set `par_space` as the vector `c(0, 2*pi)`, set the size of bootstrap `n_boost=1000`, vary the desired number of probes `num_prob`, set the initial unitary vectors for the probe and the axis of rotation `n a`, and vary the index `strategy` from 1 to 3. The index 1 represents the estimation strategy for independent covariant POVMs, 2 represents the AQSE strategy, 3 represents the restricted-AQSE strategy. To reproduce the points for our proposed estimation scheme use the function

```
ECI_Hvar(theta_real, par_space, n_boost, num_prob, n,a, C_lev, Margin_Err),
```

where you have to set the confidence level `C_lev` and marginal error `Margin_Err` to the desired values. Finally, to calculate the curve for the entangled estimation strategy use the function

Ent_Hvar(theta_real, par_space, n,a, num_prob).

References

- [1] O. E. Barndorff-Nielsen and R. D. Gill. Fisher information in quantum statistics. *J. Phys. A*, 33(24):4481–4490, 2000. DOI: [10.1088/0305-4470/33/24/306](https://doi.org/10.1088/0305-4470/33/24/306).
- [2] D. W. Berry and H. M. Wiseman. Adaptive measurements and optimal states for quantum interferometry. *Technical Digest - Summaries of Papers Presented at the Quantum Electronics and Laser Science Conference*, (5):60–61, 2001. DOI: [10.1109/QELS.2001.961853](https://doi.org/10.1109/QELS.2001.961853).
- [3] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman. How to perform the most accurate possible phase measurements. *Phys. Rev. A*, 80(5):1–22, 2009. DOI: [10.1103/PhysRevA.80.052114](https://doi.org/10.1103/PhysRevA.80.052114).
- [4] S. Boixo and R. D. Somma. Parameter estimation with mixed-state quantum computation. *Phys. Rev. A*, 77:052320, 2008. DOI: [10.1103/PhysRevA.77.052320](https://doi.org/10.1103/PhysRevA.77.052320).
- [5] S. L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439–3443, 1994. DOI: [10.1103/PhysRevLett.72.3439](https://doi.org/10.1103/PhysRevLett.72.3439).
- [6] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002. ISBN 9780534243128.
- [7] F. Chapeau-Blondeau. Optimizing qubit phase estimation. *Phys. Rev. A*, 94:022334, 2016. DOI: [10.1103/PhysRevA.94.022334](https://doi.org/10.1103/PhysRevA.94.022334).
- [8] D. Dee and A. Da Silva. Maximum-likelihood estimation of forecast and observation error covariance parameters. part I: Methodology. *Mon. Weather Rev.*, 127, 09 1998. DOI: [10.1175/1520-0493\(1999\)127<1822:MLEOFA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<1822:MLEOFA>2.0.CO;2).
- [9] M. H. DeGroot and M.J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012. ISBN 9780321500465. DOI: [10.1080/09332480.2013.845457](https://doi.org/10.1080/09332480.2013.845457).
- [10] J. P. Dowling. Quantum optical metrology – the lowdown on high-N00N states. *Contemporary Physics*, 49(2):125–143, 2008. DOI: [10.1080/00107510802091298](https://doi.org/10.1080/00107510802091298).
- [11] R. Engle and D. McFadden. *Handbook of Econometrics*, volume 4. North Holland, 1994. ISBN 0444887660,9780444887665.
- [12] F. Fröwis, M. Skotiniotis, B. Kraus, and W. Dür. Optimal quantum states for frequency estimation. *New Journal of Physics*, 16(8):083010, 2014. DOI: [10.1088/1367-2630/16/8/083010](https://doi.org/10.1088/1367-2630/16/8/083010).
- [13] A. Fujiwara. Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. *J. Phys. A*, 39(40):12489–12504, sep 2006. DOI: [10.1088/0305-4470/39/40/014](https://doi.org/10.1088/0305-4470/39/40/014).
- [14] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum-enhanced measurements: Beating the standard quantum limit. *Science*, 306(5700):1330–1336, 2004. DOI: [10.1126/science.1104149](https://doi.org/10.1126/science.1104149).
- [15] C. W. Helstrom. Quantum detection and estimation theory. *J. Stat. Phys.*, 1(2): 231–252, 1969. DOI: [10.1007/BF01007479](https://doi.org/10.1007/BF01007479).
- [16] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. Springer Science & Business Media, 1982. DOI: [10.1007/978-88-7642-378-9](https://doi.org/10.1007/978-88-7642-378-9).
- [17] A. S. Holevo. Asymptotic estimation of a shift parameter of a quantum state. *Theory of Probability and its Applications*, 49(2):207–220, 2005. DOI: [10.1137/S0040585X97981044](https://doi.org/10.1137/S0040585X97981044).
- [18] A.S Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4):337 – 394, 1973. DOI: [10.1016/0047-259X\(73\)90028-6](https://doi.org/10.1016/0047-259X(73)90028-6).

- [19] Z. Huang, K. R. Motes, P. M. Anisimov, J. P. Dowling, and D. W. Berry. Adaptive phase estimation with two-mode squeezed vacuum and parity measurement. *Phys. Rev. A*, 95:053837, 2017. DOI: [10.1103/PhysRevA.95.053837](https://doi.org/10.1103/PhysRevA.95.053837).
- [20] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer-Verlag, New York, NY, USA, second edition, 1998.
- [21] M. Manisera and P. Zuccolotto. Identifiability of a model for discrete frequency distributions with a multidimensional parameter space. *Journal of Multivariate Analysis*, 140:302–316, 2015. DOI: [10.1016/j.jmva.2015.05.011](https://doi.org/10.1016/j.jmva.2015.05.011).
- [22] L. S. Martin, W. P. Livingston, S. Hacoheh-Gourgy, and I. Wiseman, H. M. and Siddiqi. Implementation of a canonical phase measurement with quantum feedback. *Nature Physics*, 2020. DOI: [10.1038/s41567-020-0939-0](https://doi.org/10.1038/s41567-020-0939-0).
- [23] E. Merzbacher. *Quantum Mechanics*. North Holland, 3 edition, 1998. ISBN 0471887021, 9780471887027.
- [24] A. Monras. Optimal phase measurements with pure Gaussian states. *Phys. Rev. A*, 73:033821, 2006. DOI: [10.1103/PhysRevA.73.033821](https://doi.org/10.1103/PhysRevA.73.033821).
- [25] H. Nagaoka. *On the Parameter Estimation Problem for Quantum Statistical Models*, pages 125–132. 2005. DOI: [10.1142/9789812563071_0011](https://doi.org/10.1142/9789812563071_0011).
- [26] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: [10.1017/CBO9780511976667](https://doi.org/10.1017/CBO9780511976667).
- [27] C. Oh, C. Lee, C. Rockstuhl, H. Jeong, J. Kim, H. Nha, and S. Lee. Optimal gaussian measurements for phase estimation in single-mode gaussian metrology. *npj Quantum Information*, 5(1):1–9, 2019. DOI: [10.1038/s41534-019-0124-4](https://doi.org/10.1038/s41534-019-0124-4).
- [28] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi. Experimental demonstration of adaptive quantum state estimation. *Phys. Rev. Lett.*, 109:130404, 2012. DOI: [10.1103/PhysRevLett.109.130404](https://doi.org/10.1103/PhysRevLett.109.130404).
- [29] M. G. A. Paris. Quantum estimation for quantum technology. *International Journal of Quantum Information*, 7:125–137, 2009. DOI: [10.1142/S0219749909004839](https://doi.org/10.1142/S0219749909004839).
- [30] Y. Peng and H. Fan. Feedback ansatz for adaptive-feedback quantum metrology training with machine learning. *Phys. Rev. A*, 101:022107, 2020. DOI: [10.1103/PhysRevA.101.022107](https://doi.org/10.1103/PhysRevA.101.022107).
- [31] L. Pezzè and A. Smerzi. *Quantum theory of phase estimation*, pages 691–741. 2014. DOI: [10.3254/978-1-61499-448-0-691](https://doi.org/10.3254/978-1-61499-448-0-691).
- [32] M. A. Rodríguez-García. Qubit-phase-estimation. <https://github.com/Gateishion/Quantum-Phase-Estimation.git>, 2020.
- [33] F. Toscano, W. P. Bastos, and R. L. de Matos Filho. Attainability of the quantum information bound in pure-state models. *Phys. Rev. A*, 95:042125, 2017. DOI: [10.1103/PhysRevA.95.042125](https://doi.org/10.1103/PhysRevA.95.042125).
- [34] E. P. Wigner. *Symmetries and reflections*. Ox Bow Press., reprint edition edition, 1979.
- [35] K. Yamagata, A. Fujiwara, and R. D. Gill. Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Statist.*, (4):2197–2217, 08 . DOI: [10.1214/13-AOS1147](https://doi.org/10.1214/13-AOS1147).