A Geometric Perspective on Quantum Parameter Estimation

Jasminder S. Sidhu[1, a)] and Pieter Kok[1, b)]
Department of Physics and Astronomy, The University of Sheffield, Sheffield, S3 7RH, United Kingdom.

(Dated: 3 February 2020)

Quantum metrology holds the promise of an early practical application of quantum technologies, in which measurements of physical quantities can be made with much greater precision than what is achievable with classical technologies. In this review, we collect some of the key theoretical results in quantum parameter estimation by presenting the theory for the quantum estimation of a single parameter, multiple parameters, and optical estimation using Gaussian states. We give an overview of results in areas of current research interest, such as Bayesian quantum estimation, noisy quantum metrology, and distributed quantum sensing. We address the question how minimum measurement errors can be achieved using entanglement as well as more general quantum states. This review is presented from a geometric perspective. This has the advantage that it unifies a wide variety of estimation procedures and strategies, thus providing a more intuitive big picture of quantum parameter estimation.

CONTENTS

I. INTRODUCTION

When we measure a physical quantity, we need to quantify the error in that measurement, for example via the Mean Square Error (MSE). The actual MSE is hard to calculate directly, because it depends on the true unknown value of the quantity we want to measure. Remarkably, we can bound the MSE such that for any given experiment it cannot be smaller than a value that we *can* calculate. In addition, we can establish general conditions under which this bound can be saturated, i.e., the minimum MSE is in fact achieved. This is in broad strokes what parameter estimation theory is about.

With the classical theory of parameter estimation well-established in terms of the probability distribution of the observed data given the value of the physical quantity, we turn our attention to the space of probability distributions. Intuitively, when probability distributions move a large distance in this space under a change in value of the physical quantity, our measurements can pick up this change more easily than when the probability distribution moves hardly at all. This provides the link between measurement sensitivity (with correspondingly small MSE) and distance functions in the space of

[a)]Electronic mail: jsmdrsidhu@gmail.com
[b)]Electronic mail: p.kok@sheffield.ac.uk

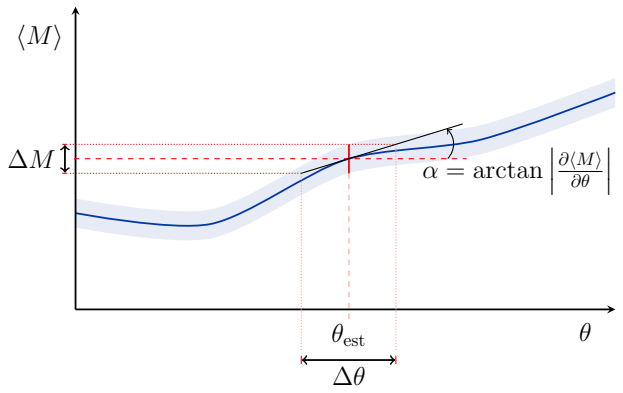


FIG. 1. We measure an observable M to obtain an estimate θ_{est} for a parameter of interest θ. We can relate the error $\Delta\theta$ in the estimate to the standard deviation ΔM by relating the tangent at θ_{est} with angle α to the ratio between ΔM (red vertical line) and the error $\Delta\theta$.

probabilities. Our task is to find a distance measure that allows us to make this link quantitative such that the distance measure can be directly related to the MSE. This gives us a metric in the space of probabilities, which is called the Fisher information. We will review key properties of the Fisher information.

The extension to quantum mechanics requires that we consider density operators instead of classical probability distributions. These operators also live in a linear vector space, and we can again construct distance measures in the space that connects the density operator to the minimal MSE in a physical experiment. However, since the density operators have a richer structure than the classical probability distributions (e.g., allowing non-commutative operators), the corresponding distance measures are also more complex. Based on the choice of inner product in the space of density operators, different distance measures can be constructed that have subtly different physical interpretations. The direct generalisation of the Fisher information, the *quantum* Fisher information, is a well-studied quantity that provides an attainable bound for the estimation of a single parameter that is imprinted on the quantum state via a unitary evolution. When multiple parameters come into play, the picture complicates considerably due to the potential non-commutativity of the observables that best estimate the individual parameters. Here, there are close links to generalised uncertainty relations. Alternative metrics can be more appropriate for different uses. For example, the Kubo-Mori information will be shown to have a particular relevance when thermal states are considered, and when we are interested in the relation between parameter estimation and conserved quantities we may turn to the Wigner-Yanase information.

Once we have established lower bounds for the error in the parameters, we consider how we can attain these bounds. For quantum estimation, entanglement plays an important, albeit subtle role. We review two important special cases in quantum estimation theory, namely estimation procedures using Gaussian states in quantum optics, and the extension of quantum estimation to the case where the evolution does not have a simple phase-like unitary structure. The final two sections are devoted to current areas of research interest, including Bayesian quantum estimation, noisy quantum metrology and some fault tolerant solutions, and distributed quantum sensing. Throughout this review, we aim to emphasise the bigger picture of quantum parameter estimation and how it very naturally arises from a geometrical view of quantum states. To achieve this we have prioritised practical examples and intuition over mathematical rigour. Readers interested in the technical mathematical details are referred to the references, and more mathematically inclined readers are referred to the collection of selected papers on quantum statistical inference by Hayashi[1].

II. CLASSICAL ESTIMATION THEORY

In this section, we briefly review classical estimation theory. We introduce the expectation value and the variance between the measured and the true value. We formulate a family of generic lower bounds that constrain the variance of parameter estimates, culminating in the well-known Cramér-Rao bound. For a general introduction to classical estimation theory, see Kay (1993)[2].

Several features of estimation theory can be understood by considering the following heuristic argument[3]: given a measurement of the observable M whose outcomes depend on a parameter θ, we can associate an estimate and error to θ. The average value $\langle M\rangle$ over many measurements of M depends on θ, and the error formula can be obtained from Fig. 1 as

$$\delta\theta = \frac{\Delta M}{|\partial_\theta\langle M\rangle|}, \tag{1}$$

where $\partial_\theta = \partial/\partial\theta$, and $\Delta M = [\langle M^2\rangle - \langle M\rangle^2]^{\frac{1}{2}}$ is the standard deviation in the measurement outcomes. The denominator $|\partial_\theta\langle M\rangle|$ can be viewed as a local correction in the units of $\delta\theta$. The steeper the tangent at θ_{est}, the more precise the estimate of θ at that point, i.e., the smaller $\delta\theta$. On the other hand, the larger the standard deviation ΔM, the lower the precision.

To illustrate the concepts of classical estimation theory, we use the example of the Mach-Zehnder interferometer (MZI), illustrated in Fig. 2. It can be used to make high precision measurements of a relative phase difference between two beams of light derived from some optical input state. This phase referenced method has become a standard tool in estimation theory and has received considerable attention given its applications in enhanced phase estimations in optical interferometry[4,5], frequency measurements[6,7], and biosensors[8,9].

Consider an experiment to estimate the relative phase difference in the interferometer, where a single photon is sent into one input mode of the MZI, and the two output modes are monitored with photodetectors. Each run of the experiment provides two pieces of data $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ that measure the photon counts in each detector, D_1 and D_2 respectively. Assuming no losses in the interferometer and ideal detectors, the probability distribution function $p(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}|\theta)$ has the following

possible outcomes:

$$p(1,1|\theta) = 0, \quad p(1,0|\theta) = \cos^2\left(\frac{\theta}{2}\right),$$
$$p(0,0|\theta) = 0, \quad p(0,1|\theta) = \sin^2\left(\frac{\theta}{2}\right), \tag{2}$$

with $\theta = \Phi - \phi$ the phase difference that we are interested in. If we count N_1 photons in detector D_1 and N_2 in detector D_2, a suitable estimator for θ is

$$\check{\theta} = \arccos\left(\frac{N_1 - N_2}{N}\right), \tag{3}$$

where we denote the estimator of θ by $\check{\theta}$ (the notation $\hat{\cdot}$ is reserved for quantum mechanical operators). The estimator is a function of the data $\check{\theta}(\mathbf{x})$ and returns a value for θ. A generalisation of this to an N-photon Fock state in one input mode of the MZI distributes the photons binomially over the output modes[10], with a more complicated corresponding estimator. If we inject a classical coherent state with intensity I_0 into one mode and measure the intensities I_1 and I_2 in detectors D_1 and D_2, the estimator becomes

$$\check{\theta} = \arccos\left(\frac{I_1 - I_2}{I_0}\right), \tag{4}$$

which is the continuous version of equation (3).

To calculate the precision of the measurement of θ in this experiment, we use the error propagation formula in Eq. (1), and the measurement operator is $M = \hat{n}_1 - \hat{n}_2$, with $\hat{n}_1$ and $\hat{n}_2$ the photon number operators for the output modes. Using the probability distribution in Eq. (2) we find that for N photons sent into the interferometer

$$\langle M \rangle = N\cos\theta \quad \text{and} \quad (\Delta M)^2 = N\sin^2\theta\,. \tag{5}$$

This yields a precision of

$$\delta\theta = \frac{\sqrt{N}\sin\theta}{|N\partial_\theta\cos\theta|} = \frac{1}{\sqrt{N}}\,. \tag{6}$$

This is the so-called shot noise limit for interferometry.

A. Fundamentals of estimation theory

The problem of estimating the value of a vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D)^\top$ from a set of observed data $\mathbf{x} = (x_1, \ldots, x_D)^\top$ is formally addressed in parameter estimation theory. Here, $\top$ denotes the transpose. Owing to experimental uncertainties and errors, the inference of parameters is related to the measurement outcomes through some conditional probability distribution $p(\boldsymbol{\theta}|\mathbf{x})$ that is derived from a model of the physical system under consideration. The main question is how well we can estimate these parameters. In other words, what is the best possible precision that we can achieve? There are generally two possibilities for $\boldsymbol{\theta}$:

1. $\boldsymbol{\theta}$ are not random, but unknown;

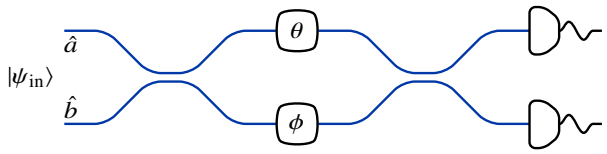


FIG. 2. The Mach-Zehnder interferometer, drawn with two evanescently coupled 50:50 beam splitters. Here, $\hat{a}$ and $\hat{b}$ define two input modes. An optical path difference between the internal arms of the interferometer results in a relative phase difference. The objective is to estimate the relative phase $\theta = \Phi - \phi$ by measuring the photon intensities at the output using the photodetectors on the right. Figure adapted with permission from Sidhu, J., Quantum metrology of grid deformations and squeezed light: with applications in quantum imaging & quantum information, Doctoral Thesis, University of Sheffield, Copyright 2018[11].

2. $\boldsymbol{\theta}$ are random and unknown.

In the first case we speak of *Fisher estimation*, while in the second case we have *Bayesian estimation*. An advantage of using Bayesian methods is that they do not rely on asymptotics to provide optimal performance, a property not enjoyed by Fisher estimation[12]. The optimal performance for both methods coincide in the large sample scenario[1]. Research efforts in quantum metrology and estimation theory have been dominated by work in the Fisher regime. To reflect this, we focus our review mainly on the estimation of non-random parameters. The subtle difference between Fisher and Bayesian estimation will be covered in subsection V H.

Given that the probability density function (PDF) of the data $p(\mathbf{x}|\boldsymbol{\theta})$ is known (i.e., we can model the physical process), the set of parameters $\boldsymbol{\theta}$ may be extracted from a set of observation data $\mathbf{x}$ via an estimator $\check{\boldsymbol{\theta}}$, which is a function of the observed data only. We will see that the estimators are also used to find the estimation errors.

An important estimator for a non-random $\boldsymbol{\theta}$ is the maximum likelihood estimator

$$\check{\boldsymbol{\theta}}(\mathbf{x}) = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x}), \tag{7}$$

where the likelihood function $p(\boldsymbol{\theta}|\mathbf{x})$ is the probability of $\boldsymbol{\theta}$ being the true value given the data set $\mathbf{x}$. This requires a model for the process in order to generate the probability $p(\boldsymbol{\theta}|\mathbf{x})$.

While the true value of $\boldsymbol{\theta}$ is an array of numbers, its estimates are *random variables*. This is due to the probabilistic nature of the data; two runs of an experiment with equal parameters $\boldsymbol{\theta}$ will not generate equal data due to statistical fluctuations: $\mathbf{x}_1 \neq \mathbf{x}_2$. Hence the estimates for both runs will differ from the actual values, and not be equal to each other. Given a very large measurement data set, an estimator that generates the true values of the parameters is referred to as a *perfect estimator*. In the next subsection, we introduce the covariance matrix as a natural measure of the performance of an estimator.

B. Expectation values and covariance

A natural figure of merit to quantify the performance of an estimator is the variance of a parameter estimate with respect

to its true value. Hence, we can characterise the estimation performance by searching for an estimator that has the smallest variance in parameter estimates. Although various other methods to characterise the performance exist, this is a natural choice that was first introduced by H. Cramér and C. R. Rao[13,14]. In the remainder of this section we assume that the measured data $\mathbf{x}$ is continuous without loss of generality.

For a multi-parameter estimation of $\boldsymbol{\theta}$, we may define the natural optimality criterion as the difference between the estimator and the true value of $\boldsymbol{\theta}$, $\Delta\boldsymbol{\theta} = \check{\boldsymbol{\theta}} - \boldsymbol{\theta}$. However, since this relies directly on the unknown true value $\boldsymbol{\theta}$, it is often more convenient to define the variation in the estimates of $\boldsymbol{\theta}$ as

$$\delta\check{\boldsymbol{\theta}} = \check{\boldsymbol{\theta}} - \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\check{\boldsymbol{\theta}}(\mathbf{x})\right], \tag{8}$$

where

$$\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\check{\boldsymbol{\theta}}(\mathbf{x})\right] = \int \mathrm{d}\mathbf{x}\; p(\mathbf{x}|\boldsymbol{\theta})\,\check{\boldsymbol{\theta}}(\mathbf{x}) \tag{9}$$

is the *expectation value* of the estimator $\check{\boldsymbol{\theta}}(\mathbf{x})$ with respect to the probability distribution $p(\mathbf{x}|\boldsymbol{\theta})$. When the estimator is unbiased (i.e., there is no systematic error in the estimator), the expectation value approaches the true value and $\delta\check{\boldsymbol{\theta}}$ reduces to $\Delta\boldsymbol{\theta}$ in the limit of large data sets.

We define the covariance matrix as

$$\begin{aligned}\mathrm{Cov}(\check{\boldsymbol{\theta}}) &= \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\delta\check{\boldsymbol{\theta}}(\mathbf{x})\;\delta\check{\boldsymbol{\theta}}(\mathbf{x})^{\top}\right] \\ &= \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\left(\check{\boldsymbol{\theta}} - \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\check{\boldsymbol{\theta}}\right]\right)\left(\check{\boldsymbol{\theta}} - \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\check{\boldsymbol{\theta}}\right]\right)^{\top}\right]. \end{aligned} \tag{10}$$

The covariance matrix depends on the parameters via the expectation value of the estimators.

The diagonal elements of the covariance matrix are the variances of the different parameters θ_j with $j \in \{1, \ldots, D\}$. Notice that the estimator's covariance is not the same as its mean square error matrix:

$$\begin{aligned}\mathbb{E}_{\mathbf{x}|\theta}\left[(\Delta\check{\boldsymbol{\theta}})(\Delta\check{\boldsymbol{\theta}})^{\top}\right] &= \mathbb{E}_{\mathbf{x}|\theta}\left[(\check{\boldsymbol{\theta}} - \boldsymbol{\theta})(\check{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\top}\right] \\ &= \mathrm{Cov}[\check{\boldsymbol{\theta}}] + b(\check{\boldsymbol{\theta}})b(\check{\boldsymbol{\theta}})^{\top},\end{aligned} \tag{11}$$

where $b(\check{\boldsymbol{\theta}}) = \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}[\check{\boldsymbol{\theta}}(\mathbf{x})] - \boldsymbol{\theta}$ is the bias of the estimator. We see that the mean square error is equal to the variance if and only if we have an *unbiased estimator*: $b(\check{\theta}_j) = 0$. Unbiased estimators ensure that the average of the estimates converge to the true value of the parameter: $\langle\check{\boldsymbol{\theta}}\rangle = \boldsymbol{\theta}$. While we include the effect that biased estimators have on the estimation precision in subsection IV E, we assume unbiased estimators for the rest of the review. In the next subsection we define the expectation value and covariance of estimators.

C. Bounds on the covariance matrix

It is typically not possible to calculate the exact values of the covariance matrix. We can often only hope to place some limits on the MSE, the variance, and other quantities. In this section we will use the structure of the covariance matrix and arguments from information geometry to formulate generic bounds on the (elements of the) covariance matrix[15,16].

For notational convenience we abbreviate $\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}$ by $\mathbb{E}$ for the remainder of this section. For a positive semi-definite matrix product XY we have that $\mathbb{E}[XY] \geq 0$. We choose

$$X = f - Ag \qquad \text{and} \qquad Y = X^{\top}, \tag{12}$$

with f a D-dimensional real vector of the same size as $\boldsymbol{\theta}$, g an R-dimensional real vector, and A an $R\times D$ matrix that depends only on $\boldsymbol{\theta}$. We then find

$$\mathbb{E}\left[[f(\mathbf{x},\boldsymbol{\theta}) - Ag(\mathbf{x},\boldsymbol{\theta})][f(\mathbf{x},\boldsymbol{\theta}) - Ag(\mathbf{x},\boldsymbol{\theta})]^{\top}\right] \geq 0\,. \tag{13}$$

Since the expectation value is linear, we can expand this into

$$\mathbb{E}\left[ff^{\top}\right] - \mathbb{E}\left[fg^{\top}A^{\top}\right] - \mathbb{E}\left[Agf^{\top}\right] + \mathbb{E}\left[Agg^{\top}A^{\top}\right] \geq 0, \tag{14}$$

and extract the matrix A, which is a constant with respect to the expectation:

$$\mathbb{E}\left[ff^{\top}\right] \geq \mathbb{E}\left[fg^{\top}\right]A^{\top} + A\,\mathbb{E}\left[gf^{\top}\right] - A\,\mathbb{E}\left[gg^{\top}\right]A^{\top}. \tag{15}$$

When we redefine

$$T = \mathbb{E}\left[fg^{\top}\right] \qquad \text{and} \qquad G = \mathbb{E}\left[gg^{\top}\right], \tag{16}$$

we arrive at

$$\mathbb{E}\left[ff^{\top}\right] \geq TA^{\top} + AT^{\top} - AGA^{\top}. \tag{17}$$

This is a bound on the expectation value of $ff^{\top}$ (which at this point can be anything that is consistent with the general definition of f), and to make the bound as tight as possible we must maximise the right-hand side of Eq. (17).

Since T and G do not depend on $\mathbf{x}$ (they are averaged over the data set), we can choose $A = TG^{-1}$ (where we require that the inverse of G exists; this places a restriction on g). This leads to the compact form

$$\mathbb{E}\left[ff^{\top}\right] \geq TG^{-1}T^{\top}. \tag{18}$$

Next we choose $f = \check{\boldsymbol{\theta}}(\mathbf{x}) - \mathbb{E}[\check{\boldsymbol{\theta}}]$, so that $\mathbb{E}[ff^{\top}] = \mathrm{Cov}(\check{\boldsymbol{\theta}})$ and therefore

$$\mathrm{Cov}(\check{\boldsymbol{\theta}}) \geq TG^{-1}T^{\top}. \tag{19}$$

This expression is valid for *any* estimator $\check{\boldsymbol{\theta}}(\mathbf{x})$. The matrix G is called the *information matrix*. Different definitions of g (and thus G) will produce different bounds that may have various advantages (computational, tightness, etc.). The matrix G is the expectation of a projector $gg^{\top}$ that may not be full rank (and therefore has no inverse). This typically happens when the estimator does not have enough degrees of freedom and therefore cannot provide estimates of all parameters.

For the above choice of f the T matrix becomes

$$\begin{aligned}T &= \mathbb{E}\left[\left(\check{\boldsymbol{\theta}}(\mathbf{x}) - \mathbb{E}[\check{\boldsymbol{\theta}}]\right)g^{\top}(\mathbf{x},\boldsymbol{\theta})\right] \\ &= \mathbb{E}\left[\check{\boldsymbol{\theta}}(\mathbf{x})g^{\top}(\mathbf{x},\boldsymbol{\theta})\right] - \mathbb{E}\left[\boldsymbol{\theta}g^{\top}(\mathbf{x},\boldsymbol{\theta})\right].\end{aligned} \tag{20}$$

We restrict ourselves to choices of g that satisfy the condition $\mathbb{E}_{\boldsymbol{\theta}|\mathbf{x}}[g^\top(\mathbf{x},\boldsymbol{\theta})] = 0$. Therefore, the first term in T becomes

$$\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\check{\boldsymbol{\theta}}(\mathbf{x})g^\top(\mathbf{x},\boldsymbol{\theta})\right] = \mathbb{E}_{\mathbf{x}}\left[\check{\boldsymbol{\theta}}(\mathbf{x})\,\mathbb{E}_{\boldsymbol{\theta}|\mathbf{x}}\left[g^\top(\mathbf{x},\boldsymbol{\theta})\right]\right] = 0\,. \quad (21)$$

As a result, the bounds on the covariance matrix are determined by the estimator $\check{\boldsymbol{\theta}}(\mathbf{x})$ and our choice of the function $g(\mathbf{x},\boldsymbol{\theta})$. Under suitable smoothness conditions on the probability distributions, we can define a family of definitions for the Fisher information to obtain different bounds[17]. In the next subsection, we introduce the most famous choice of g leading to the Fisher information and the Cramér-Rao bound.

D. The Cramér-Rao bound

The choice for $g(\mathbf{x},\boldsymbol{\theta})$ we consider here is

$$g(\mathbf{x},\boldsymbol{\theta}) = \frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,, \quad (22)$$

which requires that the first derivative of $p(\mathbf{x}|\boldsymbol{\theta})$ exists and is absolutely integrable. The function $g(\mathbf{x},\boldsymbol{\theta})$ is a natural choice in that it is additive for independent samples due to the logarithm (since independent events multiply probabilities), and the derivative measures the rate of change of the probability distribution with respect to the parameter of interest. The intuition is that a fast changing probability distribution with θ will produce a clearer change in measurement outcomes $\mathbf{x}$ as we vary θ.

As an example, we consider a single parameter θ such that g is a scalar function. Then we can evaluate T explicitly via partial integration:

$$T = -\,\mathbb{E}_{\mathbf{x}|\theta}\left[\theta\frac{\partial \ln p(\mathbf{x}|\theta)}{\partial\theta}\right] = 1\,. \quad (23)$$

The information matrix G becomes

$$G = \mathbb{E}_{\mathbf{x}|\theta}\left[\left(\frac{\partial \ln p(\mathbf{x}|\theta)}{\partial\theta}\right)^2\right] \equiv I(\theta), \quad (24)$$

which is better known as the Fisher information $I(\theta)$. The classical Fisher information is generally dependent on θ. If we have a model for the process under study, we can find $p(\mathbf{x}|\theta)$ and calculate the classical Fisher information directly. The variance in θ can then be bounded by

$$(\delta\theta)^2 \geq I^{-1}\,. \quad (25)$$

This is the Cramér-Rao Bound (CRB)[14,18]. It is saturated when

$$\check{\theta}(\mathbf{x}) - \theta = \frac{\partial \ln p(\mathbf{x}|\theta)}{\partial\theta}\,. \quad (26)$$

For a general D-parameter problem, the CRB is a matrix inequality

$$[\mathrm{Cov}(\check{\boldsymbol{\theta}})]_{ij} \geq \left[I(\boldsymbol{\theta})^{-1}\right]_{ij}\,, \quad (27)$$

where the inequality means that $\mathrm{Cov}(\check{\boldsymbol{\theta}}) - I(\boldsymbol{\theta})^{-1}$ is a positive semi-definite matrix. This bound is typically attainable using a Maximum Likelihood estimator in the asymptotic regime of many independent samples. Under the local asymptotic normality and quadratic mean differentiability conditions, it can be shown that no estimator can attain a smaller MSE that the Fisher information[17,19].

The Fisher information matrix is symmetric and positive, and it can be interpreted as the metric tensor in the parameter space. In particular, this means that when we reparameterize the space and wish to estimate the parameters $\boldsymbol{\vartheta} = (\vartheta_1 \ldots, \vartheta_D)$, with $\vartheta_j(\boldsymbol{\theta})$ some function of the original parameters $\boldsymbol{\theta}$, the corresponding transformed Fisher information matrix becomes

$$I(\boldsymbol{\vartheta}) = J^\top I(\boldsymbol{\theta}) J \quad \text{and} \quad J_{jk} = \frac{\partial\theta_j}{\partial\vartheta_k}\,, \quad (28)$$

where J is the Jacobian of the parameter transformation.

An important example for the Fisher information matrix is for a Gaussian (normal) distribution, which has a closed form[2]. Let the distribution be characterised by mean values $\boldsymbol{\mu}(\boldsymbol{\theta})$ and covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^D \det\boldsymbol{\Sigma}}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]\,. \quad (29)$$

The Fisher information matrix then takes the following closed form

$$[I(\boldsymbol{\theta})]_{jk} = \frac{\partial\boldsymbol{\mu}^\top}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta_k} + \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\right]. \quad (30)$$

Often, when estimating parameters there are several *nuisance* parameters that must also be estimated. We are not intrinsically interested in these parameters, but the estimators of the parameters of interest depend on them. We can separate the tuple of parameters $\boldsymbol{\theta}$ into genuine and nuisance parameters, $\boldsymbol{\theta} = (\boldsymbol{\theta}_g, \boldsymbol{\theta}_n)$. The Fisher information matrix can then be written in block form:

$$I(\boldsymbol{\theta}) = \begin{pmatrix} I(\boldsymbol{\theta}_g,\boldsymbol{\theta}_g) & I(\boldsymbol{\theta}_g,\boldsymbol{\theta}_n) \\ I(\boldsymbol{\theta}_n,\boldsymbol{\theta}_g) & I(\boldsymbol{\theta}_n,\boldsymbol{\theta}_n)\end{pmatrix}\,. \quad (31)$$

The CRB is still given by the inverse of I, but now we can use the inverse of a block matrix,

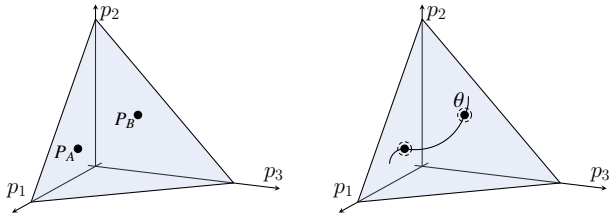
$$P_A\begin{pmatrix} A & D^\top \\ D & B\end{pmatrix}^{-1}P_A = \left(A - D^\top B^{-1} D\right)^{-1}\,, \quad (32)$$

where P_A is the projection operator onto the subspace occupied by A, to establish how the nuisance parameters affect the CRB:

$$\mathrm{Cov}(\check{\boldsymbol{\theta}}_g) \geq \left[I(\boldsymbol{\theta}_g,\boldsymbol{\theta}_g) - I(\boldsymbol{\theta}_g,\boldsymbol{\theta}_n)I(\boldsymbol{\theta}_n,\boldsymbol{\theta}_n)^{-1}I(\boldsymbol{\theta}_n,\boldsymbol{\theta}_g)\right]^{-1}\,. \quad (33)$$

In other words, the nuisance parameters *lower* the Fisher information matrix compared to the Fisher information matrix for the genuine parameters alone, as expected.

There are other choices for the function $g(\mathbf{x},\boldsymbol{\theta})$ that lead to different bounds. For more details on classical and Bayesian bounds, see Van Trees and Bell (2007)[20].



(a) The simplex. (b) Paths in the simplex.

FIG. 3. Fig. 3a: the probability simplex where the measurement has three outcomes with probabilities p_1, p_2, and p_3. Every probability distribution is represented by a point in the simplex (P_A and P_B). Fig. 3b: θ parametrises a path through the probability simplex.

III. GEOMETRY OF ESTIMATION THEORY

The Fisher information in Eq. (24) followed from our choice of $g(\mathbf{x}, \theta)$. In this section we will give an intuitive geometric derivation[21,22] that will help us with the derivation of the quantum Fisher information in the next section. We relate parameter estimation to methods of distinguishing probability distributions, including the Fisher information and the relative entropy.

A. The probability simplex

Any experiment used to infer a value of $\boldsymbol{\theta}$ will return different measurement outcomes $\mathbf{x} = (x_1, \ldots, x_D)$. Assuming that different values of $\boldsymbol{\theta}$ produce variations in the measurement outcomes (otherwise this particular measurement would not be useful in extracting a value of $\boldsymbol{\theta}$), we can posit a probability distribution $p(\mathbf{x}|\boldsymbol{\theta})$, which may originate from some physical model. The problem of finding the value of $\boldsymbol{\theta}$ is then reduced to telling the difference between two probability distributions $p(\mathbf{x}|\boldsymbol{\theta}_A)$ and $p(\mathbf{x}|\boldsymbol{\theta}_B)$. In other words, how many times do we have to sample the system (i.e., what is D) in order to tell the difference between $p(\mathbf{x}|\boldsymbol{\theta}_A)$ and $p(\mathbf{x}|\boldsymbol{\theta}_B)$?

The probability distributions $p(\mathbf{x}|\boldsymbol{\theta})$ form a space called a *probability simplex* (see Fig. 3 for a simple example). The probability distributions $p(\mathbf{x}|\theta)$ for a single parameter θ typically form a curve through the simplex that is parametrised by θ. In order to tell how many measurements we need to make in order to distinguish two probability distributions on the curve we need some distance measure (a metric) on the simplex that fits naturally with statistics. This metric can then be used to tell how far away two distributions are from each other. In turn this will allow us to infer how many measurements we need to make to distinguish between the two distributions. In what follows we first specifically consider a single parameter θ for simplicity.

In its general discrete form we can write the infinitesimal distance ds on the simplex in terms of incremental probability changes dp and a metric h:

$$\mathrm{d}s^2 = \sum_{jk} h_{jk}\ \mathrm{d}p^j\,\mathrm{d}p^k\,, \tag{34}$$

where we used contravariant elements for the probability increments and the covariant form of the metric. The metric tensor obeys

$$h_{jk}h^{kl} = \delta_{jl}\,, \tag{35}$$

where we use Einstein's summation convention and where δ_{jl} is the Kronecker delta. We have to derive a natural form for h.

B. The Fisher-Rao metric and statistical distance

For this section, we follow the procedure in Bengtsson and Życzkowski[23] to formulate an appropriate metric on the space of probability distributions. A similar procedure is taken by Kok and Lovett[24]. A natural scalar product on the simplex and its dual space of classical random variables forms an expectation value:

$$\langle A\rangle = \sum_j A_j p^j\,. \tag{36}$$

The correlation between two classical random variables A and B is then

$$\langle AB\rangle = \sum_j A_j B_k h^{jk} = \sum_j A_j B_j p^j\,. \tag{37}$$

Using the relation in Eq. (35) we find that

$$h_{jk} = \frac{\delta_{jk}}{p^j}\,, \tag{38}$$

which leads to the so-called *Fisher-Rao metric* (FR)[23]

$$\mathrm{d}s^2_{\mathrm{FR}} = \sum_{jk}\mathrm{d}p^j\,\mathrm{d}p^k h_{jk} = \sum_j \frac{\mathrm{d}p^j\,\mathrm{d}p^j}{p_j}. \tag{39}$$

This defines the statistical distance between two probability distributions in the probability simplex. The generalisation of this metric for continuous probability density functions (PDF) can be written as[23]

$$h_{ab} = \frac{1}{4}\int_\Omega \mathrm{d}\mathbf{x}\frac{\partial_a p(\mathbf{x})\,\partial_b p(\mathbf{x})}{p(\mathbf{x})}, \tag{40}$$

where Ω defines a finite dimensional sub-manifold of the probability simplex with coordinates θ^a, and $\partial_a \equiv \partial/\partial\theta^a$. For simplicity we will mostly use the discrete form in the remainder of this section.

The probability simplex with the FR metric exhibits strong curvature. Note that the statistical distance in Eq. (39) diverges when one of the probabilities p_j tends towards zero. This gives us a clue how to interpret the distance between two distributions: when the probability of one of the measurement outcomes is strictly zero, then obtaining that measurement

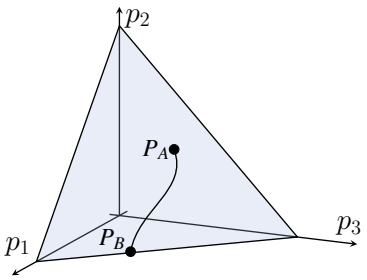


FIG. 4. The distance between probability distributions P_A and P_B diverges when one of them (B) lies on the hull of the simplex.

outcome will allow us to infer *with certainty* that the system is governed by the other probability distribution (see Fig. 4).

Next, we consider the displacement $\mathrm{d}s$ in the probability simplex along a line element $\mathrm{d}\theta$. We can write

$$\frac{\mathrm{d}s^2}{\mathrm{d}\theta^2} = \sum_j \frac{1}{p^j}\frac{(\mathrm{d}p^j)^2}{\mathrm{d}\theta^2} = \sum_j p_j \left(\frac{\partial \ln p^j}{\partial\theta}\right)^2 . \tag{41}$$

Comparing this with Eq. (24) we see that this is the Fisher information

$$I(\theta) = \left(\frac{\mathrm{d}s}{\mathrm{d}\theta}\right)^2 , \tag{42}$$

and in the case of continuous data sets

$$I(\theta) = \int \mathrm{d}\mathbf{x}\; p(\mathbf{x}|\theta)\left(\frac{\partial \ln p(\mathbf{x}|\theta)}{\partial\theta}\right)^2 . \tag{43}$$

Therefore, the Fisher information measures how fast the probability distribution changes along paths parametrised by θ. In order to tell the difference between two values θ and θ' a higher Fisher information will be beneficial. We can think of the Fisher information $I(\theta)$ as the average amount of information about θ in a single measurement.

For small finite distances induced by a shift $\delta\theta$, and starting at θ, we can express the statistical distance as a Taylor expansion

$$s(\theta+\delta\theta) = s(\theta) + \delta\theta \left.\frac{\mathrm{d}s}{\mathrm{d}\theta}\right|_\theta + O(\delta\theta^2) . \tag{44}$$

Defining $\delta s = s(\theta+\delta\theta) - s(\theta)$, we obtain $\delta s = \delta\theta\sqrt{I(\theta)}$ up to first order in $\delta\theta$. We postulate that two probability distributions are distinguishable after N measurements on independent identically prepared systems if the resulting distance crosses some threshold α:

$$N\delta s^2 \geq \alpha , \tag{45}$$

where usually we set $\alpha = 1$. We can eliminate δs from Eq. (45) to obtain

$$(\delta\theta)^2 \geq \frac{1}{NI(\theta)} , \tag{46}$$

which is strongly reminiscent of the Cramér-Rao bound. The difference is that $\delta\theta$ here is the segment of the path in the probability simplex, rather than the variance in the estimator of θ. Nevertheless, Eq. (45) is a powerful method for working out the number of measurements that are required to see a difference $\delta\theta$ in the data.

Eq. (39) can also be re–expressed in a more convenient way if we introduce a new coordinate system $(x^j)^2 = p^j$. This transforms the FR metric to the Euclidean metric:

$$\mathrm{d}s^2 = 4\sum_j \mathrm{d}x^j\,\mathrm{d}x^j . \tag{47}$$

The factor 4 can be absorbed in a change of units, but we choose to keep it here since it will reappear in the quantum extension later on. While in classical estimation theory the occurrence of (real) probability amplitudes $x_i = \sqrt{p_i}$ is something of a curiosity, in the quantum extension to estimation theory this allows us to construct the Fubini-Study metric[25].

C. Relative entropy

In addition to distance measures, probability distributions are conveniently characterised by entropic functions. The Shannon entropy of a random variable P[26],

$$S(P) = -\sum_{i=1}^{N} p_i \log_2 p_i , \tag{48}$$

measures the average amount of information in an event sampled from a system described by the probability distribution p_i. The units are bits, and in the remainder of this review, logarithms are base 2, unless stated otherwise.

To compare two probability distributions, we can make use of the Kullback-Leibler, or *relative* entropy[27,28]

$$D(P\,\|\,Q) = \sum_{i=1}^{N} p_i \log\frac{p_i}{q_i} . \tag{49}$$

While this is not a metric since it is not symmetric under exchange of p_i and q_i, it remains sufficient as a distinguishability measure between the two distributions. Eq. (49) describes the information gain when a prior distribution Q is updated to the posterior distribution P.

A Taylor expansion to first order of the logarithmic term of the relative entropy in Eq. (49) yields

$$D(P\,\|\,P+\mathrm{d}P) = \sum_{i=1}^{N} p^i \log\frac{p^i}{p^i+\mathrm{d}p^i} \approx \sum_i \frac{\mathrm{d}p^i\,\mathrm{d}p^i}{2p^i} . \tag{50}$$

The last term is identical to the statistical distance in Eq. (39) up to a factor two. The Fisher information must therefore be closely related to the relative entropy D in Eq. (49). Assuming that the two distributions in the relative entropy are connected by a curve θ, we may label the distributions by $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. We

find that the Fisher information matrix can be approximated by the second derivative of the relative entropy

$$[I(\boldsymbol{\theta})]_{ij} = \left(\frac{\partial^2}{\partial\theta'_i\partial\theta'_j} D(\boldsymbol{\theta}\,\|\,\boldsymbol{\theta}')\right)_{\theta'=\theta} = -\int \mathrm{d}\mathbf{x}\; p(\mathbf{x},\boldsymbol{\theta})\frac{\partial^2 \log p(\mathbf{x},\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}. \tag{51}$$

Conversely, the relative entropy can be written in terms if the Fisher information matrix

$$D(\boldsymbol{\theta}\,\|\,\boldsymbol{\theta}') = \frac{1}{2}(\boldsymbol{\theta}'-\boldsymbol{\theta})^\top I(\boldsymbol{\theta})(\boldsymbol{\theta}'-\boldsymbol{\theta}), \tag{52}$$

up to higher order corrections in $(\boldsymbol{\theta}'-\boldsymbol{\theta})$. The relative entropy has a number of advantages over the Fisher information in that it is not affected by changes in parameterisation, it can be used even if the distributions are not all members of a parametric family, and fewer smoothness conditions on the probability densities are needed.

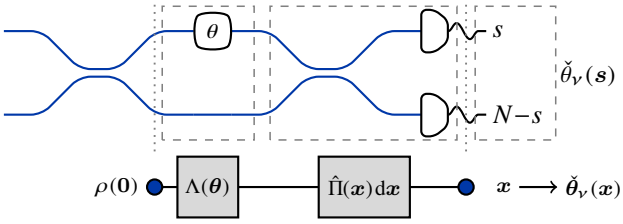


FIG. 5. General channel parameter estimation scheme with the possibility of adaptive control, shown in parallel with the phase estimation procedure using a Mach-Zehnder interferometer. The quantum channel $\Lambda(\theta)$ imparts parameters $\boldsymbol{\theta}$ on the input probe state. The probe is measured by an operator of the form $\Pi(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and distributes estimates according to Born's rule. By data processing the measurement outcomes, we obtain our estimate $\check{\theta}$. The goal of quantum metrology is in finding both the optimal probe state and observable that minimises the covariance matrix of unbiased estimates. Figure adapted with permission from Sidhu, J., Quantum metrology of grid deformations and squeezed light: with applications in quantum imaging & quantum information, Doctoral Thesis, University of Sheffield, Copyright 2018[11].

IV. SINGLE PARAMETER QUANTUM ESTIMATION

In this section we review single-parameter quantum estimation theory and derive the quantum Cramér-Rao bound. We show how the quantum fisher information can be obtained as a limiting case of the classical Fisher information, and we provide an interpretation for the general parameter estimation scheme illustrated in Fig. 5. We give various closed forms for the quantum Fisher information.

A. Quantum model of precision measurements

Any estimation strategy is described through a probe preparation stage with state $\rho(0)$, followed by an evolution that imprints the parameters of interest through a quantum channel $\rho(\theta) = \Lambda[\rho(0)](\theta)$, and a measurement stage by a self-adjoint observable $X = \int \mathrm{d}\mathbf{x}\,\mathbf{x}\,\Pi(\mathbf{x})$ where $\Pi(\mathbf{x})\,\mathrm{d}\mathbf{x}$ is a positive operator valued measure (POVM) element. This archetypal schema is illustrated in Fig. 5 (in principle, a feedback mechanism can be included). For any given interaction, this protocol describes a two-step optimisation problem; an experimenter must make a suitable choice of probe state that is sensitive to changes in the parameters to assimilate maximal information, and they must make an appropriate measurement that maximises the information extracted from the probe. Analytically, this can be modelled by describing the evolved state of the system through $\rho(\theta)$, and by associating the measurement yielding the data $\mathbf{x}$ through the POVM $\Pi(\mathbf{x})\,\mathrm{d}\mathbf{x}$. The probability distribution $p(\mathbf{x}|\theta)$ is then given by Born's rule

$$p(\mathbf{x}|\theta)\,\mathrm{d}\mathbf{x} = \mathrm{Tr}[\Pi(\mathbf{x})\rho(\theta)]\,\mathrm{d}\mathbf{x}, \tag{53}$$

where $\int \mathrm{d}\mathbf{x}\,\Pi(\mathbf{x}) = \mathbb{I}$. Born's rule gives the probability distribution function (PDF) that distributes the measurement outcomes $\mathbf{x}$, given the parameterisation θ. The state captures uncertainties associated with the state-preparation procedure, while the POVM captures those associated with the measurement stage. Together with Born's rule, they model the probabilistic nature of the measurement data.

The bounds in the previous section were derived for probability distributions $p(\mathbf{x}|\theta)$. This immediately generalises to the case where the probability distribution results from some quantum mechanical process described above through Born's rule. A natural question is what is the best possible precision in θ that can be obtained from (many copies of) $\rho(\theta)$? By optimising over all measurement strategies, the result to this question is the quantum Cramér-Rao bound (QCRB), which lower bounds the variance of any unbiased estimator that maps measured data from quantum measurements to parameter estimations[29]. It is of fundamental interest since it can be regarded as an intrinsic property of the system, and is determined entirely by the quantum Fisher information (QFI), which depends only on the state $\rho(\theta)$. In the following section, we define the quantum mechanical version of the Fisher information. We can derive this via geometrical arguments on the probability space similar to section III.

B. The quantum Fisher information

We will find an expression for the QFI inspired by the above derivation of the classical Fisher information[22,30]. Notable contributions to this extension were made by Wootters[21], Hilgevoord and Uffink[31], and Braunstein and Caves[22]. We restrict ourselves again to the case of a single parameter θ. In the quantum mechanical case the expectation value of an operator is given by the Born rule

$$\langle A\rangle = \mathrm{Tr}[\rho A], \tag{54}$$

where ρ is the quantum state of the system and A is a self-adjoint operator. In order to find the single parameter QFI we again define a metric via the correlation between two observables A and B. There is, however, a complication. Since observables in quantum mechanics generally do not commute, the product AB is often not self-adjoint and cannot be considered an observable: $(AB)^\dagger = B^\dagger A^\dagger = BA \neq AB$. A natural remedy for this problem is to use the anti-commutator $\{A, B\} = AB + BA$ as the observable for the correlation:

$$\begin{aligned}\left\langle \tfrac{1}{2}\{A, B\}\right\rangle &= \frac{1}{2}\mathrm{Tr}[\rho\{A, B\}] = \frac{1}{2}\mathrm{Tr}[A\{\rho, B\}] \\ &\equiv \mathrm{Tr}\left[A\,\mathscr{R}_\rho(B)\right], \end{aligned} \tag{55}$$

where we included a factor $\frac{1}{2}$ for normalisation, and we defined the super-operator

$$\mathscr{R}_\rho(B) \equiv \frac{1}{2}\{\rho, B\}, \tag{56}$$

which plays the role of the metric with raised indices. We can find the statistical distance by constructing the lowering operator $\mathscr{L}_\rho = \mathscr{R}_\rho^{-1}$ such that $\mathscr{L}_\rho(\mathscr{R}_\rho(B)) = B$. Explicit verification shows that $\mathscr{L}_\rho$ satisfies

$$\mathscr{L}_\rho(B) = 2\sum_{jk}\frac{B_{jk}}{p_j + p_k}\,|j\rangle\langle k|\,, \tag{57}$$

where $\{|j\rangle\}$ is the eigenbasis of ρ, with eigenvalues p_j, and B_{jk} is the matrix element of B corresponding to j and k. It is similarly easy to show that $\mathscr{R}_\rho(\mathscr{L}_\rho(B)) = B$, which proves that $\mathscr{R}_\rho$ and $\mathscr{L}_\rho$ are each others' inverse operations. Note that in this form the set of pure states is excluded from the space of density matrices. However, we will show in section IV D that we can also include pure states.

The lowering operator allows us to construct a scalar product between density operators $\mathrm{Tr}[\sigma_1\mathscr{L}_\rho(\sigma_2)]$. In particular, for small changes in the density operator $\mathrm{d}\rho$ we can construct the infinitesimal quantum statistical distance $\mathrm{d}s_Q$

$$\mathrm{d}s_Q^2 = \mathrm{Tr}\left[\mathrm{d}\rho\,\mathscr{L}_\rho(\mathrm{d}\rho)\right]. \tag{58}$$

The QFI is then given by the change of the quantum statistical distance along the curve θ:

$$I_Q(\theta) \equiv \left(\frac{\mathrm{d}s_Q}{\mathrm{d}\theta}\right)^2 = \mathrm{Tr}\left[\partial_\theta\rho\,\mathscr{L}_\rho(\partial_\theta\rho)\right], \tag{59}$$

where again $\partial_\theta \equiv \partial/\partial\theta$.

The QFI in Eq. (59) is a Riemannian metric on the quantum state space. Nagaoka[32] and Braunstein and Caves[22] show that this quantum Fisher information can be attained by a judicially chosen measurement. In other words, consider a measurement M expressed in terms of its POVM elements:

$$M = \int \mathrm{d}\mathbf{x}\; m(\mathbf{x})\;\Pi(\mathbf{x}), \tag{60}$$

where $m(\mathbf{x})$ is the real eigenspectrum of M. For the optimal choice of M the QFI in Eq. (59) coincides with the Fisher information in Eq (42), where $p(\mathbf{x}|\theta) = \mathrm{Tr}[\Pi(\mathbf{x})\rho(\theta)]$ is the probability distribution over the measurement outcomes of M. The two necessary and sufficient conditions for M to be optimal are that for all $\mathbf{x}$

$$\begin{aligned} &(1) \qquad \mathrm{Im}\,\mathrm{Tr}\left[\rho\Pi(\mathbf{x})\mathscr{L}_\rho(\partial_\theta\rho)\right] = 0\,, \\ &(2) \qquad \frac{\sqrt{\Pi(\mathbf{x})}\sqrt{\rho}}{\mathrm{Tr}[\Pi(\mathbf{x})\rho]} = \frac{\sqrt{\Pi(\mathbf{x})}\,\mathscr{L}_\rho(\partial_\theta\rho)\,\sqrt{\rho}}{\mathrm{Tr}[\rho\Pi(\mathbf{x})\mathscr{L}_\rho(\partial_\theta\rho)]}\,, \end{aligned} \tag{61}$$

where Im denotes the imaginary part.

The QFI has a number of interesting properties[33,34]. First, it is convex in the quantum states This means that for any two states ρ_1 and ρ_2 we have

$$I_Q(p_1\rho_1 + p_2\rho_2) \leq p_1 I_Q(\rho_1) + p_2 I_Q(\rho_2), \tag{62}$$

with probabilities p_1 and p_2 such that $p_1 + p_2 = 1$. Moreover, the QFI is additive for independent measurements

$$I_Q(\rho_1 \otimes \rho_2, \theta) = I_Q(\rho_1, \theta) + I_Q(\rho_2, \theta), \tag{63}$$

as well as for direct sums:

$$I_Q(p_1\rho_1 \oplus p_2\rho_2, \theta) = p_1 I_Q(\rho_1, \theta) + p_2 I_Q(\rho_2, \theta)\,. \tag{64}$$

Second, for unitary evolutions $U = \exp(-i\theta G)$ generated by a Hermitian operator G, the QFI does not depend on the position along the orbit of U:

$$I_Q(U\rho U^\dagger, G) = I_Q(\rho, U^\dagger G U) = I_Q(\rho, G)\,. \tag{65}$$

It does not increase under CPTP maps $\mathscr{E}[\rho]$ that do not depend on the parameter of interest:

$$I_Q(\rho, \theta) \geq I_Q\left(\mathscr{E}[\rho], \theta\right), \tag{66}$$

and tracing out a subsystem cannot increase the QFI:

$$I_Q(\rho, G_1 \otimes \mathbb{I}_2) \geq I_Q(\mathrm{Tr}_2[\rho], G_1)\,. \tag{67}$$

Adding white noise to the state of N particles, each described in a d-dimensional Hilbert space, is equivalent to mixing in the identity matrix, such that

$$\rho = p\,|\psi\rangle\langle\psi| + (1-p)\frac{\mathbb{I}}{d^N}\,. \tag{68}$$

The QFI then becomes

$$I_Q(\rho, \theta) = \frac{p^2}{p + \frac{1-p}{2}d^{-N}} I_Q(\psi, \theta)\,. \tag{69}$$

Third, we can write the QFI for a unitarily evolved state in terms of the generator G as

$$\begin{aligned} I_Q(\theta) &= 4\sum_{j,k} p_j |\langle j|G|k\rangle|^2 - 8\sum_{j,k}\frac{p_j p_k}{p_j + p_k}|\langle j|G|k\rangle|^2 \qquad (70) \\ &= \langle G^2\rangle - 8\sum_{j,k}\frac{p_j p_k}{p_j + p_k}|\langle j|G|k\rangle|^2, \qquad (71) \end{aligned}$$

where we used the diagonal form of the density operator $\rho = \sum_j p_j |j\rangle\langle j|$. The QFI does not depend on the diagonal elements of G, so we can write

$$I_Q(\rho, G) = I_Q(\rho, G + D), \tag{72}$$

where D is any diagonal matrix. Another way to express the QFI for mixed states is via purifications[35]:

$$I_Q(\rho, \theta) = \min_{\Psi} \left(\langle \dot{\Psi} | \dot{\Psi} \rangle - |\langle \dot{\Psi} | \Psi \rangle|^2 \right), \tag{73}$$

where $|\Psi\rangle$ is a purification of ρ. Another form is due to Fujiwara and Imai[36]:

$$I_Q(\rho, \theta) = \min_{\Psi} \langle \dot{\Psi} | \dot{\Psi} \rangle \ . \tag{74}$$

Eqs. (73) and (74) are equivalent for the purification that achieves the minimum. In other words, $|\langle \dot{\Psi} | \Psi \rangle|^2 = 0$ for the minimal purification state $|\Psi_{\min}\rangle$.

C. Distance measures in quantum estimation

We momentarily return to the role of distance measures in the definition of the QFI, and its relation to other distance functions in classical and quantum parameter estimation.

The classical statistical distance induces a curvature in the probability simplex that can be removed by introducing probability amplitudes, as shown in Eq. (47). Introducing a complex phase in the amplitudes, $x_j \to x_j e^{i\phi_j}$, allows us to relate these amplitudes to normalised vectors $|\psi\rangle$ in a complex Hilbert space $\mathscr{H}$ whose distance to other vectors in $\mathscr{H}$ is then given by the *angle* between them. In infinitesimal form, this gives the Wootters distance[21,37]

$$\mathrm{d}s^2_{\mathrm{W}} = (\arccos|\langle \psi | \psi + \delta\psi \rangle|)^2, \tag{75}$$

and the metric is called the Fubini-Study metric on the projective space $R(\mathscr{H}) = \mathscr{H} \otimes \mathscr{H}$. The pullback metric[38], one which pulls the metric from the codomain of $R(\mathscr{H})$ to the domain of $\mathscr{H}$, for the Fubini-Study metric is given by[25]

$$h_{\mathrm{FS}} = \frac{\langle \partial_\theta \psi | \partial_\theta \psi \rangle}{\langle \psi | \psi \rangle} - \frac{\langle \partial_\theta \psi | \psi \rangle \langle \psi | \partial_\theta \psi \rangle}{\langle \psi | \psi \rangle^2}, \tag{76}$$

where $|\partial_\theta \psi\rangle \equiv \partial_\theta |\psi\rangle$. Assuming $\langle \psi | \psi \rangle = 1$ then up to a factor 4 this equals the QFI for pure states[39]

$$I_Q(\theta) = 4h_{\mathrm{FS}} = 4 \left[\langle \partial_\theta \psi | \partial_\theta \psi \rangle - |\langle \partial_\theta \psi | \psi \rangle|^2 \right] . \tag{77}$$

When we extend the Fubini-Study metric to density operators in $R(\mathscr{H})$ we obtain the Bures metric[15,25,40] given in Eq. (58)

$$4\,\mathrm{d}s^2_{\mathrm{Bures}} = \mathrm{d}s^2_Q = \mathrm{Tr}\left[\mathrm{d}\rho \mathscr{L}_\rho(\mathrm{d}\rho) \right] . \tag{78}$$

The origin of the factor 4 is the same as in Eq. (47).

The Wootters distance between two quantum states $|\psi\rangle$ and $|\phi\rangle$ is closely related to the fidelity $F = |\langle \psi | \phi \rangle|^2$ between the states, $\mathrm{d}s^2_{\mathrm{W}} = (\arccos \sqrt{F})^2$. The corresponding fidelity for mixed states is the Uhlmann fidelity[41,42]

$$F(\rho, \sigma) = \left(\mathrm{Tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right)^2, \tag{79}$$

and we can express the quantum statistical distance $\mathrm{d}s^2_Q$ as

$$\mathrm{d}s^2_Q(\rho, \rho + \mathrm{d}\rho) = 8 \left[1 - \sqrt{F(\rho, \rho + \mathrm{d}\rho)} \right] . \tag{80}$$

The QFI can then be written in terms of the quantum fidelity as[22]

$$I_Q(\theta) = \frac{8}{\delta\theta^2} \left[1 - \sqrt{F(\rho, \rho + \delta\rho)} \right], \tag{81}$$

which will allow for analytic expressions in a variety of cases. Sometimes the square root of the Uhlmann fidelity in Eq. (79) is used as an alternative definition. This form is referred to as the quantity fidelity[43] and has been shown to have better fit with other quantities in quantum metrology[44] and quantum hypothesis testing[45,46].

D. The Symmetric Logarithmic Derivative

The classical Fisher information is the expectation value of the squared derivative of the logarithm of the probability distribution, as shown in Eq. (24). In quantum estimation theory, we can define a similar quantity, now an operator L_θ, that is called the *symmetric logarithmic derivative*[47,48] (SLD), which is equal to the lowering operator in Eq. (57) of the derivative of ρ

$$L_\theta = \mathscr{L}_\rho(\partial_\theta \rho) . \tag{82}$$

Moreover, the SLD is implicitly defined by the relation

$$\partial_\theta \rho = \frac{1}{2}(\rho L_\theta + L_\theta \rho) . \tag{83}$$

The symmetric form of this definition is directly related to the symmetrized definition of the correlation between quantum observables $\frac{1}{2}\langle\{A, B\}\rangle$ in Eq. (55) via the identification with $\mathscr{L}_\rho(\partial_\theta \rho)$, which is a metric operator derived directly from the inner product $\frac{1}{2}\langle\{A, B\}\rangle$. Some intuition for the definition of L_θ can be gained from the classical logarithmic derivative $L_{\mathrm{cl}} = \partial_\theta \log p(\mathbf{x}|\theta)$, which gives the relation $\partial_\theta p(\mathbf{x}|\theta) = p(\mathbf{x}|\theta) L_{\mathrm{cl}}$ (Note that L_{cl} is equal to the function $g(\mathbf{x}, \theta)$ in Eq. (22)). Replacing the classical probability distribution $p(\mathbf{x}|\theta)$ with a density operator introduces an ambiguity in the operator order of ρ and L_θ, which is resolved by taking the anti-commutator in Eq. (83).

To prove relation Eq. (82), we write L_θ in the eigenbasis of ρ:

$$L_\theta = \sum_{jk} L_{jk} |j\rangle\langle k| \quad \text{and} \quad \rho = \sum_l p_l |l\rangle\langle l| , \tag{84}$$

and construct the operator form of L_θ

$$\frac{1}{2}\left(L_\theta\rho+\rho L_\theta\right)=\frac{1}{2}\sum_{jk}\left(p_j+p_k\right)L_{jk}\left|j\right\rangle\left\langle k\right| = \sum_{jk}(\partial_\theta\rho)_{jk}\left|j\right\rangle\left\langle k\right| . \tag{85}$$

Each matrix element of $\partial_\theta\rho$ must match that of $\frac{1}{2}(L_\theta\rho+\rho L_\theta)$, and therefore we have

$$L_{jk}=\frac{2(\partial_\theta\rho)_{jk}}{p_j+p_k} . \tag{86}$$

Substituting this back into L_θ we see that the SLD in Eq. (86) takes the same form as $\mathscr{L}_\rho(\partial_\theta\rho)$ in Eq. (57), and the identity in Eq. (82) is proved. This identification unifies the geometric interpretation of the QFI with its interpretation as a limiting case of the classical Fisher information in the next subsection.

The QFI can now be written in terms of the SLD L_θ by noting that

$$\begin{aligned} I_Q(\theta) &= \mathrm{Tr}\left[(\partial_\theta\rho)\mathscr{L}_\rho(\partial_\theta\rho)\right] = \mathrm{Tr}[(\partial_\theta\rho)L_\theta] \\ &= \frac{1}{2}\mathrm{Tr}[\left(L_\theta\rho+\rho L_\theta\right)L_\theta] \\ &= \mathrm{Tr}\left[\rho L_\theta^2\right] . \end{aligned} \tag{87}$$

In the eigenbasis of the density operator ρ and using Eq. (59), the QFI can then be written as

$$I_Q(\theta)=2\sum_{jk}\frac{|\langle j|(\partial_\theta\rho)|k\rangle|^2}{p_j+p_k} . \tag{88}$$

The sum extends over all j with nonzero p_j. For vanishing probabilities p_j, the QFI becomes ill-defined and we have to find an alternative way to define it[49,50], such as through Eq. (77). Alternatively, we review a regularisation procedure in section VI which can be used to determine the QFI for pure states from expressions valid for mixed states. From Eq. (88), we observe that the QFI is dependent on the quantum state and its derivative only, and not on the measurement that is performed. In this sense, the QFI is a property of the state.

We also note that the QFI is the expectation value of the square of the SLD (i.e., its second moment), which prompts us to ask what is the first moment of L_θ. Using the fact that $\mathrm{Tr}[\partial_\theta\rho]=0$, it is straightforward to show that $\mathrm{Tr}[\rho L_\theta]=0$. This leads us to the important relations

$$\langle L_\theta\rangle=0 \qquad \text{and} \qquad I_Q(\theta)=\langle L_\theta^2\rangle=(\Delta L_\theta)^2 , \tag{89}$$

where $(\Delta A)^2\equiv\langle A^2\rangle-\langle A\rangle^2$ is the variance of an operator A.

Next, we will show that the SLD form of the QFI in Eq. (87) originates from the maximisation over all possible measurements in an estimation procedure[22,32]. This connects the geometric interpretation of the QFI as the Bures metric in the space of density operators to the statistical interpretation of the QFI as the maximum amount of information about θ that can be extracted on average in an optimal measurement. Using the Born rule and the fact that the SLD is traceless, the classical Fisher information can be written as[51]

$$I(\theta)=\int \mathrm{d}\mathbf{x}\,\frac{\left(\mathrm{Re}\left\{\mathrm{Tr}[\rho(\theta)\Pi(\mathbf{x})L_\theta]\right\}\right)^2}{\mathrm{Tr}[\Pi(\mathbf{x})\rho(\theta)]}, \tag{90}$$

Next, we maximise this quantity over all possible POVMs $\Pi(\mathbf{x})$. Given the complex vectors $\alpha,\beta\in\mathbb{C}^2$, and $\mathrm{Re}[\alpha]\mathrm{Re}[\beta]\leq|\alpha\beta|$, we develop Eq. (90) into

$$\begin{aligned} I(\theta) &\leq \int \mathrm{d}\mathbf{x}\left|\frac{\mathrm{Tr}[\rho(\theta)\Pi(\mathbf{x})L_\theta]}{\sqrt{\mathrm{Tr}[\Pi(\mathbf{x})\rho(\theta)]}}\right|^2 \\ &= \int \mathrm{d}\mathbf{x}\left|\mathrm{Tr}\left[\frac{\sqrt{\rho(\theta)}\sqrt{\Pi(\mathbf{x})}}{\sqrt{\mathrm{Tr}[\Pi(\mathbf{x})\rho(\theta)]}}\sqrt{\Pi(\mathbf{x})}L_\theta\sqrt{\rho(\theta)}\right]\right|^2 , \end{aligned} \tag{91}$$

where equality holds if and only if $\mathrm{Im}[\mathrm{Tr}[\rho(\theta)\Pi(\mathbf{x})L_\theta]]=0$, i.e. if the vectors lie in the real space $\mathbb{R}^2$. This requires the SLD to be Hermitian. Introducing $(\Delta L_\theta)^2$ as the variance of the SLD, we use the Schwartz inequality for the trace

$$\left|\mathrm{Tr}\left[AB^\dagger\right]\right|^2\leq\mathrm{Tr}\left[AA^\dagger\right]\mathrm{Tr}\left[BB^\dagger\right], \tag{92}$$

to obtain:

$$\begin{aligned} I(\theta) &\leq \int \mathrm{d}\mathbf{x}\,\mathrm{Tr}\left[\sqrt{\rho(\theta)}L_\theta\sqrt{\Pi(\mathbf{x})}\sqrt{\Pi(\mathbf{x})}L_\theta\sqrt{\rho(\theta)}\right], \\ &= \int \mathrm{d}\mathbf{x}\,\mathrm{Tr}[\rho(\theta)L_\theta\Pi(\mathbf{x})L_\theta], \\ &= \mathrm{Tr}\left[\rho(\theta)L_\theta^2\right], \end{aligned} \tag{93}$$

where the final equality arises since integrating the POVM gives the identity. This completes the maximisation of the CFIM over all possible measurements. It shows that the classical Fisher information for any measurement is upper bounded by the quantum Fisher information (QFI)

$$I(\theta)\leq I_Q(\theta)=\mathrm{Tr}\left[\rho(\theta)L_\theta^2\right]=\mathrm{Tr}[\partial_\theta\rho(\theta)L_\theta]. \tag{94}$$

Note that the QFI is independent of the POVM and is a function of the state only.

As we observe from Eq. (84), the SLD for mixed states often requires diagonalising the density matrix. For arbitrarily large s-dimensional states, this becomes increasingly difficult. To address this difficulty, alternative methods at determining L_θ have been developed. For example, it has been shown that evaluating L_θ is isomorphic to solving a set of linear algebraic equations[52].

The implicit definition of the SLD in Eq. (83) is a basis-independent Lyapunov matrix equation that has the general solution[51]

$$L_\theta=2\int_0^\infty \mathrm{d}s\,\exp\left[-\rho(\theta)s\right]\partial_\theta\rho(\theta)\exp\left[-\rho(\theta)s\right] . \tag{95}$$

When ρ is not full rank we can still define L_θ in this way, but some care needs to be taken in order to show that the QFI is still well-defined[53]. The SLD is not uniquely defined when the state is not full-rank, since the part of the operator acting

on the null-space of ρ is not specified. The QFI based on this expression for the SLD can then be written as

$$I_Q(\theta) = 2\int_0^\infty \mathrm{d}s\ \mathrm{Tr}[(\partial_\theta\rho)e^{-\rho s}(\partial_\theta\rho)e^{-\rho s}]\,. \tag{96}$$

The Lyapunov representation proves to be very useful for scenarios in which a periodic nature is observed for the anti-commutator of the density matrix and its partial derivative[53].

When the evolution imparting the parameter θ onto the quantum state $\rho(\theta)$ is a unitary transformation of the form $U = \exp(-i\theta G)$ with G a Hermitian observable and the eigenvalues of ρ independent of θ, we can use the Lyapunov form to find a particularly elegant expression for the QFI.

$$\begin{aligned} I_Q(\theta) &= 2\int_0^\infty \mathrm{d}s\ \mathrm{Tr}[(\partial_\theta\rho)e^{-\rho s}(\partial_\theta\rho)e^{-\rho s}] \\ &= -2\int_0^\infty \mathrm{d}s\ \mathrm{Tr}\Big[([G,\rho]e^{-\rho s})^2\Big]. \end{aligned} \tag{97}$$

Braunstein, Caves and Milburn showed how the QFI can be lower bounded by the generator through[30]

$$I_Q(\theta) \le 4(\Delta G)^2\,. \tag{98}$$

When the probe state is pure ($\rho^2 = \rho$), we obtain

$$I_Q(\theta,\psi) = 4(\Delta G)^2\,, \tag{99}$$

which can often be found analytically.

Finally, we address the issue that the SLD, as defined in Eq. (86) becomes singular for pure states. Nevertheless, there is a simple expression for the QFI for pure states, as shown in Eq. (99). We can also find a simple expression for the SLD when ρ is pure. Using $\rho^2 = \rho$, differentiating with respect to the parameters θ and comparing with the definition Eq. (83) we arrive at

$$L_\theta = 2\partial_\theta\rho = 2i[\rho, G]\,, \tag{100}$$

where we relate the SLD to the von Neumann equation of motion describing the dynamics of the system.

The calculation of the QFI for any physical system is at the heart of quantum metrology and is typically a difficult task. Determining the QFI using the SLD operator is particularly suited to unitary quantum metrology. It is less suited for noisy processes, where the calculation involves complex optimisation procedures[35,54]. To address this, an extended Hilbert space approach may be taken where information about the parameter is obtained by observing both the system and its environment[55]. This method prescribes the QFI in terms of the state evolving Hamiltonian, and is well suited to many physical implementations of parameter estimations, including open quantum systems[56–59].

E. The quantum Cramér-Rao bound

In this section we will derive QCRB (or the *Helstrom* bound) for a single parameter θ. We follow Helstrom's original derivation[47] that employs the SLD. First, we consider a measurement M on a system in state ρ that serves as an estimator for θ. In other words, the expectation value $\mathrm{Tr}[\rho M]$ provides an estimate for the particular measurement result with a possible bias $b(\theta) = \mathrm{Tr}[\rho(M-\theta)]$. We can take the derivative of the bias with respect to θ to obtain

$$\partial_\theta b(\theta) = \mathrm{Tr}[(M-\theta)\partial_\theta\rho] - 1\,, \tag{101}$$

where we assume that M does not itself depend on θ, since the measurement should work for a sufficiently wide range of parameters around θ. This is an important caveat that we will return to when we wish to determine the optimal estimator in section V C. We next square Eq. (101), and together with the SLD and the Schwarz inequality for the trace we get

$$\begin{aligned} (1+\partial_\theta b)^2 &= \mathrm{Tr}[(M-\theta)\partial_\theta\rho]^2 \\ &= \frac{1}{4}\mathrm{Tr}[(M-\theta)\,(L_\theta\rho+\rho L_\theta)]^2 \\ &= (\mathrm{Re}\,\mathrm{Tr}[(M-\theta)\rho L_\theta])^2 \\ &\le |\mathrm{Tr}[(M-\theta)L_\theta\rho]|^2 = \left|\mathrm{Tr}\left[\sqrt{\rho}(M-\theta)L_\theta\sqrt{\rho}\right]\right|^2 \\ &\le \mathrm{Tr}\left[\rho L_\theta^2\right]\mathrm{Tr}\left[\rho(M-\theta)^2\right]. \end{aligned} \tag{102}$$

Identifying $\mathrm{Tr}[\rho(M-\theta)^2]$ with the MSE in θ, we arrive at the quantum Cramér-Rao bound (QCRB)

$$\begin{aligned} \mathrm{Var}\,\theta &= \mathbb{E}_\rho\left[(M-\theta)^2\right] \\ &\ge \frac{(1+\partial_\theta b)^2}{\mathrm{Tr}[\rho L_\theta^2]} = (1+\partial_\theta b)^2 I_Q^{-1}(\theta), \end{aligned} \tag{103}$$

where we have identified $\mathrm{Tr}[\rho L_\theta^2]$ with the QFI $I_Q(\theta)$. Note that for types of bias with negative derivatives the MSE appears to be better than in the case of an unbiased estimator ($b = 0$). This can cause some confusion when comparing the MSE with a pre-calculated value of the QFI. When N independent measurements are made using an unbiased estimator, additivity of the QFI implies that the resulting bound is given by

$$\mathrm{Var}\,\theta \ge \frac{1}{N I_Q}\,. \tag{104}$$

This is the form of the CRB for a single parameter that is mostly used.

The QCRB, together with the bound in Eq. (98) on the QFI, leads immediately to a familiar result. For a single shot measurement ($N = 1$) Eq. (104) becomes

$$\mathrm{Var}\,\theta \ge \frac{1}{4(\Delta G)^2}\,. \tag{105}$$

When we define $\mathrm{Var}\,\theta = (\delta\theta)^2$, this leads to[30]

$$\delta\theta\,\Delta G \ge \frac{1}{2}\,, \tag{106}$$

and can be interpreted as an uncertainty relation for a quantity θ and its generator of translations G. While Heisenberg's uncertainty relations are typically derived for conjugate observables like position and momentum, Eq. (106) allows us

to define uncertainty relations between energy and time, or angular momentum and rotation angles where Robertson inequalities cannot be constructed due to a lack of self-adjoint operators for time and rotation angles.

The next question to address is how to find the optimal estimator that saturates the QCRB. First, does there exist a measurement for which the QFI equals the classical Fisher information? And if so, what is this measurement? To answer these questions, we recall the error propagation formula from Eq. (1):

$$\mathrm{Var}\,\theta = \frac{(\Delta M)^2}{|\mathrm{d}\langle M\rangle/\mathrm{d}\theta|^2}\,. \tag{107}$$

This formula relates the variance in the parameter θ to the variance of the operator M that is used to estimate θ. Helstrom[48] states that the QCRB is saturated if and only if $L_\theta = k(\theta)(M_{\rm opt} - \theta)$, with $M_{\rm opt}$ the optimal measurement strategy that saturates the QCRB and $k(\theta)$ some function that does not depend on $M_{\rm opt}$. Choosing

$$M_{\rm opt} = \theta\,\mathbb{I} + \frac{L_\theta}{I_Q(\theta)}\,, \tag{108}$$

with $k(\theta) = I_Q(\theta)$, we can prove that $\mathrm{Var}\,\theta$ saturates the QCRB. Notice that although any general measurement will not depend on the true parameter, the optimal measurement may. For the optimal measurement in Eq. (108) the bias vanishes: $b(\theta) = \mathrm{Tr}[\rho(M_{\rm opt} - \theta)] = 0$. From Eq. (89) we calculate that

$$\langle M_{\rm opt}\rangle = \theta \qquad \text{and} \qquad \frac{\mathrm{d}\langle M_{\rm opt}\rangle}{\mathrm{d}\theta} = 1\,, \tag{109}$$

and

$$\langle M_{\rm opt}^2\rangle = \left\langle \theta^2 + 2\theta\frac{L_\theta}{I_Q} + \frac{L_\theta^2}{I_Q^2}\right\rangle = \theta^2 + \frac{1}{I_Q}\,. \tag{110}$$

From this, we find that $\mathrm{Var}\,\theta = I_Q^{-1}$, which saturates Eq. (104). In addition, for the optimal measurement to saturate the QCRB, the right-hand side of Eq. (101) must be zero. Substituting Eq. (109) into Eq. (101) shows that this is indeed the case. The dependence of $M_{\rm opt}$ on θ ensures that when θ changes, the expectation value of the estimate changes accordingly.

So there indeed does exist an estimator that saturates the QCRB, but it generally depends on the unknown parameter θ.

It was shown by Braunstein and Caves[22] that the optimal measurement for θ is a von Neumann measurement that consists of projections onto the eigenstates of the SLD L_θ. However, L_θ generally depends on the unknown value of θ, and it may not be possible to choose the optimal estimator at the outset[60]. Since the QCRB, like the classical CRB, is an asymptotic bound on the variance of θ, many measurements must be made before the bound is saturated, and this allows for adaptive measurements that converge to the optimal measurement[1,32,61].

F. The role of entanglement

Consider an experiment that estimates a parameter θ. The experiment is repeated N times under identical conditions, and each time the average information that is extracted about θ is given by the QFI. Since the Fisher information is additive for independent measurements, the total information in the N experiments is NI_Q, leading to the CRB in Eq. (104). The Root Mean Square Error (RMSE) then behaves as

$$\delta\theta \geq \frac{1}{\sqrt{NI_Q}}\,. \tag{111}$$

The square-root scaling $1/\sqrt{N}$ of $\delta\theta$ is called the Standard Quantum Limit (SQL), or shot-noise limit. This is the best possible performance for a classical estimation procedure, i.e., estimation procedures that do not employ entangled probe states.

To see how we can improve over Eq. (111), we consider the case where the parameter is imparted on the quantum state via the unitary evolution $\exp(-i\theta G)$, such that the QFI takes the form

$$I_Q^{(\max)}(\theta) = 4(\Delta G)^2\,. \tag{112}$$

To maximise the QFI is therefore to maximise $(\Delta G_N)^2$ over the state of the N physical systems. Clearly, when each experiment is independent, the variances $(\Delta G_N)^2$ add, such that $(\Delta G_N)^2 = N(\Delta G)^2$ and we recover the SQL in Eq. (111). However, we can also prepare the N systems in a suitable entangled state. This will allow us to increase $(\Delta G_N)^2$ substantially, scaling instead with N^2. The QCRB then becomes

$$\mathrm{Var}\,\theta \geq \frac{\alpha}{N^2}\,, \tag{113}$$

where α is some constant, typically of order unity. This leads to an RMSE that scales with N^{-1}:

$$\delta\theta \sim \frac{1}{N}\,. \tag{114}$$

This is called the *Heisenberg limit*, and it is the ultimate limit for quantum parameter estimation[62].

To see how such a precision can be achieved, we consider the optical NOON state[6,7,63–65]

$$|\psi\rangle = \frac{|N,0\rangle + |0,N\rangle}{\sqrt{2}}\,, \tag{115}$$

where $|N\rangle$ is the N-photon Fock state. This is a two-mode entangled state that is extremely challenging to make in the lab[66,67], but it serves as a clear proof of principle. Assuming a simple phase shift θ in the second mode, the NOON state evolves to

$$|\psi(\theta)\rangle = \frac{|N,0\rangle + \mathrm{e}^{iN\theta}\,|0,N\rangle}{\sqrt{2}}\,, \tag{116}$$

where each photon in the second mode picks up a phase $\mathrm{e}^{iN\theta}$. We calculate the QFI using the expression in Eq. (77)

$$I_Q(\theta) = 4\langle\partial_\theta\psi(\theta)|\partial_\theta\psi(\theta)\rangle - 4\,|\langle\partial_\theta\psi(\theta)|\psi(\theta)\rangle|^2\,. \tag{117}$$

The derivative of the state is given by

$$|\partial_\theta\psi(\theta)\rangle = -\frac{iN\mathrm{e}^{iN\theta}}{\sqrt{2}}\,|0,N\rangle\,, \tag{118}$$

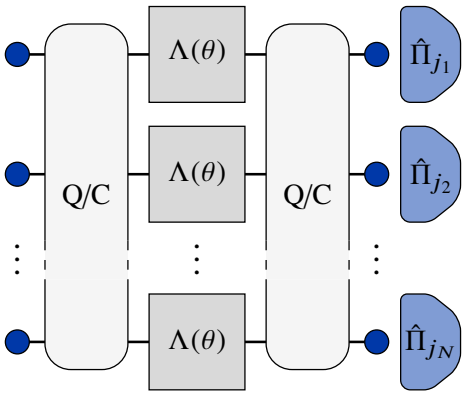


FIG. 6. General parallel parameter estimation strategies. The initial probe states, $\rho(0)$ are represented by the circles on the left. The probe states may be entangled before passing through an evolving channel, illustrated by the grey central squares. The state passes through another potentially entangling evolution and is finally measured using local POVMs Π_j, represented by the blue polygons on the right. This setup allows for four classes of estimation, where the input states many be entangled or not (Q/C on the left), and where the input state may be subjected to entangled or separable measurement observables (Q/C on the right). Figure adapted with permission from Sidhu, J., Quantum metrology of grid deformations and squeezed light: with applications in quantum imaging & quantum information, Doctoral Thesis, University of Sheffield, Copyright 2018[11].

and the QFI becomes $I_Q = N^2$, as required. To complete the estimation procedure, practical measurement observables were proposed by Pryde *et al.*[68] and Cable *et al.*[69]. A similar argument can be constructed using Greenberger-Horne-Zeilinger (GHZ) states[70].

To understand more generally how quantum entanglement can improve the estimation precision, Giovannetti, Lloyd and Maccone classified various metrology approaches[62]. They unified parallel and adaptive sequential strategies into a general framework shown in Fig. 6. The unitary evolutions U imparting the parameter are again of the form $U = \exp(-i\theta G)$. This classification allows for classical or quantum state preparation and measurement procedures, resulting in four different classes of experiments: classical states and classical measurements (CC), classical states and quantum measurements (CQ), quantum states and classical measurements (QC), and both quantum states and measurements (QQ). Here, we understand by "classical" that the input state is separable, and the measurement projects onto separable POVM elements. Since the QCRB is determined by the quantum state via the QFI, no quantum entangling strategies at the measurement stage can introduce further enhancements to the estimation procedure than what is already present in the quantum state. Therefore the CC and CQ strategies will always yield at best the SQL[71]. Any resolution enhancements must then be sourced from the probe preparation. Bound entanglement can also surpass the shot noise limit[72,73].

Toth[74] showed that for qubits, genuine multi-partite entanglement is required to achieve the Heisenberg limit. Considering three possible parameters θ_x, θ_y and θ_z generated by the Pauli operators σ_x, σ_y, and σ_z, the following general results hold:

1. for N-qubit separable states the QFI is bounded by $\sum_j I_Q(\theta_j) \leq 2N$ and for a single parameter $I_Q(\theta_j) \leq N$;
2. for general N-qubit quantum states $\sum_j I_Q(\theta_j) \leq N(N+2)$,
3. for k-producible states, where a pure state is k-producible if it is a tensor product state of at most k qubits, $I_Q(\theta_j) \leq nk^2 + (N-nk)^2$, where n is the integer part of N/k.
4. the sum of the QFI's for each parameter is bounded by

$$\sum_j I_Q(\theta_j) \leq nk(k+2) + (N-nk)(N-nk+2) \quad (119)$$

if $N - nk \neq 1$, and

$$\sum_j I_Q(\theta_j) \leq nk(k+2) + 2 \quad (120)$$

if $N - nk = 1$;
5. for multi-partite quantum states with M unentangled particles $\sum_j I_Q(\theta_j) \leq M + (N-M)(N-M+2)$.

The broader question of how useful quantum states are for quantum metrology was answered by Oszmaniec *et al.*[75], who showed that pure states chosen randomly from the symmetric subspace typically achieve the optimal Heisenberg scaling without the need for local unitary optimisation. Further, an explicit non-random choice of symmetric probe states with error correction capabilities has recently been demonstrated useful for robust metrology[76]. In this work, it was shown that if the probe state lies within the code space of certain permutation-invariant quantum codes[77], a precision enhancement is possible even in the presence of noise. Other symmetric states have been considered for robust quantum metrology[78]. These studies reflect the current area of intense research on error correction inspired robust metrology, where quantum error correction is not applied. This removes the requirement for feed-forward and error-correction, which reduces the difficulty and practicality of implementing practical quantum metrology. These methods are motivated by the near term emergence of noisy quantum devices; the so-called noisy intermediate-scale quantum (NISQ) era. We review fault-tolerant quantum metrology methods later in section VII.

The entanglement requirements above can be turned on their head, such that a QFI in excess of a certain value indicates that at least genuine k-partite entanglement must be present in the quantum state. This is a so-called *entanglement witness*[79–82]. In other words, the difference in precision scaling can be used to deduce whether entanglement is present in the probe state. Relaxing the Heisenberg limit to $\delta\theta \sim 1/N^{1-\epsilon}$ for any $\epsilon > 0$, the amount of entanglement required can be made arbitrarily small[83].

The bounds established by Toth on the QFI holds for qubits undergoing unitary evolutions $\exp(-i\theta_j\sigma_j/2)$. The question remains whether other types of evolution can lead to a different

scaling. Indeed, when the generator of translations in θ is a multi-particle Hamiltonian, the estimation precision in an experiment using N particles can scale with N^{-m} for some $m > 1$, or even 2^{-N}, as shown by Boixo *et al.*[84] and Roy and Braunstein[85], respectively. The key feature of these schemes is that the norm of the Hamiltonian is large. These results cannot be compared directly with the Heisenberg limit, since the evolution in those generalised scaling laws is generated by fundamentally different physical processes than the typical single-particle evolution $\exp(-i\theta G)$.

One way to understand these limits is via the query complexity of the estimation procedure[86]. In the standard parameter estimation procedure each particle evolves according to the effective Hamiltonian G. However, for systems with a bi-partite Hamiltonian $H_{jk} = G_j \otimes G_k$ each evolution requires a *pair* of particles. Each pair is now a *query* of the parameter θ, and for N particles there are $Q = \frac{1}{2}N(N-1)$ queries. This structure generalises to multi-partite Hamiltonians. The precision limit always scales at most linearly with the number of queries, not the number of particles[86–89].

G. Non-entangling strategies

The entanglement strategies in the previous section refer to systems consisting of distinguishable particles. A different situation arises in quantum optics, where at least in principle, non-entangled states can achieve sub-shot noise precision. In particular, the squeezed vacuum can be used to suppress the fluctuations due to shot noise[90,91]. Other states that have been used to attain the Heisenberg limit are the class of entangled coherent states (ECS)[92–94]. For phase estimations in a Mach-Zehnder interferometer, ECS have demonstrated better precision scalings than NOON states[95]. Even in a lossy interferometer, ECS can still beat the shot-noise limit for modest loss rates[96].

Nevertheless, generating highly entangled states is practically difficult for two main reasons. First, the photonic overhead increases exponentially with the number of entangled modes and second, the fidelity decreases due to decoherence processes[97]. Furthermore, despite the results in the previous section the use of quantum entanglement as a resource is still poorly understood. Specifically, the performance of NOON states for optical phase imaging performs worst in comparison with the class of other states including entangled coherent states, entangled squeezed coherent states, and entangled squeezed vacuum[98]. This suggests that mode entanglement alone is not sufficient to provide certain precision enhancements. Indeed, too much entanglement is detrimental to attaining the Heisenberg scaling in the estimation of unitarily generated parameters[99]. Similarly, the estimation precision of n optical phase differences can be enhanced by using an n-mode entangled state as input in a multi-mode interferometer[100–102]. However, this precision enhancement has also been matched using separable states with equal number of modes[103]. This has been demonstrated theoretically for spatial distinguishability of different light emitters[104]. Alternative approaches to achieve quantum enhanced measurements have been investigated[105]. These methods rely on the use of quantum correlations, and nontrivial Hamiltonian extensions.

H. Optimal estimation strategies

Once the fundamental limits to the precision of parameter estimations have been determined, a natural question that arises is given that all classical noise has been eliminated, how can we identify the measurement(s) that practically saturate these bounds? Optimal measurements can be constructed from the eigenstates of the SLD[51]. In almost all cases, determining the measurement that corresponds to this theoretical description is difficult. Generally it depends on the parameters that we would like to estimate.

Adaptive strategies have been suggested to circumvent the parameter dependence of the optimal measurements[106]. Wiseman showed that feedback control of the phase of a local oscillator can approximate the measurement of the phase quadrature in an optical mode[107]. This was demonstrated experimentally by Armen *et al.*[108] By adaptively changing the phase in one arm of a Mach-Zehnder interferometer, nearly optimal measurement of the relative phase given N input photons can be achieved[109,110]. This technique was extended to narrowband squeezed beams by Berry and Wiseman[111,112]. Fujiwara proved that a sequence of maximum likelihood estimators is asymptotically efficient for adaptive quantum parameter estimation[113]. Okamoto *et al.* use this technique to estimate the phase between left- and right-handed circular polarisation of single photons[114]. For a review of quantum feedback control techniques, see Serafini[115]. Palittapongarnpim and Sanders proposed tests to see whether adaptive strategies in quantum metrology are robust against phase noise[116].

Achieving the Heisenberg limit is state dependent. However, the probe state chosen should be tailored to achieve the best practical precision for a specific parameter. For example, squeezed light is routinely used for phase estimations[117]. A natural question to ask is what is the optimal probe state that maximises the QFI for a parameter estimation protocol? This was answered by Braunstein, Caves and Milburn[30] and Giovannetti, Lloyd and Maccone[62] in the context of unitary evolutions $\exp(-i\theta G)$. The optimal probe state is an equal superposition of eigenstates corresponding the minimum and maximum eigenvalues of the generator G. These states are generally difficult to prepare.

Only very few experiments have reported a Heisenberg limit scaling for parameter estimates[118,119]. This is generally due to two factors. First, achieving the SQL is already practically difficult since it requires eliminating all non-intrinsic system noises. Second, state entanglement of multipartite systems is challenging to realise due to their increasing susceptibility to environmental losses with increasing particle number. For example, the path-entangled NOON states in Eq. (115) can be shown to achieve the Heisenberg limit resolution scaling for phase measurements in optical interferometers[24], but for larger photon number the loss of a single photon becomes increasingly likely, and this completely destroys the capability of measuring the parameter θ.

Even modest Markovian noise reduces the Heisenberg limit scaling achievable by highly entangled states to scalings proportional to the SQL[7,35,120]. Common decoherences include depolarisation, dephasing and amplitude damping. Owing to the difficulty and stabilisation of highly entangled states, alternative approaches to achieve quantum enhanced measurements have been investigated[105]. These methods rely on the use of quantum correlations and identical particles such as photons.

I. Numerical approaches

As has been shown so far, if the QFI is known, the fundamental precision bound is known and the optimal measurement strategy can be determined. Often however, it is not possible to find the QFI analytically, for example when probe states with a large rank are used. In those cases a numerical approach may be better suited.

Saturating the QCRB requires a suitable choice of estimator, which can be found numerically. Unfortunately, the numerical method required to determine a well-behaved and efficient estimator depends on the estimation problem, since the procedure will generally depend on how the parameters are encoded in the state. However, a common procedure used is maximum likelihood estimation, which given its simplicity, has found widespread use in estimation theory. The maximum likelihood procedure attempts to find the values of the parameters $\boldsymbol{\theta}$ that maximise the log likelihood function $\ln[p(\boldsymbol{x}|\boldsymbol{\theta})]$. In some circumstances this may be as simple as taking the derivative of the likelihood function and equating it to zero to find the maximum. However, this is not always a straightforward operation and alternative methods for obtaining the maximum likelihood estimator must be used.

Numerically, the maximum likelihood estimator may be implemented via an iterative scoring algorithm[2]. Defining the kth iteration of the estimator by $\check{\boldsymbol{\theta}}^{(k)}$, the scoring algorithm proceeds according to the iterative equation

$$\check{\boldsymbol{\theta}}^{(k+1)}(\mathbf{x}) = \check{\boldsymbol{\theta}}^{(k)}(\mathbf{x}) + I_Q^{-1}\left[\check{\boldsymbol{\theta}}^{(k)}\right] \frac{\partial \ln\left[p(\mathbf{x}|\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\check{\boldsymbol{\theta}}^{(k)}}. \quad (121)$$

Based on any information on the system, by taking an initial guess of the parameters $\check{\boldsymbol{\theta}}^{(0)}$, successive iterations of the scoring algorithm generate estimates that more closely approximates the true value. For open quantum systems the Markov chain Monte Carlo integration and Metropolis Hastings algorithm are better suited than the scoring algorithm[121].

V. MULTI-PARAMETER QUANTUM ESTIMATION

In this section, we review enhanced quantum parameter estimation of multiple parameters simultaneously. Many practical high-precision estimation protocols require a multi-parameter estimation approach. This includes the estimation of multiple phases[100,122–124], characterisation of multidimensional fields[99,125,126], and Hamiltonian tomography[127,128].

Multi-parameter quantum metrology raises two important questions. First, what is the attainability of the multi-parameter QCRB. If the SLDs for each parameter are mutually compatible—that is they commute with each other—then a simultaneous, optimal estimate for all of the parameters can be made in their common eigenbasis. If the optimal measurements corresponding to the different SLDs for each parameter do not commute, a compromise between the estimation precision for each parameter must be addressed. Second, what is the tradeoff between estimation precision enhancement and the physical resources used to attain it[129]? Specifically, is it better to estimate a tuple of parameters simultaneously or sequentially? Addressing these questions will help in the design of novel estimation schemes that propel precisions closer to the Heisenberg limit.

In this section we will first derive the QFI matrix for multiple parameters, and construct the corresponding Cramér-Rao bound. We consider alternatives to the QFI, based on the right logarithmic derivative, as well as the Kubo-Mori information and the Wigner-Yanase skew information. We conclude this section with a discussion of the Holevo bound and the general attainability of the QCRB.

A. The quantum Fisher information matrix

The QFI in section IV produces a real positive number associated with a single parameter θ that can be written as an inner product

$$I_Q(\theta) = \mathrm{Tr}\left[\partial_\theta \rho\, \mathscr{L}_\rho(\partial_\theta \rho)\right]. \quad (122)$$

For multiple parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D)$ the QFI becomes a matrix, since the generalisation to the D-dimensional parameter space creates a natural two-form[40]

$$[I_Q(\boldsymbol{\theta})]_{jk} = \mathrm{Tr}\left[\partial_j \rho\, \mathscr{L}_\rho(\partial_k \rho)\right], \quad (123)$$

where $j, k = \{1, 2, \ldots, D\}$. If we furthermore associate a new SLD $L_j = \mathscr{L}_\rho(\partial_j \rho)$ with each parameter θ_j:

$$\partial_j \rho = \frac{1}{2}\left(L_j \rho + \rho L_j\right), \quad (124)$$

where $\partial_j \equiv \partial/\partial\theta_j$, then in terms of the SLDs the QFI matrix becomes

$$[I_Q(\boldsymbol{\theta})]_{jk} = \frac{1}{2}\mathrm{Tr}\left[\rho\{L_j, L_k\}\right]. \quad (125)$$

The anti-commutator appears due to the possibility of a nonzero commutator between L_j and L_k. Since the optimal estimator for θ_j is given by the projectors along the eigenvectors of L_j, it is clear that in general non-commuting L_j and L_k will cause trouble for the simultaneous estimation of θ_j and θ_k. Nevertheless, the QFI matrix is well-defined, and we can write for the matrix elements

$$[I_Q(\boldsymbol{\theta})]_{lm} = \sum_{jk} \frac{2}{p_j + p_k} \langle j|(\partial_l \rho)|k\rangle\langle k|(\partial_m \rho)|j\rangle, \quad (126)$$

which again is hard to calculate in general, and does not include pure states.

When $\rho = |\psi\rangle\langle\psi|$ is a pure state and the evolution of the parameters is given by $\exp(-i\theta_j G_j)$ with G_j the self-adjoint generators of θ_j, the state $|\psi\rangle$ obeys the Schrödinger-like equations

$$i\,\partial_j|\psi\rangle = G_j|\psi\rangle\ . \tag{127}$$

The SLD L_j can then be written as

$$L_j = 2\partial_j\rho = 2\left(|\partial_j\psi\rangle\langle\psi| + |\psi\rangle\langle\partial_j\psi|\right), \tag{128}$$

with $|\partial_j\psi\rangle = \partial_j|\psi\rangle$. By substituting the SLD into Eq. (125), we obtain

$$[I_Q(\boldsymbol{\theta})]_{jk} = 4\mathrm{Re}\left(\langle\partial_j\psi|\partial_k\psi\rangle - \langle\partial_j\psi|\psi\rangle\langle\psi|\partial_k\psi\rangle\right)\ . \tag{129}$$

This is the multi-parameter quantum Fisher information for pure states and simple unitary evolution. We obtain a particularly useful form for $I_Q(\psi,\boldsymbol{\theta})$ when we relate the derivatives of $|\psi\rangle$ to the generators G_j associated with θ_j. Eq. (125) becomes

$$\begin{aligned}[I_Q(\boldsymbol{\theta})]_{jk} &= 4\left(\langle\psi|\frac{1}{2}\{G_j,G_k\}|\psi\rangle - \langle\psi|G_j|\psi\rangle\langle\psi|G_k|\psi\rangle\right)\\ &\equiv 4\,\mathrm{Cov}_S(\mathbf{G})_{jk}\,,\end{aligned} \tag{130}$$

where we defined $\mathbf{G} = (G_1,\ldots,G_D)$, and

$$\mathrm{Cov}_S(\mathbf{G})_{jk} = \langle\psi|\frac{1}{2}\{G_j,G_k\}|\psi\rangle - \langle\psi|G_j|\psi\rangle\langle\psi|G_k|\psi\rangle \tag{131}$$

is the symmetrised covariance matrix for $\mathbf{G}$. This can also be written in terms of a non-symmetrised covariance matrix $\mathrm{Cov}(\mathbf{G})$ according to

$$\mathrm{Cov}_S(\mathbf{G})_{jk} = \frac{1}{2}\left[\mathrm{Cov}(\mathbf{G})_{jk} + \mathrm{Cov}(\mathbf{G})_{kj}\right], \tag{132}$$

where $\mathrm{Cov}(\mathbf{G})_{jk} = \langle\psi|G_jG_k|\psi\rangle - \langle\psi|G_j|\psi\rangle\langle\psi|G_k|\psi\rangle$. An alternative form for $I_Q(\boldsymbol{\theta})$ is then

$$[I_Q(\boldsymbol{\theta})]_{jk} = 4\mathrm{Re}\ \mathrm{Cov}(\mathbf{G})_{jk}\ . \tag{133}$$

When all G_j commute with each other, the covariance matrix $\mathrm{Cov}(\mathbf{G})$ is real. The expressions in Eqs. (130) and (133) are the multi-parameter generalisations of Eq. (99). The multi-parameter QFI is a manifestly symmetric positive semi-definite matrix, and like the classical Fisher information matrix it transforms as a tensor under re-parameterisation. Defining again a new set of variables $\boldsymbol{\vartheta}$ through some invertible transformation Jacobian matrix, J, such that $\boldsymbol{\vartheta} = J\boldsymbol{\theta}$, then the QFIM for the new parameters may be written

$$I_Q(\boldsymbol{\vartheta}) = J^\top I_Q(\boldsymbol{\theta})J, \quad \text{with} \quad J_{kl} = \frac{\partial\theta_k}{\partial\vartheta_l}\ . \tag{134}$$

We can bound the QFI for general mixed states by the covariance matrix, just as we did for a single parameter in Eq. (98).

We start with the general definition in Eq. (125) and note that we can modify the derivative of ρ according to

$$\partial_j\rho = -i[G_j,\rho] = -i[G_j - \langle G_j\rangle,\rho] \equiv -i[\Delta G_j,\rho], \tag{135}$$

where we defined $\langle G_j\rangle = \mathrm{Tr}[\rho\,G_j]$. Using $\Delta G \equiv G - \langle G\rangle$, the SLD for θ_k is then

$$\begin{aligned}\mathscr{L}_\rho(\partial_k\rho) &= -2i\sum_{lm}\frac{[\Delta G_k,\rho]_{lm}}{p_l+p_m}|l\rangle\langle m|\\ &= 2i\sum_{lm}\frac{p_l-p_m}{p_l+p_m}\langle l|\Delta G_k|m\rangle|l\rangle\langle m|\ .\end{aligned} \tag{136}$$

Similarly, we find

$$\partial_j\rho = -i\sum_n p_n\left(\Delta G_j|n\rangle\langle n| - |n\rangle\langle n|\Delta G_j\right)\ . \tag{137}$$

Putting this together in Eq. (126), we obtain

$$[I_Q(\boldsymbol{\theta})]_{jk} = 2\sum_{lm}(p_l+p_m)\left(\frac{p_l-p_m}{p_l+p_m}\right)^2\langle l|\Delta G_k|m\rangle\langle m|\Delta G_j|l\rangle\ . \tag{138}$$

Writing this in symmetric form and noting that

$$\left(\frac{p_l-p_m}{p_l+p_m}\right)^2 \le 1 \quad \text{and} \quad \sum_{lm}(p_l+p_m)A_{lm} = 2\sum_{lm}p_lA_{lm}\,, \tag{139}$$

for any symmetric matrix A and probabilities $0 \le p_m \le 1$, we infer that

$$[I_Q(\boldsymbol{\theta})]_{jk} \le 4\mathrm{Tr}\left[\rho\frac{1}{2}\{\Delta G_j,\Delta G_k\}\right]. \tag{140}$$

When ρ is a pure state, the trace reduces to the matrix element $\mathrm{Cov}_S(\mathbf{G})_{jk}$ in Eq. (131). We can therefore define a more general symmetrized covariance matrix for the generators of translation $\mathrm{Cov}_S(\mathbf{G})_{jk} = \mathrm{Tr}[\rho\frac{1}{2}\{\Delta G_j,\Delta G_k\}]$. Similarly it is easy to show that we can cast the QFI matrix in non-symmetric form using the real part of a suitably generalised non-symmetric covariance matrix for the generators.

Returning for the moment again to the QFI matrix for pure states, another useful form for $I_Q(\boldsymbol{\theta})$ in Eq. (129) is[130]

$$[I_Q(\boldsymbol{\theta})]_{jk} = 4\partial_j^{(1)}\partial_k^{(2)}\ \log|\langle\psi(\boldsymbol{\theta}^{(1)})|\psi(\boldsymbol{\theta}^{(2)})\rangle|\Big|_{\boldsymbol{\theta}^{(1)}=\boldsymbol{\theta}^{(2)}=\boldsymbol{\theta}}\ . \tag{141}$$

This can be easily shown by explicitly evaluating the derivatives. The expression then reduces to Eq. (129).

In Lyapunov form, the QFI for multiple parameters becomes

$$[I_Q(\boldsymbol{\theta})]_{jk} = 2\int_0^\infty \mathrm{d}s\ \mathrm{Tr}\left[(\partial_j\rho)e^{-\rho s}(\partial_k\rho)e^{-\rho s}\right]. \tag{142}$$

We can use this form to construct an analytic expression for the QFI without having to evaluate the integral[131]. It makes use of the concept of *vectorisation* of matrices, denoted by vec A

of a matrix A, where the columns of a matrix are put below each other in a single column:

$$[I_Q(\boldsymbol{\theta})]_{jk} = 2\,\mathrm{vec}(\partial_j\rho)^\dagger\,(\rho^*\otimes\mathbb{I}+\mathbb{I}\otimes\rho)^{-1}\,\mathrm{vec}(\partial_k\rho)\,. \quad (143)$$

Similarly, we can construct the SLD:

$$\mathrm{vec}\,L_j = 2\,(\rho^*\otimes\mathbb{I}+\mathbb{I}\otimes\rho)^{-1}\,\mathrm{vec}(\partial_j\rho)\,. \quad (144)$$

These expressions are valid for finite-dimensional systems, and can be calculated directly based on matrix forms of the density matrix ρ and its derivatives $\partial_j\rho$. When ρ is pure, the inverse $(\rho^*\otimes\mathbb{I}+\mathbb{I}\otimes\rho)^{-1}$ does not exist. However, we can circumvent this problem by using instead the density matrix

$$\rho_\epsilon = (1-\epsilon)\rho + \epsilon\,\frac{\mathbb{I}}{d}\,, \quad (145)$$

where d is the dimension of the Hilbert space of ρ. Calculating the QFI and SLD based on ρ_ϵ, and taking the limit of $\epsilon\to 0$ then retrieves the forms of the QFI and SLD for pure states[131].

B. The quantum Cramér-Rao bound

Now that we have the QFI matrix in terms of the SLDs L_j associated with θ_j, as given in Eq. (136), we can derive the multi-parameter quantum Cramér-Rao bound. This inequality was originally first derived by Helstrom[13,48] and is occasionally referred to as the Helstrom bound. We start with the following identity for *unbiased* estimators

$$\mathrm{Tr}\big[(\partial_j\rho)(M_k-\theta_k)\big] = \delta_{jk}\,, \quad (146)$$

where M_k is the estimator for θ_k and δ_{jk} is the Kronecker delta. Using the SLD, this can be written as

$$\delta_{jk} = \frac{1}{2}\mathrm{Tr}\big[(\rho L_j + L_j\rho)(M_k-\theta_k)\big]\,. \quad (147)$$

Next, we introduce two real-valued vectors $\mathbf{y}$ and $\mathbf{z}$ such that

$$\begin{aligned}\sum_k z_k y_k &= \frac{1}{2}\sum_{jk}\mathrm{Tr}\big[z_j(\rho L_j + L_j\rho)(M_k-\theta_k)y_k\big]\\ &= \mathrm{Re}\,\mathrm{Tr}\left[\rho\left(\sum_j z_j L_j\right)\left(\sum_k y_k(M_k-\theta_k)\right)\right].\end{aligned} \quad (148)$$

We can square both sides of the equation, and note that $(\mathrm{Re}\,u)^2 \le |u|^2$, with equality if and only if $\mathrm{Im}\,u = 0$:

$$(\mathbf{z}^\top\mathbf{y})^2 \le \left|\mathrm{Tr}\left[\rho\left(\sum_j z_j L_j\right)\left(\sum_k y_k(M_k-\theta_k)\right)\right]\right|^2\,, \quad (149)$$

where $\mathbf{z}^\top\mathbf{y}$ is the standard dot product between two real-valued vectors. Using the Schwarz inequality for traces in Eq. (92), we can write Eq. (149) as

$$\begin{aligned}(\mathbf{z}^\top\mathbf{y})^2 &\le \mathrm{Tr}\left[\rho\left(\sum_j z_j L_j\right)^2\right]\mathrm{Tr}\left[\rho\left(\sum_k y_k(M_k-\theta_k)\right)^2\right]\\ &\equiv (\mathbf{z}^\top I_Q\mathbf{z})(\mathbf{y}^\top\,\mathrm{Cov}(\boldsymbol{\theta})\,\mathbf{y})\,,\end{aligned} \quad (150)$$

where we identified the QFI matrix and defined the covariance matrix

$$\mathrm{Cov}(\boldsymbol{\theta})_{jk} = \mathrm{Tr}\big[\rho(M_j-\theta_j)(M_k-\theta_k)\big]\,. \quad (151)$$

For the choice of $\mathbf{z} = I_Q^{-1}\mathbf{y}$, we obtain the inequality

$$\mathbf{y}^\top\,\mathrm{Cov}(\boldsymbol{\theta})\,\mathbf{y} \ge \mathbf{y}^\top I_Q^{-1}\mathbf{y}\,. \quad (152)$$

This bound is valid for any real vector $\mathbf{y}$, and therefore simplifies to

$$\mathrm{Cov}(\boldsymbol{\theta}) \ge I_Q^{-1}, \quad (153)$$

where the matrix inequality $A\ge B$ means that $A-B$ is a positive semi-definite matrix. This is the famous multi-parameter quantum Cramér-Rao bound. Given that the QFI elements transform as a metric tensor, the QFI for the vector of parameters $\check{\boldsymbol{\vartheta}}$ can be written in terms of the QFI for $\boldsymbol{\theta}$, and the QCRB becomes

$$\mathrm{Cov}(\check{\boldsymbol{\vartheta}}) \ge I_Q(\boldsymbol{\vartheta})^{-1} = J^{-1}I_Q(\check{\boldsymbol{\theta}})^{-1}[J^\top]^{-1}. \quad (154)$$

where J is the transformation Jacobian with matrix elements $J_{kl} = \partial\theta_k/\partial\vartheta_l$.

We can immediately infer the MSE of a parameter θ_j as

$$\mathrm{Var}\,\theta_j \ge [I_Q^{-1}]_{jj}\,, \quad (155)$$

since the variances are the diagonal elements of the covariance matrix. Note that $[I_Q^{-1}]_{jj} \ge [I_Q]_{jj}^{-1}$, since the QFI is a positive definite matrix. Given a covariance matrix and a positive-definite risk matrix $\mathbf{R}$, we can balance the precision of the various parameters. This leads to the inequality

$$C \equiv \mathrm{Tr}[\mathbf{R}\,\mathrm{Cov}(\boldsymbol{\theta})] \ge \mathrm{Tr}\left[\mathbf{R}I_Q^{-1}\right] \equiv C_Q\,. \quad (156)$$

The QCRB for a single parameter can in principle be achieved asymptotically by a suitable measurement. The question is whether the multi-parameter QCRB can be attained. We explore this in the next subsection.

C. Saturating the quantum Cramér-Rao bound

The the optimal unbiased quantum estimators that saturate the QCRB takes the form

$$M_j(\boldsymbol{\theta}) = \theta_j\,\mathbb{1} + [\boldsymbol{I}_Q(\boldsymbol{\theta})^{-1}\cdot\mathbf{L}]_j, \quad (157)$$

which form a set of self-adjoint operators[51]. They are linear combinations of the SLDs $\mathbf{L} = (L_1,\ldots,L_D)$. Determining the measurement $M_j(\boldsymbol{\theta})$ is typically a difficult task, since it depends on $\boldsymbol{\theta}$. To overcome this difficulty, adaptive measurements have been suggested[109,110]. An important caveat to the multi-parameter QCRB is that the QCRB for multiple parameters is generally not saturable, since the optimal observables in Eq. (157) may not be compatible. It is easy to see that this may occur when the SLDs associated with the parameters do not commute. However, there is a bit more to it than that.

To explore the attainability of the multi-parameter QCRB, we consider a more general bound derived by Holevo[32,132].

$$C \geq \min_{\{M_j\}} \left(\mathrm{Tr}[\mathbf{R}\,\mathrm{Re}\,J] + \left\| \sqrt{\mathbf{R}}\,(\mathrm{Im}\,J)\sqrt{\mathbf{R}} \right\|_1 \right) \equiv C_H\,, \quad (158)$$

where $\boldsymbol{R}$ denotes a positive definite weight matrix, and $J_{jk} = \mathrm{Tr}[M_j M_k \rho(\boldsymbol{\theta})]$ and $\|\cdot\|_1$ is the trace norm[133]. The bound in Eq. (158) is the Holevo Cramér Rao bound (HCRB) and defines a scalar lower bound on the weighted mean square error and represents the best precision attainable with global measurements on an asymptotically large number of identical copies of a quantum state[134–137].

Despite its importance for practical metrology, the HCRB has seen limited use in multi-parameter quantum metrology. This is due to the non-trivial optimisation over a set of observables and the implementation of global measurements is a difficult task. However, few results that use the HCRB do exist for qubit state estimation[138], two-parameter estimation with pure states[1] and two-parameter displacement estimation with two-mode Gaussian states[139,140]. A recent study by Albarelli *et al.* investigated the numerical tractability of calculating the HCRB for multi-parameter estimation problems[141], and an algorithm for finding a closed analytic form for the HCRB for two parameters was presented by Sidhu *et al.*[142].

The inequality in Eq. (158) follows[32] from the lemma that given the identity $\int_\Omega f(\omega) M(\mathrm{d}\omega) = F$ we have

$$\int_\Omega |f(\omega)|^2 M(\mathrm{d}\omega) \geq F F^\dagger\,, \quad (159)$$

where M is an observable with outcomes in Ω, $f(\omega)$ a complex function on Ω, and F an operator. We choose $f(\check{\boldsymbol{\theta}}) = \sum_j \xi_j(\check{\theta}_j - \theta_j)$ and $F = \sum_j \xi_j M_j$, substitute them into Eq. (159) and take the expectation value with respect to ρ. Taking into account a risk matrix $\mathbf{R}$ and optimising over all M_j we obtain Eq. (158).

The Holevo form of the QCRB can be attained when the statistical model involves the broad class of Gaussian state shifts where the parameters are encoded in shifts of the first moment[132,135,136,143]. When we choose the measurements M_j derived from the SLDs in Eq. (157) and substitute them into C_H, we find that the first term becomes the standard QCRB: $\min_{\{M_j\}} \mathrm{Tr}[\mathbf{R}\,\mathrm{Re}\,J] \rightarrow \mathrm{Tr}[\mathbf{R}\,I_Q^{-1}]$, and the second term becomes $\|\mathbf{R}\,\mathrm{Im}\,I_Q^{-1}\|_1$. Since I_Q^{-1} is a real positive definite matrix, this term vanishes, and C_H reduces to the QCRB based on the SLDs. The significance of Eq. (158) is that for non-commuting SLDs there may be a different set of observables that outperform the M_j in Eq. (157).

Next, we note that $\mathrm{Im}\,J$ can be written in terms of the commutator

$$\mathrm{Im}\,J_{jk} = \frac{1}{2}\mathrm{Tr}[\rho[M_j, M_k]]\,. \quad (160)$$

Assuming $\mathbf{R}$ and I_Q strictly positive matrices, and noting that $\mathrm{Tr}[\rho[M_j, M_k]] = 0$ implies that $\mathrm{Tr}[\rho[L_j, L_k]] = 0$, a necessary and sufficient condition for the saturability of the multi-parameter QCRB is then[122,124,133,138,144–147]

$$\mathrm{Tr}\big[\rho[L_j, L_k]\big] = 0\,. \quad (161)$$

This is of course a weaker condition than requiring that the SLDs commute directly. The condition in Eq. (161) is necessary and sufficient for unitary evolutions on pure states[144], which is equivalent to requiring the existence of commuting generators that generate the evolution of the probe. For mixed states, the demands to realise optimal simultaneous estimation are more involved. Specifically, we require the existence of a single probe state that maximises the QFI for all values of $\boldsymbol{\theta}$, a compatible measurement that ensures saturability of the QCRB, and a diagonal QFIM, which would allow independent estimations of each parameter[133]. Alternative methods to provide better precision bounds may involve collective measurements over many independent copies of the system, which is experimentally challenging. Recently, further conditions for optimal measurements to saturate the multi-parameter QCRB with separable or collective measurements on a *small* number of copies of the state have been explored by Yang *et al.*[148].

From the above discussion we see that the SLD operator plays a pivotal role in quantum estimation theory. For a multi-parameter estimation problem, finding the SLD for each parameter in $\boldsymbol{\theta}$ is sufficient to inform whether a simultaneous, efficient estimation can be performed. It also prescribes the optimal estimator that saturates the QCRB; the fundamental limit to estimation precisions allowed by quantum mechanics. We therefore turn our attention to find a functional form for the SLD.

D. Simultaneous versus sequential estimation

Multi-parameter quantum estimation is important for modelling a wider class of physical systems. For example, it may be necessary to infer the value of a parameter by estimating a set of related but different parameters. Also, there are examples where knowledge of multiple parameters are required, such as for microscopy, optical, electromagnetic, and gravitational field imaging. One approach for the estimation of multiple parameters is to prepare individual optimal probe and measurements schemes for each parameter. However, this is generally challenging to implement experimentally. It would also be unsuitable for sensing dynamically evolving probes. Instead, a natural approach would be to simultaneously estimate each parameter at the same time. The QCRB matrix bounds the precision of simultaneous multi-parameter estimates, and it could in principle be faster to implement the measurements simultaneously with fewer resources. For example, in the case of estimating D phases, Humphreys *et al.*[100] have demonstrated that simultaneous estimation provides an intrinsic $\mathcal{O}(D)$ precision improvement over the best quantum scheme for individual measurements devised by Lee, Kok, and Dowling[64] using NOON states. A similar advantage has been demonstrated by Baumgratz and Datta for multi-field estimation[99]. Within the literature, simultaneous strategies are also referred to as parallel estimation strategies. Even for large photon losses in phase imaging applications, simultaneous estimation schemes can provide a constant factor advantage over individual schemes. This has seen a surge of recent work focused on yielding quantum enhanced sensing from simulta-

neous estimation of multiple parameters[122,126,145,149–151].

The principal difficulty of simultaneous multi-parameter estimation arises from the non-attainability of the precision bounds when the infinitesimal generators associated with the parameters do not mutually commute[152–154]. This can lead to the requirement of non-Hermitian measurements to saturate precision bounds[155], which would be difficult to realise. We also saw in the previous section that the necessary condition to saturate the QCRB asymptotically is a zero expectation value of the commutator of the SLDs, as shown in Eq. (161). If this condition holds, there exists a common optimal measurement basis for simultaneous estimates. This case is well studied with general results derived for the optimal estimation of multiple phases. For example, Macchiavello calculates the cost associated with the estimation and finds it increases with increasing number of phase measurements[156]. Ballester demonstrated that entanglement in the probe states and measurements provides no advantage for estimating multiple phases[157]. If however, Eq. (161) does not hold, there is no simultaneous eigenbasis for the optimal simultaneous estimation of all parameters. One option to counter this would be to search for alternative informative bounds. However, this depends strongly on the specific estimation problem. For example, Genoni *et al.* find that for the estimation of complex parameters, the so-called right logarithmic derivative is more informative[150] (see section V E). We have already seen alternative bounds based on the SLD and the Holevo bound.

Several studies have illuminated progress towards simultaneous estimation of parameters with non-commuting SLDs. Even when the condition in Eq. (161) is not satisfied, a precision advantage may still be granted when estimating all of the parameters simultaneously[133]. For non-commuting unitary operations, it has been shown that entanglement in both the probe states and measurements can attain the Heisenberg limit[129,158]. Constructing the optimal simultaneous projective measurements requires knowledge of the POVMs that saturate the QCRB. Knowledge of this provides an instructive guide for the design of multi-parameter estimation experiments and has been heavily investigated. For single parameter estimation, finding an optimal measurement for which the equality $I(\theta) = I_Q(\theta)$ holds is always possible[22,30]. Unfortunately, this does not trivially extend to multi-parameter, where few analytic results are known for specific physical systems. The first is the estimation of single qubit mixed states[135,159]. Simultaneous optimal estimation has also been considered for the class of pure states. Determining the attainability of the QCRB for pure states was considered by Matsumoto. Specifically, if

$$\mathrm{Im}\left[\langle\partial_l\psi_\theta|\partial_m\psi_\theta\rangle\right] = 0 \;\forall\; l, m \tag{162}$$

and the QFIM is invertible, then a measurement scheme exists for which the classical Fisher information saturates the QFIM[144]. This is weaker constraint than the requirement of commuting Hamiltonians for each parameter $[H_j, H_k] = 0$ for all j, k used for the estimation of multi-dimensional fields[99]. Further conditions on projective measurements for optimal phase-like estimation schemes using pure states has been discussed by Pezze *et al.*[160]. First, assuming we have a pure probe state $|\psi_\theta\rangle$ evolving under unitary dynamics and satisfying Eq. (162), then the projectors $\{|\Gamma_k\rangle\langle\Gamma_k|\}$ constructed from the probe state and from vectors on the orthogonal subspace always saturates the QCRB, which can be experimentally realised. Pezze *et al.* also demonstrate that there always exists a set of projectors nonorthogonal to the probe such that $\langle\Gamma_k|\psi_\phi\rangle = 0$ for all k that saturates $I(\theta) = I_Q(\theta)$ if and only if[160]

$$\mathrm{Im}\left[\langle\partial_l\psi_\theta|\Gamma_k\rangle\langle\Gamma_k|\psi_\theta\rangle\right] = |\langle\psi_\theta|\Gamma_k\rangle|^2\,\mathrm{Im}\left[\langle\partial_l\psi_\theta|\psi_\theta\rangle\right]. \tag{163}$$

for all $l = 1, 2, \ldots, D$ and all $k \neq 1$. Finally, any general projective measurement for pure probe states under unitary dynamics is optimal if

$$\lim_{\phi\to\theta}\frac{\mathrm{Im}\left[\langle\partial_l\psi_\phi|\Gamma_k\rangle\langle\Gamma_k|\psi_\phi\rangle\right]}{\left|\langle\Gamma_k|\psi_\phi\rangle\right|} = 0, \tag{164}$$

for all indices l, m and all k for which $\langle\Gamma_k|\psi_\phi\rangle = 0$ and if Eq. (163) is fulfilled for all indices l and all k for which $\langle\Gamma_k|\psi_\phi\rangle \neq 0$. These conditions are necessary and sufficient for saturating the multi-parameter QCRB provided the QFIM is invertible. Further conditions for when the information content from simultaneous estimation of multiple parameters matches or exceeds that from separable measurements for general probe states was analysed by Ragy *et al.*[133].

Generally, the conditions for optimal simultaneous measurements may be hard to realise. This makes it difficult or impossible to saturate the QCRB matrix, and several tradeoffs are considered. The first deals with the probe states, since the optimal probe is generally different for each of the D parameters[161,162]. The second tradeoff deals with the choice of measurements for each parameter, as the optimal measurements for different parameters are usually incompatible[163]. To impose meaningful quantifiers of these tradeoffs, a cost function is introduced to assign a figure of merit. A general cost function, based on the Fisher information for a particular choice of probe state and measurement strategy is

$$C_\Pi(I_Q) = \mathrm{Tr}\left[I_Q(\boldsymbol{\theta})I(\boldsymbol{\theta})^{-1}\right], \tag{165}$$

where $I(\boldsymbol{\theta})^{-1}$ is the inverse of the classical Fisher information for a given POVM, Π. This quantifier enables optimisation over probe states and measurement strategies to address general tradeoffs in multi-parameter estimation schemes with non-commuting SLDs. Although any positive definite matrix can be used as the weight matrix in the cost function in Eq. (165), the QFIM is a natural choice for state estimation since it maximises the average fidelity between the estimated state and the actual state[159]. Notice that the cost function is lower bounded by $C_\Pi(I_Q) \geq D$, with equality applicable for measurements that are simultaneously optimal for all parameters. This cost function has been used to determine general bounds for thermometry of distant black bodies[164].

Using this strategy, a tradeoff between how well different parameters may be estimated is often considered. This approach has been explored by Crowley *et al.* for the estimation of phase ϕ and loss η in optical interferometry, where saturating the QCRB for both parameters simultaneously is impossible[124]. However, in many instances there is a single measurement observable for both ϕ and η that saturates the Holevo CRB[141].

For a general two-mode pure probe state with definite photon number N, Demkowicz-Dobrzanski *et al.* show that in lossy interferometers the density matrix for the evolved state can be written as the direct sum[4] $\rho(\boldsymbol{\theta}) = \oplus_{j=0}^{N} p_j \left|\psi_j\right\rangle \left\langle\psi_j\right|$. An intuition behind the origin for the tradeoff between precisions in phase and loss estimation can be understood by noting that the derivatives $\partial_\phi \left|\psi_j\right\rangle$ and $\partial_\eta \left|\psi_j\right\rangle$ are the same up to an imaginary constant[124]. Hence, estimating $\boldsymbol{\theta} = (\phi, \eta)^\top$ corresponds to measuring along the axes given by $\left|\psi_j\right\rangle \pm \partial_\phi \left|\psi_j\right\rangle$ and $\left|\psi_j\right\rangle \pm i\partial_\phi \left|\psi_j\right\rangle$[154]. Using a similar quantifier as in Eq. (165), the QFI for phase measurements can be saturated at the cost of the following diminished information on the loss parameter[124]

$$I(\eta) = \left[I_Q(\boldsymbol{\theta})\right]_{\eta\eta} - \frac{1}{4\eta^2}\left[I_Q(\boldsymbol{\theta})\right]_{\phi\phi}. \tag{166}$$

For the same probe state, a similar estimation trade-off has been explored for the estimation of phase and phase diffusion for different levels of noise[162].

Besides optimising over probe states and measurement strategies, a further optimisation over multi-parameter encoding channels has been considered in the literature. The two channel schemes that examine this optimisation is parallel and sequential estimation schemes. An important question then is which scheme is better suited for different quantum information tasks. In quantum estimation, based on general precision bounds for single parameters in the presence of noise, parallel schemes perform better than sequential feedback schemes[35,59]. For Hamiltonian parameter estimation on D-dimensional systems, sequential feedback has been demonstrated to provide an order of $\mathcal{O}(D+1)$ improvement over parallel schemes[161]. In quantum channel discrimination, the comparison of these two strategies has drawn a lot of attention[165–167]. In this scenario, sequential feedback schemes outperform parallel schemes, where the use of ancillary systems was necessary[168,169]. However, the use of ancillary systems or entanglement was relaxed in a separate protocol by Duan *et al.*, who developed the optimal sequential scheme to discriminate between different quantum channels[170].

E. The Right Logarithmic Derivative

The classical Fisher information is the unique Riemannian metric on the space of classical probability distributions that has the property of contraction under coarse graining. However, in the quantum-mechanical case the possibility of non-commuting operators breaks this uniqueness, and instead we obtain a family of metrics[171]. It is therefore interesting to explore some alternative quantum extensions of the Fisher information.

The QFI and the SLD are derived from the symmetrised correlation $\frac{1}{2}\langle\{A, B\}\rangle$ between two observables A and B (see Eq. (55)). This correlation can be interpreted as an instance of an inner product in the space of linear operators

$$(A, B)_\rho \equiv \frac{1}{2}\mathrm{Tr}\left[\rho\left(BA^\dagger + A^\dagger B\right)\right], \tag{167}$$

which also applies to operators A and B that are not self-adjoint. As shown in section IV, this inner product defines the metric operator $\mathscr{L}_\rho$ that is used to construct the SLD and the QFI. Our choice of the inner product was motivated by the possible non-commutativity of A and B. The QFI matrix elements are defined in terms of this inner product as

$$[I_Q]_{jk} = (L_j, L_k)_\rho, \tag{168}$$

with the usual definition for the SLDs.

However, there are also other definitions for the inner product we could use. For example, we could define the inner product as[39,155]

$$\langle\langle A, B\rangle\rangle_\rho \equiv \mathrm{Tr}\left[\rho B A^\dagger\right]. \tag{169}$$

This leads to a different metric, and correspondingly to a different logarithmic derivative. Following the same procedure as in section IV, the inner product leads to a raising operator $\mathscr{R}_\rho^{(R)}(B) = \rho B$. Since the lowering operator is defined as the inverse of the raising operator, we find

$$\mathscr{L}_\rho^{(R)}(\rho B) = B. \tag{170}$$

When we let the lowering operator act on the derivative $\partial_j\rho$, we obtain the so-called *Right* Logarithmic Derivative[152,153,155] (RLD)

$$R_j \equiv \mathscr{L}_\rho^{(R)}(\partial_j\rho), \tag{171}$$

which from Eq. (170) can be written as

$$\partial_j\rho = \rho R_j = R_j^\dagger \rho. \tag{172}$$

The information matrix corresponding to these RLDs is given by

$$[I_R]_{jk} = \left\langle\left\langle R_j, R_k\right\rangle\right\rangle_\rho = \mathrm{Tr}\left[R_j^\dagger \rho R_k\right]. \tag{173}$$

This inner product is a different way to define a metric in the space of linear operators, leading to an alternative way of measuring the distance between density matrices along curves parametrised by $\boldsymbol{\theta}$. The question is when and how this is useful for metrology.

We can derive a bound on the covariance matrix for $\boldsymbol{\theta}$ analogous to the QCRB, but instead based on the RLD. Unbiased estimators M_j for θ_j obey $\mathrm{Tr}[\rho M_j] = \theta_j$, and we find

$$\mathrm{Tr}\left[(\partial_k\rho)(M_j - \theta_j)\right] = \mathrm{Tr}\left[\rho(M_j - \theta_j)R_k^\dagger\right] = \delta_{jk}. \tag{174}$$

The fact that the RLD involves a simple product of operators R_j and ρ means that we can left- and right-multiply Eq. (174) by two complex vectors $\mathbf{y} = (y_1, \ldots, y_D)$ and $\mathbf{z} = (z_1, \ldots, z_D)$, respectively:

$$\sum_{jk}\mathrm{Tr}\left[y_j^*\rho(M_j - \theta_j)R_k^\dagger z_k\right] = \sum_j y_j^* z_j. \tag{175}$$

We separate the trace into a product of operators

$$\sqrt{\rho}\sum_j y_j^*(M_j-\theta_j) \qquad \text{and} \qquad \sqrt{\rho}\sum_k z_j^* R_j\,, \tag{176}$$

and use the Schwarz inequality for the trace to obtain the following inequality:

$$\left|\mathbf{y}^\dagger\mathbf{z}\right|^2 \leq \left(\sum_{jk} z_j^*\mathrm{Tr}\left[\rho R_j R_k^\dagger\right] z_k\right)\left(\sum_{jk} y_j^*\,\mathrm{Cov}(\boldsymbol{\theta})_{jk}\, y_k\right). \tag{177}$$

We can choose $\mathbf{y}$ and $\mathbf{z}$ any way we want. Using $[I_R]_{jk} = \mathrm{Tr}[\rho R_j R_k^\dagger]$ and setting $\mathbf{z} = I_R^{-1}\mathbf{y}$, the resulting inequality is

$$\mathbf{y}^\dagger\,\mathrm{Cov}(\boldsymbol{\theta})\,\mathbf{y} \geq \mathbf{y}^\dagger I_R^{-1}\mathbf{y}\,. \tag{178}$$

Again, since this must be true for any complex-valued vector $\mathbf{y}$, we obtain the inequality

$$\mathrm{Cov}(\boldsymbol{\theta}) \geq I_R^{-1}\,. \tag{179}$$

Clearly, the matrix I_R plays a role similar to the QFI I_Q, but now using the RLD. Expressed using a cost matrix $\mathbf{R}$, the bound based on the RLD becomes

$$\mathrm{Tr}[\mathbf{R}\,\mathrm{Cov}(\boldsymbol{\theta})] \geq \mathrm{Tr}\left[\mathbf{R}I_R^{-1}\right] \equiv C_R\,. \tag{180}$$

There are instances in which the this bound is tighter (i.e., more informative) that the QCRB, or $C_R > C_Q$.

For a single parameter the QCRB based on the SLD is always higher than the bound based on the RLD[29]. To show this, we write the RLD in the diagonal basis of the density operator. First, we use the definitions

$$\partial_\theta\rho = \sum_{mn}(\partial_\theta\rho)_{mn}\,|m\rangle\,\langle n|\,. \tag{181}$$

and

$$R_\theta = \sum_{jk} R_{jk}\,|j\rangle\,\langle k|\,. \tag{182}$$

The RLD can then be written as

$$R_\theta = \sum_{jk}\frac{(\partial_\theta\rho)_{jk}}{p_j}\,|j\rangle\,\langle k|\,. \tag{183}$$

From this it follows that

$$\begin{aligned} I_R &= \mathrm{Tr}[(\partial_\theta\rho)R_\theta] \\ &= \sum_{jk}\frac{|\langle j|(\partial_\theta\rho)|k\rangle|^2}{p_j} = \sum_{jk}\frac{|\langle j|(\partial_\theta\rho)|k\rangle|^2}{p_k} \\ &= \frac{1}{2}\sum_{jk}\left(\frac{1}{p_j}+\frac{1}{p_k}\right)|\langle j|(\partial_\theta\rho)|k\rangle|^2\,. \end{aligned} \tag{184}$$

We can now compare this with the QFI for a single parameter as in Eq. (88), and note that

$$\frac{2}{p_j+p_k} \leq \frac{1}{2}\left(\frac{1}{p_j}+\frac{1}{p_k}\right). \tag{185}$$

This means that for single parameter estimation problems $I_Q \leq I_R$ (when R_θ exists), and the QCRB based on the SLD is never less than the RLD-based bound. Therefore, the RLD bound is interesting only when we consider multiple parameters.

Another constraint is that the RLD does not exist for pure states. Nevertheless, a meaningful bound may still be constructed[39], relying on the fact that the non-existence of R_j for pure states does *not* imply the non-existence of I_R^{-1}. Fujiwara gives two examples. First, I_R^{-1} for the quadrature measurements (q,p) of a pure coherent state $|\alpha\rangle$ with complex amplitude α and frequency ω is given by

$$I_R^{-1} = \frac{1}{2}\begin{pmatrix}\omega & i\\ -i & \omega^{-1}\end{pmatrix}. \tag{186}$$

The information in Eq. (186) is interesting, since the associated bound in Eq. (180) takes the form of the additive uncertainty relation[172] for the rescaled position and momentum

$$\Delta\tilde{q}^2 + \Delta\tilde{p}^2 \geq 1\,, \tag{187}$$

where $\tilde{q} = q/\sqrt{\omega}$ and $\tilde{p} = \sqrt{\omega}\,p$. This is in contrast to the QCRB using the QFI for the evolution $U = \exp(-i\tilde{q}\tilde{p})$ with $\tilde{p}$ the momentum operator, which recovers the standard form of the uncertainty relation $\Delta\tilde{q}\,\Delta\tilde{p} \geq \frac{1}{2}$.

Second, for a measurement of a rotation $R(\theta,\phi)$ of a spin-j system in a state $|j\rangle$ oriented along the z-axis the information I_R^{-1} becomes

$$I_R^{-1} = \frac{1}{\sin^2\theta}\begin{pmatrix}\sin^2\theta & -i\sin\theta\\ i\sin\theta & 1\end{pmatrix}, \tag{188}$$

where θ and ϕ are the usual spherical coordinates. Again, this information translates via Eq. (180) into an additive uncertainty relation for the two rotation angles

$$\Delta\theta^2 + \sin^2\theta\,\Delta\phi^2 \geq 1\,. \tag{189}$$

Both these examples outperform the standard QCRB, and provide the most informative bounds[32,132,155,172]. Suzuki provides an explicit formula for the bound on two-parameter estimation using qubits[138].

Finally, the RLD is generally not Hermitian. Consequently, the optimal estimator derived from these two quantities may not correspond to a physical POVM. Despite this, the use of non-self adjoint operators have been demonstrated to saturate the multi-parameter QCRB[1,155].

It is interesting to consider how the RLD bound relates to the Holevo form of the QCRB in Eq. (158). Nagaoka[32] defines the RLD bound in Eq. (180) as

$$C_R = \mathrm{Tr}\left[\mathbf{R}\,\mathrm{Re}\,I_R^{-1}\right] + \left\|\sqrt{\mathbf{R}}\,(\mathrm{Im}\,I_R^{-1})\sqrt{\mathbf{R}}\right\|_1, \tag{190}$$

which takes the form of Eq. (158) when the observables M_j are chosen such that $J = I_R^{-1}$. These are generally not the same obervables found through minimisation, and we therefore find

$$C_H \geq \max\{C_Q, C_R\}\,, \tag{191}$$

and the Holevo form is the most informative bound on the covariance matrix of $\boldsymbol{\theta}$.

F. Kubo-Mori information

When the quantum system under consideration is in a thermal state, a more natural CRB can be found based on so-called Kubo-Mori information. The density operator of a thermal state can be written in exponential form as

$$\rho = \mathrm{e}^{-\beta H - \theta A}, \tag{192}$$

where $\beta = 1/k_B T$, with k_B Boltzmann's constant, T the temperature, H the Hamiltonian, and we absorbed the partition function in H. The term θA corresponds to a force on the system generated by A.

The quantum relative entropy for two states ρ and σ is given by[173,174]

$$D(\rho \,\|\, \sigma) = \mathrm{Tr}[\rho \log \rho] - \mathrm{Tr}[\rho \log \sigma]\,. \tag{193}$$

The quantum relative entropy for two states $\mathrm{e}^{-\beta H - \theta A}$ and $\mathrm{e}^{-\beta H - \theta' B}$ then becomes

$$\begin{aligned} D(\rho(\theta, A) \,\|\, \rho(\theta', B)) = \mathrm{Tr}[\mathrm{e}^{-\beta H - \theta A}(\beta H + \theta' B)] \\ - \mathrm{Tr}[\mathrm{e}^{-\beta H - \theta A}(\beta H + \theta A)]\,. \end{aligned} \tag{194}$$

Taking the derivative of the quantum relative entropy with respect to θ and θ', we obtain the relation

$$\begin{aligned} \frac{1}{\beta^2}\frac{\partial^2 D(A \,\|\, B)}{\partial\theta\partial\theta'} &= \langle A\rangle\langle B\rangle \\ &\quad - \frac{1}{\beta}\int_0^\beta \mathrm{Tr}\left[\mathrm{e}^{-\beta H}\mathrm{e}^{xH} A \mathrm{e}^{-xH} B\right] \mathrm{d}x\,, \end{aligned} \tag{195}$$

when $\theta' \to \theta$. This motivates the Bogoliubov inner product for thermal states:

$$(A, B)_\rho^{(\mathrm{B})} = \int_0^1 \mathrm{Tr}\left[\rho^x A^\dagger \rho^{1-x} B\right] \mathrm{d}x\,, \tag{196}$$

which can be used to construct a new logarithmic derivative[175] called the Bogoliubov Logarithmic Derivative (BLD)

$$\partial_\theta \rho = \int_0^1 \rho^x B_\theta\, \rho^{1-x}\ \mathrm{d}x\,. \tag{197}$$

and

$$B_\theta = \int_0^\infty (x+\rho)^{-1}(\partial_\theta \rho)\,(x+\rho)^{-1}\,\mathrm{d}x\,, \tag{198}$$

leading to the Kubo-Mori information

$$I_{\mathrm{KM}}(\theta) = (B_\theta, B_\theta)_\rho^{(\mathrm{B})} = \int_0^1 \mathrm{Tr}\left[\rho^x B_\theta^\dagger \rho^{1-x} B_\theta\right] \mathrm{d}x\,. \tag{199}$$

The form of the Bogoliubov inner product in Eq. (196) originates in the theory of linear response of thermal systems using the canonical correlation between A and B

$$\int_0^1 \mathrm{Tr}\left[\rho^x (A - \mathrm{Tr}[\rho A])\rho^{1-x}(B - \mathrm{Tr}[\rho B])\right] \mathrm{d}x\,,$$

and therefore the use of $I_{\mathrm{KM}}(\theta)$ in Eq. (199) can be considered more natural for statistical physics applications than the QFI[176]. For further details, see chapter 4 of Kubo, Toda, and Hashitsume[177].

A Cramér-Rao-type inequality can be derived when we consider the covariance of $M_\theta = \check{\theta} - \theta$, such that

$$\mathrm{Cov}_B(\theta) = \int_0^\infty \mathrm{Tr}\left[\rho^x M_\theta \rho^{1-x} M_\theta\right] \mathrm{d}x \geq I_{\mathrm{KM}}^{-1}(\theta)\,. \tag{200}$$

For a single parameter θ, the Kubo-Mori information is equal to

$$I_{\mathrm{KM}}(\theta) = \sum_{jk} \frac{\log p_j - \log p_k}{p_j + p_k} \left|\langle j|\partial_\theta \rho|k\rangle\right|^2\,. \tag{201}$$

This is a more informative bound than the one based on the RLD in Eq. (184), but less informative than the QCRB based on the SLD in Eq. (88)[175]. The Kubo-Mori information gives the bound for consistent superefficient estimators[176], i.e., estimators that outperform maximum likelihood estimation. The Kubo-Mori information plays a central role in the construction of uncertainty relations between energy and temperature in quantum thermodynamics[178].

G. Wigner-Yanase skew information

We have seen in Eqs. (88), (184), and (201) that we can write I_Q, I_R, and I_{KM} in terms of the matrix elements of the derivative of the density operator $\partial_\theta \rho$ in the basis of ρ:

$$I_Q(\theta) = \sum_{jk} \frac{2}{p_j + p_k} \left|\langle j|(\partial_\theta \rho)|k\rangle\right|^2\,, \tag{202}$$

$$I_R(\theta) = \sum_{jk} \left(\frac{1}{2p_j} + \frac{1}{2p_k}\right) \left|\langle j|(\partial_\theta \rho)|k\rangle\right|^2\,, \tag{203}$$

$$I_{\mathrm{KM}}(\theta) = \sum_{jk} \frac{\log p_j - \log p_k}{p_j + p_k} \left|\langle j|(\partial_\theta \rho)|k\rangle\right|^2\,. \tag{204}$$

The different types of quantum information are determined by the functions $c(p_j, p_k)$ of the eigenvalues of ρ in the above equations. Morozova and Chentsov (and extended by Petz) proposed these functions as a classification of monotone Riemannian metrics on matrix spaces[171,179], and they include another interesting example:

$$I_{\mathrm{WY}}(\theta) = \sum_{jk} \frac{4}{\left(\sqrt{p_j} + \sqrt{p_k}\right)^2} \left|\langle j|(\partial_\theta \rho)|k\rangle\right|^2\,. \tag{205}$$

This is the Wigner-Yanase skew information[180]. When the parameter θ is generated unitarily by the self-adjoint operator G, the form becomes

$$\begin{aligned} I_{WY}(G) &= -\frac{1}{2}\mathrm{Tr}\left[[\sqrt{\rho}, G]^2\right] \\ &= \mathrm{Tr}\left[G^2\rho\right] - \mathrm{Tr}\left[G\rho^{\frac{1}{2}} G \rho^{\frac{1}{2}}\right], \end{aligned} \tag{206}$$

which no longer relies on finding the spectrum of ρ. In the limit of pure states, I_{WY} reduces to the variance of G. For unitary evolutions $\exp(-i\theta G)$, is was shown[181] that $I_Q(\theta) = 8I_{\text{WY}}(G)$.

Wigner and Yanase sought to capture the sense in which observables that commute with some conserved quantity G are easier to measure than observables that do not[180]. The skew information they propose measures the information in ρ about not-so-easy-to-measure observables, e.g., it provides a measure of non-commutativity between ρ and G. Their definition ensures that I_{WY} is convex, additive, and is non-increasing over time for open quantum systems.

As with most other metrics in this review, the Wigner-Yanase skew information can be used to derive uncertainty relations between observables A and B. Furuichi and Yanagi define the correlation between observables as[182]

$$\text{Cor}(A, B) \equiv \text{Tr}[\rho AB] - \text{Tr}\left[\sqrt{\rho}A\sqrt{\rho}B\right], \tag{207}$$

and prove that this is bounded by

$$\text{Cor}\,(A, B)^2 \geq I_{\text{WY}}(A)\, I_{\text{WY}}(B)\,. \tag{208}$$

There are many other uses for the Wigner-Yanase skew information, including as an entanglement witness[183].

Another way to understand the Wigner-Yanase skew information is as a measure of the *quantum* fluctuations in a state ρ for an observable A, as opposed to the variance $(\Delta A)^2$, which measures both the classical and the quantum fluctuations:

$$(\Delta A)^2 = -\frac{1}{2}\text{Tr}\left[[\rho^a, A][\rho^{1-a}, A]\right] + \text{Tr}\left[\rho^a \delta A \rho^{1-a} \delta A\right], \tag{209}$$

where $a \in [0, 1]$ and $\delta A = A - \langle A\rangle$. The first term is the Wigner-Yanase skew information generalised by Dyson, which reduces to $I_{\text{WY}}(A)$ when $a = \frac{1}{2}$. The second term is the classical uncertainty in A given the state ρ. A common approach[178,184,185] is to average the quantum fluctuations over the possible values of a:

$$-\frac{1}{2}\int_0^1 \text{Tr}\left[[\rho^a, A][\rho^{1-a}, A]\right] \text{d}a \;= \left\langle A^2\right\rangle - (A, A)_\rho^{(\text{B})}\,. \tag{210}$$

This relates the average to the Bogoliubov inner product in Eq. (196), from which the Kubo-Mori information was derived. Hence the Bogoliubov inner product is intimately related to the classical fluctuations of A given the state ρ. It is easily verified that for pure states $(A, B)_\psi^{(\text{B})} = \langle A\rangle\langle B\rangle$ and the classical uncertainty vanishes.

H. Bayesian quantum estimation theory

So far we have considered Fisher estimation theory, where $\boldsymbol{\theta}$ is an unknown parameter that is not random. However, often the parameter $\boldsymbol{\theta}$ can itself be a random variable. This is handled by the Bayesian approach to probability theory. In this case we have some probability distribution $p(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$, called the *prior*. Bayesian quantum metrology using prior information was considered by Demkowicz-Dobrzański[186], and Macieszczak, Fraas and Demkowicz-Dobrzański[187].

Let $p(\mathbf{x}, \boldsymbol{\theta})$ be the *joint* probability distribution of $\mathbf{x}$ and $\boldsymbol{\theta}$. Bayes' theorem then states that $p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$. We can change the natural expectation value to

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\,[f] = \int \text{d}\mathbf{x}\;\text{d}\boldsymbol{\theta}\; p(\mathbf{x}, \boldsymbol{\theta})\, f(\mathbf{x}, \boldsymbol{\theta}), \tag{211}$$

for some integrable function $f(\mathbf{x}, \boldsymbol{\theta})$. The Bayesian covariance matrix becomes

$$\text{Cov}(\check{\boldsymbol{\theta}}) = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\left(\check{\boldsymbol{\theta}} - \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\check{\boldsymbol{\theta}}\right]\right)\left(\check{\boldsymbol{\theta}} - \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\check{\boldsymbol{\theta}}\right]\right)^{\top}\right]. \tag{212}$$

This no longer depends directly on $\boldsymbol{\theta}$ via the probability distribution $p(\mathbf{x}|\boldsymbol{\theta})$, and it incorporates the prior information about $\boldsymbol{\theta}$. Classically, we can switch between Fisher and Bayesian estimation by simply replacing the expectation values throughout, even though the interpretations of the resulting quantities will be subtly different in important ways. For example, the classical MSE will depend on the true value of $\boldsymbol{\theta}$, while the Bayesian MSE is averaged over $\boldsymbol{\theta}$.

The optimal estimator $\check{\theta}_B$ that minimises the Bayesian MSE of a parameter θ is the mean of the posterior probability distribution[2] $p(\theta|\mathbf{x})$

$$\check{\theta}_B(\mathbf{x}) = \int \text{d}\theta\;\theta\; p(\theta|\mathbf{x})\,. \tag{213}$$

This estimator changes as additional data $\mathbf{x}$ is obtained. In addition, the choice of a prior probability distribution $p(\theta)$ is important for a successful Bayesian parameter estimation. For more details on Bayesian estimation in the classical domain, see Kay (1993)[2].

To derive the Bayesian form of the quantum Fisher information we use again the density matrix of a system $\rho(\theta)$ evolved according to a parameter θ of interest. There is now a prior probability distribution $p(\theta)$ over θ, so that we can define a prior-weighted density operator as:

$$\overline{\rho} = \int \text{d}\theta\, p(\theta)\rho(\theta)\,. \tag{214}$$

Similarly, we can define the *posterior mean operator*

$$\overline{\theta\rho} = \int \text{d}\theta\, p(\theta)\theta\rho(\theta)\,. \tag{215}$$

Just like the QFI was given by the scalar product of $\partial_\theta\rho$ with itself, $I_Q(\theta) = (\partial_\theta\rho, \partial_\theta\rho)_\rho = \text{Tr}\left[\partial_\theta\rho\,\mathscr{L}_\rho(\partial_\theta\rho)\right]$, we define a new Bayesian information I_B based on $\overline{\theta\rho}$ using the same inner product that originated from the symmetrised operator correlation in Eq. (55):

$$I_B \equiv \left(\overline{\theta\rho}, \overline{\theta\rho}\right)_{\bar{\rho}} = \text{Tr}\left[\overline{\theta\rho}\,\mathscr{L}_{\bar{\rho}}(\overline{\theta\rho})\right] \tag{216}$$

A new symmetric logarithmic posterior mean operator $S_\theta \equiv \mathscr{L}_{\bar{\rho}}(\overline{\theta\rho})$ can be constructed from

$$\overline{\theta\rho} = \frac{1}{2}\left\{\overline{\rho}, S_\theta\right\}, \tag{217}$$

which plays a similar role to the SLD in the ordinary quantum Cramér-Rao bound. Using that $A = \frac{1}{2}\{\rho, B\}$ has a general solution[188]

$$B = 2\int_0^\infty \mathrm{d}s\; e^{-\rho s} A e^{-\rho s}\,, \tag{218}$$

we find

$$S_\theta = 2\int_0^\infty \mathrm{d}s\; e^{-\bar{\rho} s}\left(\overline{\theta\rho}\right) e^{-\bar{\rho} s}\,. \tag{219}$$

The quantum Bayesian information I_B then takes the form

$$I_B = 2\int_0^\infty \mathrm{d}s\; \mathrm{Tr}\left[\left(\overline{\theta\rho}\right) e^{-\bar{\rho} s}\left(\overline{\theta\rho}\right) e^{-\bar{\rho} s}\right] = \mathrm{Tr}\left[\overline{\rho}\, S_\theta^2\right]. \tag{220}$$

Contrary to the QFI, which has units of θ^{-2}, this information has units of θ^2.

Personick[188] showed how these definitions leads to a minimum MSE for the situation where we have a prior distribution $p(\theta)$ over the parameter θ. The MSE for θ is obtained via a measurement of some observable M:

$$\begin{aligned}\mathrm{Cov_B}(M) &\equiv \int \mathrm{d}\theta\; p(\theta)\mathrm{Tr}\left[\rho(\theta)\,(M-\theta\mathbb{I})^2\right] \\ &= \mathrm{Tr}\left[\overline{\rho} M^2\right] - 2\mathrm{Tr}\left[\left(\overline{\theta\rho}\right) M\right] + \int \mathrm{d}\theta\; p(\theta)\,\theta^2,\end{aligned} \tag{221}$$

where $\overline{\rho}$ and $\overline{\theta\rho}$ are defined as above. When $M = M_{\mathrm{opt}}$ is the optimal measurement that yields the minimum MSE, we have $\mathrm{Tr}\left[\overline{\rho} M_{\mathrm{opt}}^2\right] = \mathrm{Tr}\left[\overline{\theta\rho} M_{\mathrm{opt}}\right]$, and we obtain

$$\mathrm{Cov_B}(M_{\mathrm{opt}}) = \mathbb{E}_\theta\left(\theta^2\right) - \mathrm{Tr}\left[\left(\overline{\theta\rho}\right) M_{\mathrm{opt}}\right]. \tag{222}$$

For the detailed proof, see Personick[188]. Note that we used the subscript "B" to emphasise this is a Bayesian covariance. Our task is to find the optimal operator M that achieves the minimal MSE. It is given by Eq. (219), and M is Hermitian and unique[188], at least in the single-parameter case. The proof by Personick proceeds by considering

$$\mathrm{Cov_B}(M) \leq \mathrm{Cov_B}(M + \epsilon H), \tag{223}$$

with $\epsilon \geq 0$ and H and arbitrary Hermitian operator. Using the definition in Eq. (221) and expanding the square then leads to the condition that

$$\mathrm{Tr}\left[H\left(\overline{\rho} M + M\overline{\rho} - 2\overline{\theta\rho}\right)\right] = 0, \tag{224}$$

and from Eq. (217) the identification $M = S_\theta$ follows. The Bayesian quantum Cramér-Rao bound then becomes

$$(\Delta\theta)_B^2 \geq (\Delta\theta)_p^2 - I_B\,, \tag{225}$$

where $(\Delta\theta)_p^2$ is the variance of the prior distribution

$$(\Delta\theta)_p^2 = \int \mathrm{d}\theta\; p(\theta)\theta^2, \tag{226}$$

and the quantity $(\Delta\theta)_p^2 - I_B$ is the quantum Allan variance[189].

This construction was used by Macieszczak *et al.* for Bayesian quantum frequency estimation[187], leading to the introduction of the quantum Allan variance by Chabuda, Leroux, and Demkowicz-Dobrzański[189] and their application of the quantum Allan variance to the precision of atomic clocks. Rubio and Dunningham describe quantum metrology in the presence of limited data using this formalism[190]. Bernád, Sanavio and Xuereb use the quantum Bayesian estimation technique to estimate the nonlinear opto-mechanical coupling strength[191] and the matter-field coupling strength in the dipole approximation[192]. Rządkowski and Demkowicz-Dobrzański apply Bayesian techniques to discrete phases[193].

The multi-parameter form of $\mathrm{Cov_B}$ for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D)$ with estimator observables $\boldsymbol{S} = (S_1, \ldots, S_D)$ is

$$\mathrm{Cov_B}(\boldsymbol{S})_{jk} = \mathbb{E}_{\boldsymbol{\theta}}\left(\theta_j\theta_k\right) - \mathrm{Tr}\left[\overline{\theta_j\rho}\, S_k\right] - \mathrm{Tr}\left[\overline{\theta_k\rho}\, S_j\right] \tag{227}$$

$$+ \frac{1}{2}\mathrm{Tr}\left[\overline{\rho}\{S_j, S_k\}\right], \tag{228}$$

where $j, k \in \{1, \ldots, D\}$. Furthermore, we can prove that

$$\mathrm{Tr}\left[\overline{\theta_j\rho}\, S_k\right] + \mathrm{Tr}\left[\overline{\theta_k\rho}\, S_j\right] = \mathrm{Tr}\left[\overline{\rho}\{S_j, S_k\}\right], \tag{229}$$

and we can therefore introduce the *quantum Bayes information* matrix

$$[I_B(\boldsymbol{S})]_{jk} = \mathrm{Tr}\left[\overline{\theta_j\rho}\,\mathscr{L}_{\bar{\rho}}\left(\overline{\theta_k\rho}\right)\right] = \frac{1}{2}\mathrm{Tr}\left[\overline{\rho}\,\{S_j, S_k\}\right]. \tag{230}$$

The proof of this expression for the matrix follows from the fact that the Bayesian estimation case has the same geometric structure as the Fisher estimation case. The multi-parameter Bayesian minimum MSE bound can then be written as

$$\mathrm{Cov_B}(\boldsymbol{S})_{jk} \geq \mathbb{E}_{\boldsymbol{\theta}}\left(\theta_j\theta_k\right) - [I_B(\boldsymbol{S})]_{jk}\,. \tag{231}$$

A similar result for the multi-parameter Bayesian MSE was independently obtained by Rubio and Dunningham[194].

VI. SPECIAL CASES OF QUANTUM ESTIMATION

There are a number of special applications in quantum parameter estimation. First, in quantum optics we very often deal with Gaussian states, and these admit closed forms for the QFI and the SLDs. Second, we consider the case when the evolution of the probe state does not take the form of a simple unitary $U = \exp(-i\theta G)$.

A. Estimation using Gaussian states

The practical implementation of quantum parameter estimation often involves Gaussian quantum states, due to their ubiquity, relative easy of preparation, and their admittance of closed form analytical expressions for the QFI and the SLD. Of particular importance is the use of Gaussian states in the context of

quantum optics, where coherent, squeezed, and thermal states all form part of the Gaussian family of states[195].

Consider a system comprised of n modes of a bosonic field. The Hilbert space $\mathscr{H} = \otimes_{k=1}^{n} \mathscr{H}_k$ is the tensor product of the infinite dimensional Fock space of each mode k, that is spanned by the number basis $\{|m\rangle_k\}$, with m a natural number. The creation and annihilation operators for each mode are defined by

$$\begin{aligned} \hat{a}_k |m\rangle_k &= \sqrt{m}\, |m-1\rangle_k \\ \hat{a}_k^\dagger |m\rangle_k &= \sqrt{m+1}\, |m+1\rangle_k \,, \end{aligned} \tag{232}$$

obeying the commutation relation $[a_k, a_j^\dagger] = \delta_{jk}$, with all other commutators zero. In the phase space formalism, a conventional approach is to arrange these operators into a vectorial operator $\hat{\boldsymbol{a}} := (\hat{a}_1, \hat{a}_1^\dagger, \ldots, \hat{a}_n, \hat{a}_n^\dagger)^\top$. The vector then satisfies the compact commutation relation

$$\left[\hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j\right] = \boldsymbol{\Omega}_{ij} \tag{233}$$

where the matrix $\boldsymbol{\Omega} = \oplus_{j=1}^{n} i\sigma_y$ satisfies $\boldsymbol{\Omega}^\top = -\boldsymbol{\Omega} = \boldsymbol{\Omega}^{-1}$, and σ_y is the Pauli y matrix. An equivalent description of bosonic systems can be written in terms of the dimensionless canonical quadrature field operators:

$$\hat{q}_j = \frac{1}{\sqrt{2}}\left(\hat{a}_j + \hat{a}_j^\dagger\right), \quad \text{and} \quad \hat{p}_j = -\frac{i}{\sqrt{2}}\left(\hat{a}_j - \hat{a}_j^\dagger\right). \tag{234}$$

These observables act similar to the position and momentum operators of the quantum harmonic oscillator, and satisfy the canonical commutation relations $[\hat{q}_j, \hat{p}_k] = i\delta_{jk}$. Introducing the vector of operators $\hat{\boldsymbol{R}} = (\hat{q}_1, \hat{p}_1, \ldots, \hat{q}_n, \hat{p}_n)^\top$, we similarly re-write the canonical commutation relations in the compact form

$$\left[\hat{\boldsymbol{R}}_j, \hat{\boldsymbol{R}}_k\right] = i\boldsymbol{\Omega}_{jk}. \tag{235}$$

A Gaussian n-mode quantum state ρ is a state that is completely characterised by n first moments $\boldsymbol{\lambda}$ and an $n \times n$ second moment matrix $\boldsymbol{\Sigma}$ such that

$$\boldsymbol{\lambda} = \mathrm{Tr}[\rho\, \hat{\boldsymbol{a}}] = \langle \hat{\boldsymbol{a}} \rangle, \quad \boldsymbol{\Sigma}_{jk} = \mathrm{Tr}\left[\rho \left\{\hat{\bar{a}}_j, \hat{\bar{a}}_k\right\}\right], \tag{236}$$

with the zero mean operators $\hat{\bar{a}}_j = \hat{a}_j - \langle \hat{a}_j \rangle$, and $\{A, B\} = AB + BA$ the anti-commutator of operators A and B. The characteristic function of ρ is defined as

$$\chi(\boldsymbol{\zeta}) = \mathrm{Tr}\left[\rho \exp[-\hat{\boldsymbol{a}}^\top \boldsymbol{\Omega}\, \boldsymbol{\zeta}]\right], \tag{237}$$

where $\boldsymbol{\zeta} = (\zeta_1, \zeta_1^*, \ldots, \zeta_n, \zeta_n^*)^\top \in \mathbb{R}^{2n}$. For a Gaussian density matrix with first moment $\boldsymbol{\lambda}$ and second moment $\boldsymbol{\Sigma}$, the corresponding characteristic function is also Gaussian:

$$\chi(\boldsymbol{\zeta}) = \exp\left[\frac{1}{2}\boldsymbol{\zeta}^\top(\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}^\top)\boldsymbol{\zeta} - i(\boldsymbol{\Omega}\boldsymbol{\lambda}^\top\boldsymbol{\zeta})\right], \tag{238}$$

A Gaussian transformation is any trace preserving quantum channel that preserves the Gaussian nature of the characteristic function.

A common pure Gaussian state is the vacuum $\rho_{\text{in}} = |0\rangle\langle 0|$, while for mixed states it is the thermal state:

$$\rho_{\text{th}} = \sum_{n=0}^{\infty} \frac{\overline{n}^n}{(1+\overline{n})^{1+n}} |n\rangle\langle n|, \tag{239}$$

where $\overline{n} = \langle \hat{n} \rangle$ is the expectation of the photon number $\hat{n}$. Thermal radiation exhibits a Gaussian characteristic function with zero mean $\boldsymbol{\lambda}_{\text{th}} = \boldsymbol{0}$ and covariance $\boldsymbol{\Sigma}_{\text{th}} = \frac{1}{2}(2\overline{n}+1)\mathbb{1}$. Any single-mode Gaussian state can be written in terms of a displaced squeezed thermal state[195]:

$$\rho_{\text{G}} = S_\epsilon D_\alpha \rho_{\text{th}} D_\alpha^\dagger S_\epsilon^\dagger = U \rho_{\text{th}} U^\dagger, \tag{240}$$

where the Gaussian unitaries

$$D_\alpha = \exp\left[\alpha \hat{a}^\dagger - \alpha^* \hat{a}\right], \tag{241}$$

$$S_\epsilon = \exp\left[\frac{1}{2}(\epsilon^* \hat{a}^2 - \epsilon \hat{a}^{\dagger 2})\right], \tag{242}$$

are defined as the displacement operator and single mode squeezing operator, respectively. The ordering of a displacement and squeezing operations can be reversed according to

$$D_\alpha S_\epsilon = S_\epsilon D_\gamma, \tag{243}$$

with $\gamma = \alpha \cosh r_\epsilon - \alpha^* \exp[i\vartheta_\epsilon] \sinh r_\epsilon$. States described by Eq. (240) are readily prepared in laboratories and their unitary evolutions realised by use of lasers, linear optical elements, and squeezing mediums[196]. The multi-mode generalisation of Eq. (240) can be found using the matrix generalisation[197] of the operators in Eq. (242).

Both the mean and second moment of a Gaussian quantum state may depend on parameters $\boldsymbol{\theta}$ that we wish to estimate. To bound the precision of the estimation procedure, we aim to calculate the QFI of Gaussian quantum states. This is not always an easy task. All of the work described in the previous sections relied on knowledge of the density matrix, and for Gaussian states this is most conveniently addressed in the phase-space formalism by mapping transformations of the state to transformations of the moments[198].

The first work in this direction was completed by Pinel *et al.*[199] who derived the expression for the ultimate limit to parameter estimations using pure Gaussian states of arbitrarily many modes. A more general approach to finding the QFI and the SLD for Gaussian states is through the mean displacement and covariance matrix. This approach has been taken by a number of authors. The first result in this direction was by Monras in 2013, who derived the form of the SLD for general Gaussian states evolving under Gaussian unitaries[200]. By taking an SLD ansatz that is quadratic in the quadrature operators, the SLD and QFI are written as an infinite series solution to the Stein-equation. A similar approach by Jiang confirmed Monras' result and gave the SLD for states in exponential form in terms of the generator and its moments[201]. Gao and Lee followed an alternative method to derive the SLD and the QFI for multi-mode Gaussian states. The necessity of inverting relatively large matrices is a drawback of this method[202]. Expressions for the QFIM for multi-mode Gaussian states were reported by Šafránek, Lee, and Fuentes[203]. Banchi, Braunstein

and Pirandola also derived analytical forms for the fidelity between two arbitrary Gaussian states[46], which can be related to the QFI via the quantum fidelity introduced in Eq. (81).

A unification of these results and a resolution of the problematic divergence behaviour of the QFIM for pure states has been addressed[204]. Specifically, by applying the regularisation procedure R to expressions of the QFIM valid for mixed Gaussian states only, the QFIM for Gaussian states with any number of pure modes can be immediately determined. Defining $I_Q[\rho_G;\{\boldsymbol{\lambda},\boldsymbol{\Sigma}\}]$ as the QFIM for a mixed Gaussian state with mean $\boldsymbol{\lambda}$ and covariance $\boldsymbol{\Sigma}$, the regularisation procedure R for some regularisation parameter $\xi > 1$ proceeds according to

$$\mathsf{R}: \lim_{\xi\to 1} I_Q\left[\rho'_G;\{\boldsymbol{\lambda},\xi\boldsymbol{\Sigma}\}\right]. \tag{244}$$

The resulting value is the correct QFIM for any Gaussian state and does not suffer from divergences. A similar regularisation has been also been used to regularise the QFI for non-Gaussian states[49]. Marian and Marian gave a full analysis of the QFI for two-mode Gaussian states[205,206].

Due to the structure of Gaussian states, the SLD is at most quadratic in the bosonic (or equivalently the canonical) mode operators. Hence, for the SLD of parameter θ_k we make the following Ansatz:

$$L_k = \frac{1}{2}A^{(k)}_{\alpha\beta}\left\{\hat{\boldsymbol{a}}^\alpha,\hat{\boldsymbol{a}}^\beta\right\} + B^{(k)}_\alpha\hat{\boldsymbol{a}}^\alpha - \frac{1}{2}\mathrm{Tr}\left[A^{(k)}\boldsymbol{\Sigma}\right], \tag{245}$$

for the SLD for Gaussian states, where Greek indices imply Einstein summation convention. Notice that the first term in Eq. (245) is composed of quadratic (squared bosonic operators) and bilinear terms (product between a creation operator with an annihilation operator). The quadratic contribution describes active devices such as squeezers and down-converters, while the bilinear contributions to the first term corresponds to linear unitary operations such as beam-splitters, half and quarter wave plates, and phase-shifters. The second term in Eq. (245) has a linear dependence on the operators. These operations describe displacements in phase space. The coefficients $A^{(k)}$ and $B^{(k)}$ are completely determined from the first and second moments of the probe state and their derivatives. Finally, we include a constant term $-\frac{1}{2}\mathrm{Tr}[A^{(k)}\boldsymbol{\Sigma}]$.

Monras[200] found a closed form for these coefficients as a solution to the Stein-equation

$$A^{(k)} = \sum_{j=0}^{\infty}\boldsymbol{F}^{\top j}\partial_k(\boldsymbol{\Sigma}^{-1})\boldsymbol{F}^j, \quad B^{(k)} = 2\boldsymbol{\Sigma}^{-1}(\partial_k\boldsymbol{\lambda}), \tag{246}$$

with $\boldsymbol{F} = (i\boldsymbol{\Sigma}\boldsymbol{\sigma}_y)^{-1}$. Notice that the constant $A^{(k)}$ is defined only for non-singular states, which excludes pure states. The QFIM then reads

$$[I_Q]_{jk} = \frac{1}{2}\frac{\bar{n}^2}{1+\bar{n}^2}\mathrm{Tr}\left[(\partial_j\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\partial_k\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right] + 2(\partial_j\boldsymbol{\lambda})^\top\boldsymbol{\Sigma}^{-1}(\partial_k\boldsymbol{\lambda}), \tag{247}$$

where $\bar{n}$ is the average photon number of the thermal states before symplectic transformations. The QFI in Eq. (247) is not valid for all mixed states, but only for 'isotropic' or 'isothermal' states that have the same symplectic eigenvalues for all modes, which includes the class of pure state models. An elegant form for the QFIM for arbitrary mixed Gaussian states was derived by Šafránek[204], from which the origin of precision enhancements were assigned to three qualitatively different terms: changes to the orientation and squeezing of the Gaussian state, changes in purity, and changes in displacement.

Independently, Gao and Lee used phase space methods to derive an exact form of the SLD and the QFI[202]. Their use of a quadratic Ansatz for the SLD lead to the result

$$L_j = \frac{1}{2}\mathfrak{M}^{-1}_{\gamma\kappa,\alpha\beta}\left(\partial_j\boldsymbol{\Sigma}^{\alpha\beta}\right)\left(\hat{\boldsymbol{a}}^\gamma\hat{\boldsymbol{a}}^\kappa - \boldsymbol{\Sigma}^{\gamma\kappa}\right) + \boldsymbol{\Sigma}^{-1}_{\mu\nu}\left(\partial_j\boldsymbol{\lambda}^\nu\right)\hat{\boldsymbol{a}}^\mu, \tag{248}$$

where $\mathfrak{M} = \boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma} + \boldsymbol{\Omega}\otimes\boldsymbol{\Omega}/4$. The QFI for any n-mode bosonic Gaussian system was then determined through application of Wicks theorem for Gaussian states to yield

$$\left[I_Q\right]_{jk} = \frac{1}{2}\mathfrak{M}^{-1}_{\alpha\beta,\mu\nu}\partial_j\boldsymbol{\Sigma}^{\alpha\beta}\partial_k\boldsymbol{\Sigma}^{\mu\nu} + \boldsymbol{\Sigma}^{-1}_{\mu\nu}\partial_j\boldsymbol{\lambda}^\mu\partial_k\boldsymbol{\lambda}^\nu. \tag{249}$$

Calculating the QFI increases exponentially in computational time with the number of modes owing to the inversion of large dimensional matrices. This result is limited to mixed Gaussian states.

A closed form of the SLD has recently also been derived using the quantum fidelity between two arbitrary Gaussian states[207]. The optimal measurement described by these SLDs does not always correspond to Gaussian measurements[208].

B. Hamiltonians with non-multiplicative factors

So far we have considered the estimation precision of a given probe into the dynamics that imprints the parameter θ. This is a channel estimation scheme, where an optimisation over the possible input probe states can improve the estimation precision[209]. Within this regime, we have reviewed parameter estimation schemes where the Hamiltonian is either a constant or has multiplicative dependence on the parameter of interest (e.g., a coupling strength). This regime is known as the phase-shift or phase-like Hamiltonian estimation, where the parameter to estimate multiplies a parameter-independent Hermitian generator G[30,210]. Under the unitary evolution $U = \exp[-iH]$, the generator of parameter translation in θ is formally defined through

$$G = i\left(\partial_\theta U\right)U^\dagger. \tag{250}$$

The phase-like Hamiltonian regime, where $U = \exp(-i\theta G)$, has been extensively studied, and these single parameter phase-shift Hamiltonians define the optimal probe state[71] as discussed in subsection IV H. It has been applied for Hamiltonian characterisation in the absence of noise [211–214], and frequency measurements and atomic spectroscopy in the presence of decoherence[7,35].

More general Hamiltonians have recently started to attract attention[215], since they permit the application of quantum metrology to a more general class of problems such as time-varying fields[216,217] and in gradient magnetometry[218]. It is

well known that a pure state parameterised by a multiplicative Hamiltonian of the form $H_j = \theta_j G_j$ for a time t, the QFI is given by[30,125]

$$I_Q(\theta_j) = 4t^2(\Delta G_j)^2. \tag{251}$$

The definition of the generator when the state is parameterised by a more general Hamiltonian for the form $H(\theta)$ becomes unclear. Assuming the Hamiltonian has n unique eigenvalues E_j with $j = 1, \ldots, n$, degeneracy d_j and corresponding eigenvectors $|E_j^{(k)}\rangle$, $k \in \{1, \ldots, d_j\}$ satisfying $\langle E_\alpha^{(\beta)}|E_\gamma^{(\delta)}\rangle = \delta_{\alpha\gamma}\delta_{\beta\delta}$, then the generator of translations in θ_j can be written as[125,219,220]

$$G_j(\boldsymbol{\theta}) = \sum_{k=1}^{n_g} \partial_j E_k P_k + 2\sum_{k\neq l}\sum_{m=1}^{d_k}\sum_{n=1}^{d_l} \exp\left[-i(E_k - E_l)/2\right] \times \sin\left[\frac{E_k - E_l}{2}\right] \left\langle E_l^{(n)}\middle|\partial_j E_k^{(m)}\right\rangle \left|E_k^{(m)}\right\rangle \left\langle E_l^{(n)}\right|, \tag{252}$$

where $P_k = \sum_j |E_k^{(j)}\rangle \langle E_k^{(j)}|$ is the projector onto the E_k-eigenspace. The form of the generator in Eq. (252) implies that the QFI can be separated into two parts. The first part is due to the dependence of the eigenvalues on θ_j, and the second due to the dependence of the eigenstates on θ_j. An upper bound on the QFI was derived by Pang and Brun[125]

$$I_Q^{\max} \leq 2t^2 \sum_{j=1}^{n} d_j(\partial_j E_j)^2 + 8\sum_{k\neq l}\sum_{m=1}^{d_k}\sum_{n=1}^{d_l} \left|\sin\left[\frac{(E_k - E_l)t}{2}\right]\right|^2 \left|\left\langle E_l^{(n)}|\partial_j E_k^{(m)}\right\rangle\right|^2. \tag{253}$$

We observe that the QFI is finite for all time t, which is in contrast to the QFI for multiplicative factor in Eq. (251). Increasing the channel QFI can be achieved by enhancing the sensitivity of the generator through additional terms in the Hamiltonian[209]. Specifically, for time evolutions with parameter independent eigenvalues, Eq. (252) exhibits a periodic time dependence of the channel QFI[125]. Generally, this alone does not saturate the Heisenberg limit precision, but can with feedback controls[161,221]. In section VII, we see how the precision achievable in channel estimation schemes can be improved with ancilla states.

The general result in Eq. (252) implies that for any unitary encoding, the optimal probe states can be assigned through the prescription described by Giovannetti *et al.*[71]. However, this method requires that the spectrum for the Hamiltonian is known, and this is generally difficult. This complication can be avoided if the non-multiplicative estimation problem can be related to an equivalent, phase-like estimation problem through re-parameterisation of the parameters $\boldsymbol{\theta}$, since the QFI matrix elements transforms as a metric tensor[222], as we have seen in section V A. If a nonlinear function of the parameters can be transformed into a phase-like scheme this approach may simplify the calculation of the QFI. We will return to functions of multiple parameters in section VIII B.

VII. NOISY QUANTUM METROLOGY

We have seen that quantum resources can improve sensing capabilities over classical methods. However, very few systems can be prepared in complete isolation from its environment in practice. Therefore, applications of quantum metrology must include a description of the interaction of a system with its environment as illustrated in figure 7. Generally, the performance of sensors are limited by decoherence in a familiar fashion to most applications of quantum technologies.

A. Metrology in noisy quantum channels

Various methods to calculate parameter precision bounds in the presence of noise have been developed. A common approach is to use the Kraus decomposition of the quantum channel[35,54,57]. The completely positive, trace preserving (CPTP) channel $\rho(\theta) = \Lambda[\rho(0)](\theta)$, satisfying $\Lambda[\rho(0)](\theta) \geq 0$ and $\mathrm{Tr}[\rho(\theta)] = \mathrm{Tr}[\rho(0)]$, can be expanded in terms of the Kraus operators[223,224]

$$\rho(\theta) = \Lambda\left[\rho(0)\right] = \sum_{j=1}^{q} K_j(\theta)\rho(0)K_j^\dagger(\theta). \tag{254}$$

The set of q Kraus operators $\mathscr{K}(\theta) = \{K_j(\theta)\}$ is referred to as a q-Kraus decomposition of Λ, satisfying $\sum_{j=1}^{q} K_j^\dagger(\theta)K_j(\theta) = \mathbb{1}$. This representation is not unique, since for every q-Kraus decomposition of a channel, all the other q-Kraus decompositions can be constructed via a unitary transformation

$$\sum_k u_{jk}(\theta)K_j(\theta),$$

where $u_{ij}(\theta)$ is a unitary matrix. The set of all q-Kraus decompositions of a channel is called the q-Kraus ensemble and is noted $\mathscr{K}_q(\theta)$. The smallest possible number q of Kraus operators is known as the Kraus rank. It can be obtained as the number of non-vanishing eigenvalues of the Choi-matrix of the channel[23]. The generality of this representation is that it can be used to model any system decoherence. Specifically, adding arbitrary environmental degrees of freedom to the probe system, the probe dynamics can be written as a unitary evolution for the combined probe-bath system. This process is not unique since it requires the environment to be defined; the probe evolution depends on the unitary evolution associated with the enlarged Hilbert space. A fixed choice of the unitary evolution and environmental degrees of freedom corresponds to choosing a Kraus operator $K_j(\theta)$, which leads to an upper bound for the QFI[35]

$$I_{Q,\max}\left(\rho(0); K_j(\theta)\right) \geq 4\left\{\langle H_1(\theta)\rangle - \langle H_2(\theta)\rangle^2\right\}, \tag{255}$$

where

$$H_1(\theta) = \sum_j \frac{\mathrm{d}K_j^\dagger(\theta)}{\mathrm{d}\theta}\frac{\mathrm{d}K_j(\theta)}{\mathrm{d}\theta}, \tag{256}$$

$$H_2(\theta) = i\sum_j \frac{\mathrm{d}K_j^\dagger(\theta)}{\mathrm{d}\theta}K_j(\theta), \tag{257}$$

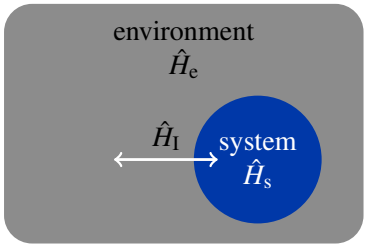


FIG. 7. System-bath interaction model where H_s, H_e, and H_I are the system, environment, and interaction Hamiltonians.

and the expectation is with respect to the initial probe state. For unitary processes, Eq. (255) reduces to the common generator variance bound. The Kraus decomposition can be applied to ancilla assisted schemes, where the bound can be similarly determined from the reduced density matrix of the probe. Determining the QFI requires optimising Eq. (255) over all Kraus operators, which is generally difficult. Alternative methods have been introduced to determine error bounds in noisy systems. This includes using a variational approach[55], a generalised Bures angle approach to quantum channels[225], and through the geometry of quantum channels, which has been used to explore the effects depolarisation, dephasing, spontaneous emission and photon loss channels[57,226].

For non-unitary probe encoding, few analytic expressions for the QFI exist[4,213,227] given that the calculation involves complex optimisation procedures which becomes increasingly difficult with increasing system size. In this case, it is more informative to report on asymptotic lower bounds to the estimation precision for specific noise models. For the estimation of some parameter θ, the error in its estimate depends on the initial probe state. Noise generally degrades the optimal quadratic enhancement of precision estimates to one that is a constant improvement over classical methods[226]. In fact, it has been demonstrated that for parallel Markovian dephasing noise in the probe preparation, the precision scaling is reduced to the classical SQL $\Delta\theta \sim 1/\sqrt{N}$ scaling[7,35,226], and transverse Markovian dephasing noise reduces the scaling bound to $\Delta\theta \sim 1/N^{5/6}$ if an entangled probe is used[228]. For the entangled GHZ probe, these scalings were generalised to Markovian noise in nonlinear quantum metrology with many-body open systems by Beau and del Campo[229]. Specifically, for a k-dimensional Hamiltonian and p-dimensional Lindblad operator, the variance of a Hamiltonian parameter scales as $N^{-[k-(p/2)]}$, which surpasses the shot-noise limit for $2k > p+1$. The system-environment coupling parameter can be estimated with a precision that scales as $N^{-(p/2)}$, while many-body decoherence enhances the precision to N^{-k} in the noise-amplitude estimation of a fluctuating k-body Hamiltonian. For non-Markovian noise, the precision bound has a $\Delta\theta \sim 1/N^{3/4}$ scaling[56,230]. Although these reduced precision scalings lower the prospects of quantum metrology and sensing, some gain has been demonstrated with non-Markovian dephasing and probe readout schemes[231]. We also review alternative methods to attempt to diminish the effects of noise later in this section.

Most of these established precision bounds are based on educated guesses or numerical methods on the integrated form of the probe dynamics. For non-unitary dynamics, where extra terms have to be added to fully describe the dynamics of the probe state, even the integration is not straightforward. The specific dynamics are described by the Lindblad equation[232]

$$\partial_t \rho = -i[\rho, H(\theta)] + \sum_j \mathsf{L}_j \rho \mathsf{L}_j^\dagger - \frac{1}{2}\{\rho, \mathsf{L}_j^\dagger \mathsf{L}_j\}, \tag{258}$$

where L_j denotes the Lindblad operators. Eq. (258) is often referred to as the Master Equation. One method to obtain general precision bounds from Eq. (258) is through numerical methods, which does not match the insight provided by analytical approaches. Fortunately, this limitation has been addressed by Sekatski *et al.*, who developed a framework to provide analytic results for general decoherence models described through the master equation[233]. This was followed by a more qualitative approach that addresses the effects of arbitrary Markovian dynamics on the precision scaling of unitary parameter estimation problems. Specifically, if the generator of translations G in θ can be written in terms of the operators[234,235]

$$\left\{\mathbb{1}, \mathsf{L}_j^{\mathrm{H}}, i\mathsf{L}_j^{\mathrm{AH}}, (\mathsf{L}_j^\dagger \mathsf{L}_j)^{\mathrm{H}}, i(\mathsf{L}_j^\dagger \mathsf{L}_j)^{\mathrm{AH}}\right\}, \tag{259}$$

where the superscript 'H' and 'AH' denote the Hermitian and the anti-Hermitian part of an operator respectively, then for some interrogation time T, the precision scales as at most

$$\Delta\theta \geq \frac{\alpha}{\sqrt{T}}, \tag{260}$$

where α is a real constant. Equivalently, the precision scales no better than $1/\sqrt{N}$ with N quantifying the resources used. This method does not depend on the integrated form of the master equation, but only on the geometric properties on the Lindblad operators. Furthermore, if the generator is not in the span of the Lindblad operators then the Heisenberg scaling can be recovered with application of quantum error correction procedures. This was realised experimentally by Kessler *et al.*[236], Dür *et al.*[237], and Sekatski *et al.*[233]. We will review this in further detail in subsection VII C.

One approach to quantify and understand the impact of decoherence on parameter estimates is to have access to the environmental degrees of freedom to make estimates of both Hamiltonian and bath-coupling parameters. Monras and Paris showed that for Gaussian probes, this approach makes no improvement[227] in that there is always an environment such that observing the combined probe-environment state does not provide additional information gain when compared to simply observing the probe itself[35]. For small losses, the QFI scales with the loss parameter itself, which demonstrates a qualitative improvement over the shot-noise limit.

The impact of loss on optical interferometry has been explored in detail in the field. We observed in subsection IV F that the highly entangled NOON state saturates the Heisenberg limit on precision for phase measurements. However, this class of states is extremely susceptible to losses, and are outperformed by purely classical states for moderate losses[7,238–240]. Although for lossy interferometry the Heisenberg limit is not

attainable[120,241], certain quantum states have been engineered to outperform both standard and NOON states for phase measurements in the presence of losses[242]. Note that this strategy differs from quantum error correction techniques used for quantum computing and protecting quantum memories, where the approach is to protect the information encoded in light[243]. This demonstrates that a quantum enhancement is possible even in the presence of decoherences. Given that a diverse range of different physical quantities can be determined through phase measurements, it is important to place general bounds for phase estimation with lossy optical interferometers. A systematic approach to address the optimal states with definite photon numbers for interferometry in the presence of losses was first considered by Dorner *et al.*[213]. This idea was later extended to find the optimal input state for two-mode interferometry using a numerical optimisation of the QFI[4]. Due to the convexity of the QFI that we observed in subsection IV B, miniminising the estimation error is a convex optimisation problem. The corresponding precision beats classical methods and lies between the SQL and the Heisenberg limit, depending on the loss rates. It can not be improved by considering probes with indefinite photon number and states with photons distributed between distinguishable time bins.

Besides interferometry, further applications of noisy metrology have been considered. For example, the effects of phase flip and amplitude damping decoherence channels with N-qubit GHZ probes have been explored. As we observed in subsection IV F, in the absence of noise, the Heisenberg limit can be achieved through rotations along the Z direction. For a phase noise channel, the QFI decreases with increasing decoherence rate due to information leak to the environment, although it can exhibit revivals[244]. For an amplitude damping channel, the QFI can be enhanced by adjusting the temperature of the environment. In bosonic quantum metrology, a constant scaling improvement over the SQL was observed by Spedalieri *et al.*, who demonstrated that correlated thermal states outperform coherent states for the estimation of a loss parameter[245]. Also, the case of simultaneous estimation of multi-parameters in the presence of noise has been investigated[102,246]. Since preparing noiseless probe states and encoding channels is experimentally challenging, efforts to view noise as a utility to introduce correlations into the system have been explored[247–249]. These correlations can lead to entanglement[250], which may provide precision enhancements. Hence decoherences can be used to protect precisions against noise. However, too much noise is detrimental. Understanding the interplay between relaxing the noiseless criterion for ease of experimental realisation and the amount of useful noise in the system remains an open question.

B. Ancilla-assisted schemes and channel estimation

Assuming that noise cannot be eliminated from the system, a natural question to ask is whether it is possible to negate its effects. One strategy is to use ancillary systems, and allow the probe system to arbitrarily interact with the ancilla. There are two ways in which an ancillary state can be introduced to the state. In the first, the ancilla state is entangled with the probe but does not itself participate in the estimation[59]. This is known as channel extension and is illustrated in Fig. 8. The objective is to consider how changes to the dynamics that imprints the parameters to the probe state improves or limits the estimation precision. Given that quantum channels are completely positive trace preserving maps, channel extension can be described through

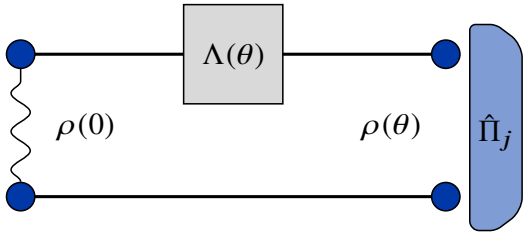


FIG. 8. Ancilla assisted quantum channel extension scheme, described by $\Lambda(\theta) \otimes \mathbb{1}$ and a POVM measurement Π_j.

$$\Lambda(\theta) \rightarrow \Lambda(\theta) \otimes A, \tag{261}$$

where A is an arbitrary channel acting on an ancilla system. The channel quantum Fisher information $I_{Q,\mathrm{ch}}(\Lambda;\theta)$ for $\Lambda(\theta)$ is defined as

$$I_{Q,\mathrm{ch}}(\Lambda;\theta) = \max_{\rho} I_Q\left(\Lambda\left[\rho\right];\theta\right). \tag{262}$$

Using the monotonicity of the QFI, it is enough to consider extensions by the identity[251]. This identity channel extension has been demonstrated to generate increased estimation precisions[252,253]. The underlying principle is to enlarge the Hilbert space of the probing system, such that noisy components can be readily separated from the signal. It has been shown that for phase estimation, the ancilla is useful for arbitrary values of the noise parameter[254,255]. This has recently been demonstrated experimentally[256,257].

The second method of extending the Hilbert space of the system through ancillary states is the Hamiltonian extension scheme and involves adding an operator to the Hamiltonian. This is a more fundamental approach, since the Hamiltonian describes the probe dynamics during the parameterisation process. It differs from channel extension since it can include an interaction term between the probe and the ancilla. This provides a general framework for the treatment of open quantum systems. However, it has been shown that this approach does not improve the sensitivity of phase shift-type measurements when considering the quantum Fisher information optimised over input states[209]. Several experiments have demonstrated the advantage of entanglement-assisted schemes in other applications including channel tomography[258], and channel capacity estimation[259].

The precision limit for more general parameter estimation schemes—including adaptive protocols—can be addressed by defining the encoding channel $\Lambda(\theta)$. Adaptive channel estimation has been explored by Pirandola and Lupo[260], Takeoka and Wilde[261], and was further developed by Laurenza *et al.*[262]. The QFI for the adaptive estimation of a noise parameter θ in

a channel that obeys so-called joint teleportation covariance is determined through[260]

$$I_{Q,\mathrm{ch}}(\Lambda;\theta) = \frac{8\left[1 - F\left(\rho_{\Lambda_\theta}, \rho_{\Lambda_{\theta+\mathrm{d}\theta}}\right)\right]}{\mathrm{d}\theta^2}, \tag{263}$$

where teleportation covariance is given by[262,263]:

$$\Lambda(\theta)\left[U\rho U^\dagger\right] = V\Lambda(\theta)[\rho]V^\dagger. \tag{264}$$

Note that V is not necessarily equal to U, and does not depend on θ. Such channels include the set of thermal-loss[264], Pauli[265], erasure[266], and noisy amplifier channels.

It is worth noting that studies of channel estimation and channel discrimination are inextricably linked. For the prototypical quantum channel $\Lambda(\theta)$, the former scheme corresponds to continuous x, where the task is to estimate the unknown parameterisation. Channel discrimination corresponds to discrete θ, where it is more meaningful to distinguish the quantum channels. The symmetric discrimination between two arbitrary qudit channels was considered by Pirandola *et. al* using a port-based teleportation scheme for channel simulation[267]. Here, the same technique was used to derive limits to adaptive quantum metrology[262]. Specifically, any adaptive protocol of channel estimation is bounded in terms of the channel's Choi matrix and saturates the HL in the number of probings.

Through bounds to state discrimination, one can immediately determine informative bounds for a range of other tasks. This has been illustrated in quantum communication and key distribution[268], estimation theory[269], quantum illumination tasks[270], and quantum hypothesis testing[261,267].

C. Fault-tolerant metrology

Tools from quantum error correction[271] can also be used to enhance and potentially recover estimation precisions in the Heisenberg scaling regime for sensing under Markovian noise. These methods are often used in conjunction with, or in addition to, ancilla-assisted schemes[235–237,272] and are referred to as fault tolerant quantum metrology. Their purpose is to correct errors in the probe state and/or the measurement procedure against noise while permitting the signal to be encoded on the sensor.

Quantum error correction (QEC) has found widespread application in quantum information processing and has evolved into a selfcontained and well developed field of active research[273,274]. However, migrating these tools to implement fault tolerant quantum metrology is not straightforward. The characteristic difficulty is in designing QEC codes that detect and correct the noise in the sensor, while leaving the encoded signal intact. This constraint is uncommon in quantum computing applications and consequently many existing QEC codes are not suitable for quantum sensing applications[275]. It is possible to identify QEC codes that can be used to protect quantum sensing protocols against noise, based on certain criteria. First, if the signal Hamiltonian and the error operators commute, it has been shown that the application of error correction can enhance the sensing precision[236,237,272]. Quantum sensing in this regime has been demonstrated using nitrogen-vacancy centers[236], trapped ions systems[276], and has been experimentally demonstrated for field sensing[277] without the use of feedback control methods. Second, provided the generator exists outside the span of the Lindblad operators and assuming access to noiseless ancillas, there exists a quantum error correction protocol that protects the sensor against noise[233–235]. This has been extended to more realistic settings, e.g., where the signal and noise are in the same direction[278], and where noiseless ancilla states were shown to be unnecessary when the signal Hamiltonian and the error operators commute[275].

The overheads necessary to realise fault-tolerant metrology schemes make them unfeasible with current technology. However, these methods are gaining increasing attention, and the next goal is to determine realisable fault tolerant quantum metrology protocols.

VIII. DISTRIBUTED QUANTUM SENSING

A natural extension to multi-parameter quantum estimation is to distribute the parameters over different sensors and use multiple probe systems to interrogate the whole system. This networked approach distributes sensing protocols over multiple resources. Distributed quantum sensing has recently gained increased attention since it provides a promising platform to address a wider class of applications and compare existing results. For example, this framework can describe distributed interferometric phase sensing[279], estimate the properties of a multi-dimensional field where each sensor is an ensemble of atoms[280], and calibrating continuous-variable quantum key distribution networks[281]. Architectures based on a network of sensors have also been reported to realise the quantum internet[282], scalable quantum computing[283], and national[284] and international[285] quantum cryptography. Proctor *et al.* developed a general framework for the treatment of photonic and atomic sensor networks[286]. This approach is also important since it helps to understand the optimal strategies that saturate the QCRB.

In section IV F we reviewed work that demonstrated the necessity of entangled probe states to facilitate HL precision scalings. However, in section IV G we saw recent work that demonstrate non-entangling strategies to achieve the same scaling as entangled probes[105]. Beyond a few counter examples that have reported too much entanglement as detrimental[99], little research to date has addressed the question when and how much entanglement is necessary as a resource in order to saturate fundamental quantum precision bounds[151]. Distributed quantum sensing holds the promise to demonstrate the utility of entanglement and elucidate its contribution in providing precision enhancements. A separate question that shares a similar dilemma is the choice between simultaneous and sequential estimation strategies. These two strategies generate different precision scalings with respect to the resource count[287–289] and time[217,221,231]. However, no clear results have been found that determine when either strategy should be preferred. In this section, we review efforts in distributed quantum sensing to address these two open questions.

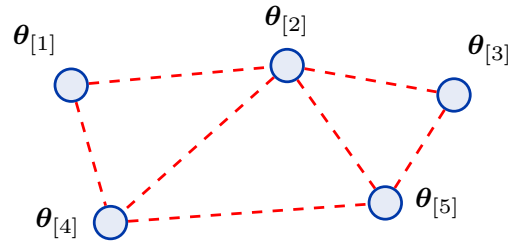


FIG. 9. Networked approach to quantum sensing for a decomposition of the parameters across 5 sensors with potentially entangled states (dashed lines). The j^{th} node is encoded by the parameter subset $\boldsymbol{\theta}_{[j]}$. The dashed lines represent (quantum) correlations between the nodes. This approach is well suited to field sensing, where each node is a spin system.

A. Framework for distributed sensing

To formally define distributed sensing, consider a quantum system comprised of n subsystems, labelled through an index $j = 1, 2, \ldots, n$, with an associated Hilbert space $\mathscr{H}_j$. The Hilbert space associated with two different subsystems are not required to be identical. Additionally, each space is not limited to a single field mode. The total Hilbert space is $\mathscr{H} = \otimes_{j=1}^{n} \mathscr{H}_j$. This product structure of the total system Hilbert space describes a spatially distributed sensor network as illustrated in figure 9. The physical system associated with each Hilbert space $\mathscr{H}_j$ is referred to as a quantum sensor, and the collection of all sensors a quantum sensing network.

Similar to single probe states, the parameters can be imprinted on the quantum network via unitary evolutions[280] and more general dynamics[290]. A unitary evolution of the total system is written as $U(\boldsymbol{\theta}) = \exp[-i\boldsymbol{H}\cdot\boldsymbol{\theta}]$ where the vector of Hermitian operators $\boldsymbol{H} = (H_1, \ldots, H_D)$ is the system Hamiltonian, and each node j has d_j parameters unitarily encoded such that $\sum_j d_j = D$. The j^{th} Hamiltonian element $H_j = h_j \otimes_{k\neq j} \mathbb{1}_k$ acts non-trivially on only one of the Hilbert spaces in $\mathscr{H}$. From this, the product nature of the unitary is explicit:

$$U(\boldsymbol{\theta}) = \bigotimes_{j=1}^{n} U_j(\boldsymbol{\theta}_{[j]}), \tag{265}$$

where $[j]$ indicates the set of d_j parameters at node j. Any system with a tensor product structure can be reconstructed in the language of quantum networks.

To understand how distributed sensing framework can be used to evaluate the role of entanglement, it is necessary to introduce local and global networked sensing strategies. For a given decomposition of a network into different sensors, an estimation procedure is local if[103] (1) the input probe state is separable with respect to the network of quantum sensors, and (2) the measurement and the construction of the estimator can be implemented with only local operations and classical communication (LOCC), along with local (classical) computations. A global estimation strategy is defined as a non-local estimation procedure, where either a probe that is entangled over the quantum sensors and/or a measurement requiring non-local quantum operations is used. Note that in local strategies, each parameter is estimated individually, while in global strategies parameters are estimated simultaneously. Local estimation strategies are more robust against local estimation failure than global strategies and often involve more realistic methods of state preparation[291], measurement, and control[292].

The use of correlations between different sensors in the network can provide precision enhancements beyond unentangled systems[281]. However, this is true only in specific scenarios. When each quantum sensor is used to estimate a single parameter, entanglement between different sensors has a detrimental impact on the estimation precision. Intuitively, when the parameters are encoded locally on each sensor, it does not make sense to use global states or measurements that exhibit entanglement between the sensors for improved precisions. Hence, under the single parameter networked sensing remit, separable state and measurements over the sensors are preferred[280]. This reflects the results of simultaneous estimation in optical multi-parameter estimation problems[100–102]. It is also consistent with the results in[99] where too much entanglement was reported as being detrimental, and the calibration of optical quantum gyroscopes, where the optimal level of entanglement depends on the parameters[151]. If instead each sensor is used to estimate multiple parameters, the effect of entanglement on the estimation precision depends on properties of each of the parameter generators[280]. If all the generators corresponding to the parameters commute, then entanglement between sensors is still detrimental to the estimation precision. In contrast, if the generators do not all commute then entanglement between sensors may in some cases give a small constant reduction in the estimation uncertainty. However, this advantage diminishes when each sensor is coupled to an ancillary system. Practically, this is important for examples where the resources used must be limited to avoid damaging a sample whose properties are being probed[293–295].

The original scheme for simultaneous quantum-enhanced phase estimation by Humphreys *et al.* was realised using highly entangled NOON states. However, the same precision enhancements exhibited by the global strategies can be obtained with mode-separable states and local measurements alone[103]. This is not surprising since in multimode optical systems with commuting phase generators, the crucial resource for enhanced metrology is a large particle number variance within each mode, which can be obtained without multimode entanglement.

B. Estimating functions of parameters

So far we have considered the multi-parameter estimation of a vector of parameters: for a well-defined decomposition of a network of quantum sensors, different subsets $\boldsymbol{\theta}_{[j]}$ of the parameters are locally encoded onto the space of the j^{th} sensor. An alternative approach to estimating the vector of parameters is to estimate a function of these elements instead. This is better suited for many practical applications and can be more informative than estimating the parameters. Immediate exam-

ples of this include vectorial magnetometry for imaging using a network of atomic sensors[296,297], the realisation of a global quantum network of clocks[298], and field interpolation[299]. Field interpolation exemplifies the advantage of estimating a function of parameters since the the objective is to estimate the field at a spatial region not covered by a sensor within the network.

Determining precision bounds for the estimation of functions requires a strategy that minimises estimate variances of general functions of $\boldsymbol{\theta}$. One can construct a D-dimensional vector of parameter functions

$$\boldsymbol{\vartheta} = (f_1(\boldsymbol{\theta}), \ldots, f_D(\boldsymbol{\theta})) \tag{266}$$

where $f_j : \mathbb{R}^D \to \mathbb{R}, j \in \{1, D\}$ are continuously differentiable functions. In section V, we observed that the QFI elements transform as a metric tensor. Hence the single-shot QCRB for an estimator of $\boldsymbol{\vartheta}$ is given by Eq. (154). The attainability of this bound for function estimation has a strict caveat. Extreme care should be taken when estimating a scalar function of multiple parameters $f(\theta_1, \ldots, \theta_D)$. This can not be treated as a single-parameter estimation f, since the mapping of many parameters to one implicitly assumes that the remaining $D-1$ parameters, whatever they are, are known perfectly. This assumption is generally not satisfied and the procedure therefore leads to unattainable bounds[300]. This unattainability has been explored for a network of single qubits by Eldredge *et al.*, where for the specific problem of nanoscale magnetic resonance imaging, GHZ and spin-squeezed states are found to be optimal[301].

Work in this direction has to date been limited to estimating the class of linear functions, where the vector $\boldsymbol{\vartheta}$ is constructed from linear transformations of $\boldsymbol{\theta}$ through $\boldsymbol{\vartheta} = J\boldsymbol{\theta}$. This class of functions is useful for modeling optical and atomic sensing applications, and the estimation of parameter averages. The last application has been considered for a network of precision clocks by Komar *et al.*[298]. To facilitate convenient comparisons between these works, Proctor *et al.* demand each row of the transformation matrix J be a normalised vector[286]. In the case of a single parameter encoded on each sensor, entanglement can help to perform better estimates of linear functions. The precise effect of entanglement depends on the explicit choice of linear function. Denoting the first row of J with the vector $\boldsymbol{v}$ such that $\vartheta_1 = \boldsymbol{v} \cdot \boldsymbol{\theta}$, then Proctor *et al.* find that an entangled probe across the sensors always provides a precision enhancement compared with a separable state, unless the function $\boldsymbol{v} = \|\boldsymbol{v}\|\boldsymbol{e}_k$[280]. Intuitively, if $\boldsymbol{v}$ contains more than one non-zero element, then ϑ_1 describes a global process across multiple sensors and naturally, a globally correlated state will be most sensitive to changes in ϑ. Conversely, if only one element is non-zero, the process is local and entanglement is not required and can indeed be detrimental for the estimation performance.

To conclude, the use of simultaneous estimation schemes and entangled resources are not always necessary to achieve quantum-enhanced sensing. In particular, the use of entanglement can be detrimental and any precision enhancements depend on whether the parameters of interest are local or global properties of a set of systems. As future work, it would be interesting to observe how these results can be extended to the class of non-linear functions. The analogy between the choice of appropriate weight/importance matrices in the class of non-distributed sensing to the freedom in choice of arbitrary functions in distributed sensing can also be explored.

IX. CONCLUSIONS AND OUTLOOK

Without exaggeration, our ability to devise novel approaches for high precision measurements has been one of the greatest drivers of fundamental science and technologies. A plethora of historical examples in optics, communication, computation, imaging, and metrology bear substantial evidence to support this statement. However, we are reaching various limits in metrology, including the shot noise limit and the diffraction limit. The latter directly impacts our ability to keep up with Moore's law in the manufacturing of electronics. Quantum metrology provides a natural extension to classical methods to continue driving performance improvements using quantum systems.

Quantum metrology has matured into a broad field with many active areas of theoretical and experimental research. In this review, we provide a summary of the main techniques in sensing and metrology. We use geometric arguments stemming from information theoretic concepts to motivate the key quantities in classical and quantum estimation theory. This is a powerful approach that provides a method to visualise the estimation process, and helps compare different sensing protocols. It is important to understand what resources are necessary to improve the precision of measurements. Precision bounds derived using geometric arguments can also provide conceptual connections between quantum estimation and general quantum information tasks. We review methods to generate general bounds for qubit and multi-particle systems using query complexity arguments for the estimation procedure.

Central to information geometry is the information matrix, which assigns a quantitative measure on the performance of parameter estimates. We show that the quantum Fisher information (QFIM) is part of a wider family of different information matrices that can be found via the Schwarz inequality. These information matrices can be used to derive bounds for elements of the covariance matrix. For multiple parameters, we review the symmetric logarithmic derivative (SLD) and the right logarithmic derivative QFIM, the Kubo-Mori and Wigner-Yanase skew information matrices. We review how they can be connected via their Riemannian metrics. We provide a comparison of these bounds and discuss their attainability. Specifically, we report that the commonly used QCRB for multiple parameters is often not simultaneously saturable, which adds to the difficulty that optimal measurements often depend on the true, unknown values of the parameters. The Holevo Cramér-Rao bound is then found to be the most informative alternative bound, although it is generally difficult to determine for an arbitrary probe. The specific case of Gaussian state estimation and non phase-like parameters is considered. Numerical methods to solve the QCRB have also been discussed. Although we focus mostly on unbiased Fisher estimation schemes, we comment on the effect that biased estimators have on the estimation

precision. Recently, Rubio and Dunningham showed how to construct a Bayesian analogue of the QFIM. In this review, we independently provide a geometric derivation of the Bayes information to arrive at a consistent result. We show that this procedure follows the same method as the Fisher estimation scheme using the prior-weighted density operator and the posterior mean operator to define a corresponding symmetric logarithmic posterior mean operator.

The prevailing thought of entanglement as a necessary resource for enhanced quantum sensing has been challenged, and several works attribute the precision enhancement instead to quantum correlations. Further, the optimal amount of entanglement in the probe state depends on the parameters and the estimation scheme. This demonstrates that entanglement provides limited use in certain cases. This could be seen as welcoming news, given the difficultly to produce robust states with high entanglement. We review efforts that explore non-entangling strategies to saturate the QCRB.

We conclude by summarising key results in current research efforts in noisy metrology schemes. A longstanding impediment of noisy quantum metrology is the reduction of quantum enhancements to constant factor improvements over classical precision bounds. This has often reduced expectations of practical applications of quantum sensing. The use of fault-tolerant schemes to reinstate quantum scaling has renewed efforts to explore the most general bounds in noisy quantum metrology with error correction. This strategy introduces its own unique difficulties: first, well-known methods in quantum error correction for quantum computing cannot be straightforwardly implemented for use in quantum metrology since they would also "correct" the signal we want to measure. Second, the physical overheads required make most uses currently impractical. We review research that addresses these issues. Another current area of intense research is distributed quantum sensing. This field demonstrates how methods in quantum estimation theory are often interwoven with related efforts in quantum communication. We summarise our current understanding of the most efficient methods to distribute resources over different nodes in a network of sensors.

The progress of practical applications presented in this review highlights quantum sensing as a frontrunner to the emergence of quantum technologies. However, developing realisable protocols remains a non-trivial problem with many open questions.

ACKNOWLEDGMENTS

The authors would like to thank Jonathan A. Gross, Rafal Demkowicz-Dobrzanski, and Jesús Rubio Jimenez for useful discussions. We would also like the thank the following for providing valuable feedback and comments on the manuscript: Francesco Albarelli, Cosmo Lupo, Paulina Marian, Changhun Oh, Michal Oszmaniec, Stefano Pirandola, Wojciech Rzadkowski, Dominik Ŝafránek, Claudio Sanavio, Mankei Tsang, Mark Wilde, and Jing Yang. JSS and PK acknowledge EPSRC for funding via the Quantum Communications Hub (EP/M013472/1).

REFERENCES

[1] M. Hayashi, *Asymptotic Theory of Quantum Statistical Inference* (World Scientific Publishing, 2008).
[2] S. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory (v. 1)*, 1st ed. (Prentice Hall, United States, 1993).
[3] G. Tóth and D. Petz, Phys. Rev. A **87**, 032324 (2013).
[4] R. Demkowicz-Dobrzanski, U. Dorner, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Phys. Rev. A **80**, 013825 (2009).
[5] L. Pezzé and A. Smerzi, Proc. Int. School Phys. Enrico Fermi **Course 188**, 691 (2014).
[6] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A **54**, R4649 (1996).
[7] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Phys. Rev. Lett. **79**, 3865 (1997).
[8] B. Luff, J. Wilkinson, J. Piehler, U. Hollenbach, J. Ingenhoff, and N. Fabricius, J. Lightwave Technol. **16**, 583 (1998).
[9] C. Yang, A. Wax, M. S. Hahn, K. Badizadegan, R. R. Dasari, and M. S. Feld, Opt. Lett. **26**, 1271 (2001).
[10] C. C. Gerry and P. L. Knight, *Introductory Quantum Optics*, 1st ed. (Cambridge University Press, United Kingdom, 2004).
[11] J. S. Sidhu, *Quantum Metrology of Grid Deformations and Squeezed Light: With applications in quantum imaging & quantum information*, Doctoral Thesis (University of Sheffield, 2018).
[12] D. McNeish, Struc. Eq. Model. **23**, 750 (2016).
[13] C. W. Helstrom, Int. J. Theor. Phys. **8**, 361 (1973).
[14] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, 1999).
[15] S. Amari, *Information Geometry and Its Applications*, 1st ed., Applied Mathematical Sciences, Vol. 194 (Springer, 2016).
[16] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer, *Information Geometry*, 1st ed. (Springer, 2017).
[17] A. W. van der Vaart, *Asymptotic Statistics* (Cambridge University Press, 2000).
[18] S. D. Gupta, *Selected Papers of C. R. Rao* (Wiley-Blackwell, 1995).
[19] L. Le Cam and G. Lo Yang, *Asymptotics in Statistics: Some Basic Concepts*, 2nd ed., Springer Series in Statistics (Springer, 2000).
[20] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (Wiley-Blackwell, 2007).
[21] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
[22] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[23] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*, 1st ed. (Cambridge University Press, New York, 2008).
[24] P. Kok and B. W. Lovett, *Introduction to Optical Quantum Information Processing*, 1st ed. (Cambridge University Press, Cambridge, UK, 2010).
[25] P. Facchi, R. Kulkarni, V. I. Man'ko, G. Marmo, E. C. G. Sudarshan, and F. Ventriglia, Physics Letters A **374**, 4801 (2010).
[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley Series in Telecommunications and Signal Processing (Wiley-Blackwell, 2006).
[27] C. E. Shannon, The Bell System Technical Journal **27**, 379 (1948).
[28] S. Kullback and R. A. Leibler, Ann. Math. Statist. **22**, 79 (1951).
[29] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press Inc., 1976).
[30] S. L. Braunstein, C. Caves, and G. Milburn, Ann. Phys. **247**, 135 (1995).
[31] J. Hilgevoord and J. Uffink, Found. Phys. **21**, 323 (1991).
[32] H. Nagaoka, *A new approach to Cramér-Rao bounds for quantum state estimation*, IT 89-42 (IEICE Technical Report, 1989) reprinted in [1], Chapter 8, pages 100–112.
[33] G. Tóth and I. Apellaniz, J. Phys. A: Math. Theoret. **47**, 424006 (2014).
[34] S. Yu, "Quantum Fisher Information as the Convex Roof of Variance," (2013), unpublished, arXiv:1302.5311.
[35] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, Nat. Phys. **7**, 406 (2011).
[36] A. Fujiwara and H. Imai, J. Phys. A: Math. Theoret. **41**, 255304 (2008).
[37] M. Hübner, Phys. Lett. A **163**, 239 (1992).

[38] J. Jost, *Riemannian Geometry and Geometric Analysis*, 6th ed. (Springer, 2011).
[39] A. Fujiwara, METR **94-8** (1994).
[40] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs, Vol. 191 (American Mathematical Society, Oxford University Press, 1993).
[41] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).
[42] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).
[43] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, 2010).
[44] H. Yuan and C.-H. F. Fung, New J. Phys. **19**, 113039 (2017).
[45] A. Chefles, Contemp. Phys. **41**, 401 (2000).
[46] L. Banchi, S. L. Braunstein, and S. Pirandola, Phys. Rev. Lett. **115**, 260501 (2015).
[47] C. W. Helstrom, Phys. Lett. A **25**, 101 (1967).
[48] C. W. Helstrom, IEEE Trans. Inf. Theory **14**, 234 (1968).
[49] D. Šafránek, Phys. Rev. A **95**, 052320 (2017).
[50] L. Seveso, F. Albarelli, M. G. Genoni, and M. G. A. Paris, J. Phys. A: Math. Theoret. **53**, 02LT01 (2019).
[51] M. G. A. Paris, Int. J. Quantum Inf. **7**, 125 (2009).
[52] E. Ercolessi and M. Schiavina, Phys. Lett. A **377**, 1996 (2013).
[53] J. Liu, J. Chen, X.-X. Jing, and X. Wang, J. Phys. A: Math. Theoret. **49**, 275302 (2016).
[54] M. Sarovar and G. J. Milburn, J. Phys. A: Math. Gen. **39**, 8487 (2006).
[55] B. M. Escher, L. Davidovich, N. Zagury, and R. L. de Matos Filho, Phys. Rev. Lett. **109**, 190404 (2012).
[56] A. W. Chin, S. F. Huelga, and M. B. Plenio, Phys. Rev. Lett. **109**, 233601 (2012).
[57] J. Kołodyński and R. Demkowicz-Dobrzański, New J. Phys. **15**, 073043 (2013).
[58] S. Alipour, M. Mehboudi, and A. T. Rezakhani, Phys. Rev. Lett. **112**, 120405 (2014).
[59] R. Demkowicz-Dobrzański and L. Maccone, Phys. Rev. Lett. **113**, 250801 (2014).
[60] O. E. Barndorff-Nielsen and R. D. Gill, J. Phys. A: Math. Gen. **33**, 4481 (2000).
[61] A. Fujiwara, J. Phys. A: Math. Gen. **39**, 12489 (2006).
[62] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).
[63] A. Boto, P. Kok, D. Abrams, S. Braunstein, C. Williams, and J. Dowling, Phys. Rev. Lett. **85**, 2733 (2000).
[64] H. Lee, P. Kok, and J. P. Dowling, J. Mod. Opt. **49**, 2325 (2002).
[65] P. Kok, H. Lee, and J. P. Dowling, Phys. Rev. A **65**, 052104 (2002).
[66] P. Walther, J. W. Pan, M. Aspelmeyer, R. Ursin, S. Gasparoni, and A. Zeilinger, Nature **429**, 158 (2004).
[67] M. W. Mitchell, J. S. Lundeen, and A. M. Steinberg, Nature **429**, 161 (2004).
[68] G. J. Pryde and A. G. White, Phys. Rev. A **68**, 052315 (2003).
[69] H. Cable and J. P. Dowling, Phys. Rev. Lett. **99**, 163604 (2007).
[70] D. M. Greenberger, M. A. Horne, and A. Zeilinger, in *Bell's Theorem, Quantum Theory and Conceptions of the Universe* (Springer, Dordrecht, Dordrecht, 1989) pp. 69–72.
[71] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photon. **5**, 222 (2011).
[72] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi, Phys. Rev. A **85**, 022321 (2012).
[73] Ł. Czekaj, A. Przysiężna, M. Horodecki, and P. Horodecki, Phys. Rev. A **92**, 062303 (2015).
[74] G. Tóth, Phys. Rev. A **85**, 195 (2012).
[75] M. Oszmaniec, R. Augusiak, C. Gogolin, J. Kołodyński, A. Acin, and M. Lewenstein, Physical Review X **6**, 0291 (2016).
[76] Y. Ouyang, N. Shettell, and D. Markham, "Robust quantum metrology with explicit symmetric states," (2019), unpublished, arXiv:1908.02378.
[77] Y. Ouyang, Phys. Rev. A **90**, 062317 (2014).
[78] B. Koczor, S. Endo, T. Jones, Y. Matsuzaki, and S. C. Benjamin, "Variational-state quantum metrology," (2019), unpublished, arXiv:1908.08904.
[79] A. Shimizu and T. Morimae, Phys. Rev. Lett. **95**, 090401 (2005).
[80] O. Gühne and G. Toth, Phys. Rep. **474**, 1 (2009).
[81] L. Pezzé and A. Smerzi, Phys. Rev. Lett. **102**, 100401 (2009).
[82] I. Apellaniz, M. Kleinmann, O. Gühne, and G. Tóth, Phys. Rev. A **95**, 032330 (2017).
[83] R. Augusiak, J. Kołodyński, A. Streltsov, M. N. Bera, A. Acin, and M. Lewenstein, Physical Review A **94**, 012339 (2016).
[84] S. Boixo, S. T. Flammia, C. M. Caves, and J. Geremia, Phys. Rev. Lett. **98**, 090401 (2007).
[85] S. M. Roy and S. L. Braunstein, Phys. Rev. Lett. **100**, 220501 (2008).
[86] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Phys. Rev. Lett. **105**, 180402 (2010).
[87] M. Zwierz, C. A. Perez-Delgado, and P. Kok, Phys. Rev. Lett. **107**, 059904 (2011).
[88] M. Zwierz, C. A. Perez-Delgado, and P. Kok, Phys. Rev. A **85**, 042112 (2012).
[89] M. J. W. Hall, D. W. Berry, M. Zwierz, and H. M. Wiseman, Phys. Rev. A **85**, 041802 (2012).
[90] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[91] L. Pezzé and A. Smerzi, Phys. Rev. Lett. **100**, 073601 (2008).
[92] X.-X. Jing, J. Liu, W. Zhong, and X.-G. Wang, Comms. in Theo. Phys. **61**, 115 (2014).
[93] J. Joo, K. Park, H. Jeong, W. J. Munro, K. Nemoto, and T. P. Spiller, Phys. Rev. A **86**, 043828 (2012).
[94] T. Ono and H. F. Hofmann, Phys. Rev. A **81**, 033819 (2010).
[95] J. Joo, W. Munro, and T. Spiller, Phys. Rev. Lett. **107**, 083601 (2011).
[96] Y. M. Zhang, X. W. Li, W. Yang, and G. R. Jin, Phys. Rev. A **88**, 043832 (2013).
[97] H. Wang, M. Mariantoni, R. C. Bialczak, M. Lenander, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, T. Yamamoto, Y. Yin, J. Zhao, J. M. Martinis, and A. N. Cleland, Phys. Rev. Lett. **106**, 060401 (2011).
[98] L. Zhang and K. W. C. Chan, Phys. Rev. A **95**, 032321 (2017).
[99] T. Baumgratz and A. Datta, Phys. Rev. Lett. **116**, 030801 (2016).
[100] P. C. Humphreys, M. Barbieri, A. Datta, and I. A. Walmsley, Phys. Rev. Lett. **111**, 070403 (2013).
[101] J. Liu, X.-M. Lu, Z. Sun, and X. Wang, J. Phys. A: Math. Theoret. **49**, 115302 (2016).
[102] J.-D. Yue, Y.-R. Zhang, and H. Fan, Sci. Rep. **4**, 5933 (2014).
[103] P. A. Knott, T. J. Proctor, A. J. Hayes, J. F. Ralph, P. Kok, and J. A. Dunningham, Phys. Rev. A **94**, 062312 (2016).
[104] J. S. Sidhu and P. Kok, Phys. Rev. A **95**, 063829 (2017).
[105] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola, Rev. Mod. Phys **90**, 035006 (2018).
[106] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, Nat. Photon. **12**, 724 (2018).
[107] H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995).
[108] M. A. Armen, J. K. Au, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Phys. Rev. Lett. **89**, 59 (2002).
[109] D. W. Berry and H. M. Wiseman, Phys. Rev. Lett. **85**, 5098 (2000).
[110] D. W. Berry and H. M. Wiseman, Phys. Rev. A **65**, 043803 (2002).
[111] D. W. Berry and H. M. Wiseman, Phys. Rev. A **73**, 013813 (2006).
[112] D. W. Berry and H. M. Wiseman, Phys. Rev. A **87**, 019901 (2013).
[113] A. Fujiwara, J. Phys. A: Math. Gen. **39**, 12489 (2006).
[114] R. Okamoto, S. Oyama, K. Yamagata, A. F. P. R. A, and 2017, Phys. Rev. Lett. **109**, 130404 (2012).
[115] A. Serafini, ISRN Optics **2012**, 275016 (2012).
[116] P. Palittapongarnpim and B. C. Sanders, Phys. Rev. A **100**, 012106 (2019).
[117] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, Nat. Photon. **7**, 613 (2013).
[118] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature **450**, 393 (2007).
[119] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, New J. Phys. **11**, 073023 (2009).
[120] J. Kołodyński and R. Demkowicz-Dobrzański, Phys. Rev. A **82**, 053804 (2010).
[121] B. Gong and W. Cui, "On the calculation of Fisher information for quantum parameter estimation based on the stochastic master equation," (2017), unpublished, arXiv:1702.08089.
[122] M. D. Vidrighin, G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. Kim, A. Datta, M. Barbieri, and I. A. Walmsley, Nat. Commun. **5**, 3532 (2014).
[123] D. W. Berry, M. Tsang, M. J. W. Hall, and H. M. Wiseman, Phys. Rev. X

5, 031018 (2015).
[124] P. J. D. Crowley, A. Datta, M. Barbieri, and I. A. Walmsley, Phys. Rev. A **89**, 023845 (2014).
[125] S. Pang and T. A. Brun, Phys. Rev. A **90**, 022117 (2014).
[126] Y. Yao, L. Ge, X. Xiao, X. Wang, and C. P. Sun, Phys. Rev. A **90**, 062113 (2014).
[127] J. Zhang and M. Sarovar, Phys. Rev. Lett. **113**, 080401 (2014).
[128] N. Kura and M. Ueda, Phys. Rev. A **97**, 012101 (2018).
[129] H. Imai and A. Fujiwara, J. Phys. A: Math. Theoret. **40**, 4391 (2007).
[130] S. Gammelmark and K. Mølmer, Phys. Rev. Lett. **112**, 170401 (2014).
[131] D. Šafránek, Physical Review A **97**, 042322 (2018).
[132] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).
[133] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, Phys. Rev. A **94**, 052108 (2016).
[134] M. u. u. u. u. Guţă and J. Kahn, Phys. Rev. A **73**, 052108 (2006).
[135] M. Hayashi and K. Matsumoto, J. Math. Phys. **49**, 102101 (2008).
[136] K. Yamagata, A. Fujiwara, and R. D. Gill, Ann. Statist. **41**, 2197 (2013).
[137] Y. Yang, G. Chiribella, and M. Hayashi, Commun. Math. Phys. **368**, 223 (2019).
[138] J. Suzuki, J. Math. Phys. **57**, 042201 (2016).
[139] M. Bradshaw, S. M. Assad, and P. K. Lam, Physics Letters A **381**, 2598 (2017).
[140] M. Bradshaw, P. K. Lam, and S. M. Assad, Phys. Rev. A **97**, 012106 (2018).
[141] F. Albarelli, J. F. Friel, and A. Datta, Phys. Rev. Lett. **123**, 200503 (2019).
[142] J. S. Sidhu, Y. Ouyang, E. T. Campbell, and P. Kok, "Tight bounds on the simultaneous estimation of incompatible parameters," (2019), unpublished, arXiv:quant-ph/1912.09218.
[143] J. Kahn and M. Guta, Commun. Math. Phys. **289**, 597 (2009).
[144] K. Matsumoto, J. Phys. A: Math. Gen. **35**, 3111 (2002).
[145] A. Monras and F. Illuminati, Phys. Rev. A **83**, 012315 (2011).
[146] R. D. Gill and M. Guta, arXiv.org (2012), 1112.2078v2.
[147] C. Vaneph, T. Tufarelli, and M. G. Genoni, Quant. Meas. Quant. Metr. **1**, 12 (2013).
[148] J. Yang, S. Pang, Y. Zhou, and A. N. Jordan, Phys. Rev. A **100**, 032104 (2019).
[149] A. Fujiwara, METR **94-9** (1994).
[150] M. G. Genoni, M. G. A. Paris, G. Adesso, H. Nha, P. L. Knight, and M. S. Kim, Phys. Rev. A **87**, 012107 (2013).
[151] P. Kok, J. Dunningham, and J. F. Ralph, Phys. Rev. A **95**, 012326 (2017).
[152] C. W. Helstrom and R. S. Kennedy, IEEE Trans. Inf. Theory **20**, 16 (1974).
[153] V. P. Belavkin, Theoret. Math. Phys. **26**, 213 (1976).
[154] A. Fujiwara, Phys. Rev. A **65**, 012316 (2001).
[155] H. P. Yuen and M. Lax, IEEE Trans. Inf. Theory **19**, 740 (1973).
[156] C. Macchiavello, Phys. Rev. A **67**, 062302 (2003).
[157] M. A. Ballester, Phys. Rev. A **70**, 032310 (2004).
[158] M. A. Ballester, Phys. Rev. A **69**, 022303 (2004).
[159] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz Tapia, Phys. Rev. A **73**, 032301 (2006).
[160] L. Pezzè, M. A. Ciampini, N. Spagnolo, P. C. Humphreys, A. Datta, I. A. Walmsley, M. Barbieri, F. Sciarrino, and A. Smerzi, Phys. Rev. Lett. **119**, 130504 (2017).
[161] H. Yuan, Phys. Rev. Lett. **117**, 160801 (2016).
[162] M. Szczykulska, T. Baumgratz, and A. Datta, Quantum Science and Technology **2**, 044004 (2017).
[163] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[164] M. E. Pearce, E. T. Campbell, and P. Kok, Quantum **1**, 21 (2017).
[165] A. Acín, Phys. Rev. Lett. **87**, 177901 (2001).
[166] G. M. D'Ariano, P. Lo Presti, and M. G. A. Paris, Phys. Rev. Lett. **87**, 270404 (2001).
[167] R. Duan, Y. Feng, and M. Ying, Phys. Rev. Lett. **98**, 100503 (2007).
[168] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 180501 (2008).
[169] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, Phys. Rev. A **81**, 032339 (2010).
[170] R. Duan, Y. Feng, and M. Ying, Phys. Rev. Lett. **103**, 210501 (2009).
[171] D. Petz, Lin. Alg. Appl. **244**, 81 (1996).
[172] L. Maccone and A. K. Pati, Phys. Rev. Lett. **113**, 296 (2014).
[173] D. Petz, J. Math. Phys. **35**, 780 (1994).
[174] V. Vedral, Rev. Mod. Phys. **74**, 197 (2001).
[175] D. Petz and G. Toth, Lett. Math. Phys. **27**, 205 (1993).
[176] M. Hayashi, J. Phys. A: Math. Gen. **35**, 7689 (2002).
[177] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics*, Springer Series in Solid-state Sciences, Vol. 31 (Springer, 1985).
[178] H. J. D. Miller and J. Anders, Nat. Commun. **9**, 349 (2018).
[179] E. A. Morozova and N. N. Chentsov, Itogi Nauki i Tekhniki **36**, 69 (1990).
[180] E. P. Wigner and M. M. Yanase, Proc. Natl. Acad. Sci. **49**, 910 (1963).
[181] S. Luo, Phys. Rev. Lett. **91**, 101 (2003).
[182] S. Furuichi and K. Yanagi, J. Math. Anal. Appl. **388**, 1147 (2012).
[183] Z. Chen, Phys. Rev. A **71**, 195 (2005).
[184] X. Li, D. Li, H. Huang, and L. C. Kwek, Eur. Phys. J. D **64**, 147 (2011).
[185] I. Frérot and T. Roscilde, Phys. Rev. B **94**, 075121 (2016).
[186] R. Demkowicz-Dobrzański, Phys. Rev. A **83**, 061802 (2011).
[187] K. Macieszczak, M. Fraas, and R. Demkowicz-Dobrzański, New J. Phys. **16**, 113002 (2014).
[188] S. D. Personick, IEEE Trans. Inf. Theory **17**, 240 (1971).
[189] K. Chabuda, I. D. Leroux, and R. Demkowicz-Dobrzański, New J. Phys. **18**, 1 (2019).
[190] J. Rubio and J. Dunningham, New J. Phys. **21**, 043037 (2019).
[191] J. Z. Bernád, C. Sanavio, and A. Xuereb, Physical Review A **97**, 063821 (2018).
[192] J. Z. Bernád, C. Sanavio, and A. Xuereb, Physical Review A **99**, 062106 (2019).
[193] W. Rządkowski and R. Demkowicz-Dobrzański, Physical Review A **96**, 032319 (2017).
[194] J. Rubio and J. Dunningham, "Bayesian multi-parameter quantum metrology with limited data," (2019), unpublished, arXiv:1906.04123v2.
[195] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Rev. Mod. Phys. **84**, 621 (2012).
[196] A. Ferraro, S. Olivares, and M. G. A. Paris, "Gaussian states in continuous variable quantum information," (2005), unpublished, arXiv:quant-ph/0503237.
[197] S. L. Braunstein, Phys. Rev. A **71**, 055801 (2005).
[198] G. Adesso, S. Ragy, and A. R. Lee, Open Syst. Inf. Dyn. **21**, 1440001 (2014).
[199] O. Pinel, J. Fade, D. Braun, P. Jian, N. Treps, and C. Fabre, Phys. Rev. A **85**, 010101 (2012).
[200] A. Monras, "Phase space formalism for quantum estimation of Gaussian states," (2013), unpublished, arXiv:arXiv:1303.3682.
[201] Z. Jiang, Phys. Rev. A **89**, 032128 (2014).
[202] Y. Gao and H. Lee, Eur. Phys. J. D **68**, 1 (2014).
[203] D. Šafránek, A. R. Lee, and I. Fuentes, New J. Phys. **17**, 073016 (2015).
[204] D. Šafránek, J. Phys. A: Math. Theoret. **52**, 035304 (2018).
[205] P. Marian and T. A. Marian, Physical Review A **93**, 99 (2016).
[206] P. Marian and T. A. Marian, Phys. Rev. A **86**, 022340 (2012).
[207] C. Oh, C. Lee, L. Banchi, S.-Y. Lee, C. Rockstuhl, and H. Jeong, Phys. Rev. A **100**, 012323 (2019).
[208] C. Oh, C. Lee, C. Rockstuhl, H. Jeong, J. Kim, H. Nha, and S.-Y. Lee, NPJ Quantum Inf. **5**, 10 (2019).
[209] J. M. E. Fraïsse and D. Braun, Phys. Rev. A **95**, 062342 (2017).
[210] A. Holevo, Reports on Mathematical Physics **13**, 379 (1978).
[211] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986).
[212] B. C. Sanders and G. J. Milburn, Phys. Rev. Lett. **75**, 2944 (1995).
[213] U. Dorner, R. Demkowicz-Dobrzanski, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Phys. Rev. Lett. **102**, 040403 (2009).
[214] M. Skotiniotis, P. Sekatski, and W. Dür, New J. Phys. **17**, 073032 (2015).
[215] D. C. Brody and E.-M. Graefe, Entropy **15**, 3361 (2013).
[216] M. Tsang, H. M. Wiseman, and C. M. Caves, Phys. Rev. Lett. **106**, 090401 (2011).
[217] E. Magesan, A. Cooper, H. Yum, and P. Cappellaro, Phys. Rev. A **88**, 032107 (2013).
[218] I. Urizar-Lanz, P. Hyllus, I. L. Egusquiza, M. W. Mitchell, and G. Tóth, Phys. Rev. A **88**, 013626 (2013).
[219] S. Pang and T. A. Brun, Phys. Rev. A **93**, 059901 (2016).
[220] J. S. Sidhu and P. Kok, ArXiv (2018), arXiv:1802.01601 [quant-ph].
[221] H. Yuan and C.-H. F. Fung, Phys. Rev. Lett. **115**, 110401 (2015).
[222] M. Tsang, Optica **2**, 646 (2015).

[223] M.-D. Choi, Lin. Alg. Appl. **10**, 285 (1975).
[224] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, 1983rd ed., edited by A. Böhma, J. D. Dollard, and W. H. Wootters, Lecture Notes in Physics (Book 190) (Springer, 1983).
[225] H. Yuan and C.-H. F. Fung, NPJ Quantum Inf. **3** (2017), 10.1038/s41534-017-0014-6.
[226] R. Demkowicz-Dobrzanski, J. Kolodynski, and M. Guta, Nat. Commun. **3**, 1063 (2012).
[227] A. Monras and M. G. A. Paris, Phys. Rev. Lett. **98**, 160401 (2007).
[228] R. Chaves, J. B. Brask, M. Markiewicz, J. Kołodyński, and A. Acin, Phys. Rev. Lett. **111**, 120401 (2013).
[229] M. Beau and A. del Campo, Phys. Rev. Lett. **119**, 010403 (2017).
[230] Y. Matsuzaki, S. Benjamin, and J. Fitzsimons, Phys. Rev. A **84**, 012103 (2011).
[231] S. Dooley, W. J. Munro, and K. Nemoto, Phys. Rev. A **94**, 052320 (2016).
[232] H. P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, 2002).
[233] P. Sekatski, M. Skotiniotis, J. Kołodyński, and W. Dür, Quantum **1**, 27 (2017).
[234] R. Demkowicz-Dobrzański, J. Czajkowski, and P. Sekatski, Phys. Rev. X **7**, 041009 (2017).
[235] S. Zhou, M. Zhang, J. Preskill, and L. Jiang, Nat. Commun. **9**, 78 (2018).
[236] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin, Phys. Rev. Lett. **112**, 150802 (2014).
[237] W. Dür, M. Skotiniotis, F. Fröwis, and B. Kraus, Phys. Rev. Lett. **112**, 080801 (2014).
[238] M. A. Rubin and S. Kaushik, Phys. Rev. A **75**, 053805 (2007).
[239] G. Gilbert, M. Hamrick, and Y. S. Weinstein, J. Opt. Soc. Am. B **25**, 1336 (2008).
[240] S. D. Huver, C. F. Wildfeuer, and J. P. Dowling, Phys. Rev. A **78**, 063828 (2008).
[241] S. Knysh, V. N. Smelyanskiy, and G. A. Durkin, Phys. Rev. A **83**, 021804 (2011).
[242] M. Kacprowicz, R. Demkowicz-Dobrzański, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Nat. Photon. **4**, 357 (2010).
[243] J. Roffe, D. Headley, N. Chancellor, D. Horsman, and V. Kendon, Quant. Sci. Tech. **3**, 035010 (2018).
[244] B. J. Falaye, A. G. Adepoju, A. S. Aliyu, M. M. Melchor, M. S. Liman, O. J. Oluwadare, M. D. González-Ramírez, and K. J. Oyewumi, Sci. Rep. **7**, 16622 (2017).
[245] G. Spedalieri, C. Lupo, S. L. Braunstein, and S. Pirandola, Quant. Sci. Tech. **4**, 015008 (2018).
[246] Ying-Wang, L.-Y. Hu, Z.-M. Zhang, and C.-P. Wei, Opt. Commun. **427**, 250 (2018).
[247] D. Braun, Phys. Rev. Lett. **89**, 277901 (2002).
[248] F. Benatti, R. Floreanini, and M. Piani, Phys. Rev. Lett. **91**, 070402 (2003).
[249] A. Streltsov, H. Kampermann, and D. Bruß, Phys. Rev. Lett. **107**, 170502 (2011).
[250] M. Piani, S. Gharibian, G. Adesso, J. Calsamiglia, P. Horodecki, and A. Winter, Phys. Rev. Lett. **106**, 220403 (2011).
[251] A. Fujiwara, Phys. Rev. A **63**, 042304 (2001).
[252] A. Fujiwara and H. Imai, J. Phys. A: Math. Gen. **36**, 8093 (2003).
[253] A. Fujiwara, Phys. Rev. A **70**, 012317 (2004).
[254] Z. Huang, C. Macchiavello, and L. Maccone, Phys. Rev. A **94**, 012101 (2016).
[255] Z. Huang, C. Macchiavello, and L. Maccone, Phys. Rev. A **97**, 032333 (2018).
[256] K. Wang, X. Wang, X. Zhan, Z. Bian, J. Li, B. C. Sanders, and P. Xue, Phys. Rev. A **97**, 042112 (2018).
[257] M. Sbroscia, I. Gianani, L. Mancino, E. Roccia, Z. Huang, L. Maccone, C. Macchiavello, and M. Barbieri, Phys. Rev. A **97**, 032305 (2018).
[258] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White, Phys. Rev. Lett. **90**, 193601 (2003).
[259] M. Hotta, T. Karasawa, and M. Ozawa, Phys. Rev. A **72**, 052334 (2005).
[260] S. Pirandola and C. Lupo, Phys. Rev. Lett. **118**, 100502 (2017).
[261] M. Takeoka and M. M. Wilde, "Optimal estimation and discrimination of excess noise in thermal and amplifier channels," (2016), unpublished, arXiv:quant-ph/1611.09165.
[262] R. Laurenza, C. Lupo, G. Spedalieri, S. L. Braunstein, and S. Pirandola, Quant. Meas. Quant. Metr. **5**, 1 (2019).
[263] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Nat. Commun. **8**, 15043 (2017).
[264] P. Papanastasiou, C. Lupo, C. Weedbrook, and S. Pirandola, Phys. Rev. A **98**, 012340 (2018).
[265] S. T. Flammia and J. J. Wallman, ArXiv (2019), arXiv:1907.12976 [quant-ph].
[266] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, Phys. Rev. Lett. **78**, 3217 (1997).
[267] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, NPJ Quantum Inf. **5**, 50 (2019).
[268] Q. Liao, Y. Guo, D. Huang, P. Huang, and G. Zeng, New J. Phys. **20**, 023015 (2018).
[269] T. Cope and S. Pirandola, Quant. Meas. Quant. Metr. **4**, 44 (2017).
[270] S. Lloyd, Science **321**, 1463 (2008).
[271] J. Roffe, Contemp. Phys. **0**, 1 (2019).
[272] G. Arrad, Y. Vinkler, D. Aharonov, and A. Retzker, Phys. Rev. Lett. **112**, 150801 (2014).
[273] S. J. Devitt, W. J. Munro, and K. Nemoto, Rep. Prog. Phys. **76**, 076001 (2013).
[274] B. M. Terhal, Rev. Mod. Phys. **87**, 307 (2015).
[275] D. Layden, S. Zhou, P. Cappellaro, and L. Jiang, Phys. Rev. Lett. **122**, 040502 (2019).
[276] F. Reiter, A. S. Sørensen, P. Zoller, and C. A. Muschik, Nat. Commun. **8**, 1822 (2017).
[277] T. Unden, P. Balasubramanian, D. Louzon, Y. Vinkler, M. B. Plenio, M. Markham, D. Twitchen, A. Stacey, I. Lovchinsky, A. O. Sushkov, *et al.*, Phys. Rev. Lett. **116**, 230502 (2016).
[278] D. Layden and P. Cappellaro, NPJ Quantum Inf. **4**, 30 (2018).
[279] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, Phys. Rev. Lett. **121**, 043604 (2018).
[280] T. J. Proctor, P. A. Knott, and J. A. Dunningham, Phys. Rev. Lett. **120**, 080501 (2018).
[281] Q. Zhuang, Z. Zhang, and J. H. Shapiro, Phys. Rev. A **97**, 032329 (2018).
[282] H. J. Kimble, Nature **453**, 1023 (2008).
[283] N. H. Nickerson, Y. Li, and S. C. Benjamin, Nat. Commun. **4**, 1756 (2013).
[284] M. Sasaki, M. Fujiwara, H. Ishizuka, W. Klaus, K. Wakui, M. Takeoka, S. Miki, T. Yamashita, Z. Wang, A. Tanaka, *et al.*, Opt. Express **19**, 10387 (2011).
[285] S.-K. Liao, W.-Q. Cai, W.-Y. Liu, L. Zhang, Y. Li, J.-G. Ren, J. Yin, Q. Shen, Y. Cao, Z.-P. Li, *et al.*, Nature **549**, 43 (2017).
[286] T. J. Proctor, P. A. Knott, and J. A. Dunningham, "Networked quantum sensing," (2017), unpublished, arXiv:1702.04271.
[287] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Phys. Rev. A **85**, 042112 (2012).
[288] N. Friis, D. Orsucci, M. Skotiniotis, P. Sekatski, V. Dunjko, H. J. Briegel, and W. Dür, New J. Phys. **19**, 063044 (2017).
[289] R. Yousefjani, R. Nichols, S. Salimi, and G. Adesso, Phys. Rev. A **95**, 062307 (2017).
[290] W. D. Pavel Sekatski, Sabine Wölk, "Optimal distributed sensing in noisy environments," (2019), unpublished, arXiv:1905.06765.
[291] J. Etesse, M. Bouillard, B. Kanseri, and R. Tualle-Brouri, Phys. Rev. Lett. **114**, 193602 (2015).
[292] P. A. Knott, T. J. Proctor, A. J. Hayes, J. P. Cooling, and J. A. Dunningham, Phys. Rev. A **93**, 033859 (2016).
[293] F. Wolfgramm, C. Vitelli, F. A. Beduini, N. Godbout, and M. W. Mitchell, Nat. Photon. **7**, 28 (2013).
[294] M. A. Taylor, J. Janousek, V. Daria, J. Knittel, B. Hage, H.-A. Bachor, and W. P. Bowen, Nat. Photon. **7**, 229 (2013).
[295] M. A. Taylor, J. Janousek, V. Daria, J. Knittel, B. Hage, H.-A. Bachor, and W. P. Bowen, Phys. Rev. X **4**, 011017 (2014).
[296] L. M. Pham, D. L. Sage, P. L. Stanwix, T. K. Yeung, D. Glenn, A. Trifonov, P. Cappellaro, P. R. Hemmer, M. D. Lukin, H. Park, A. Yacoby, and R. L. Walsworth, New J. Phys. **13**, 045021 (2011).
[297] N. Alfasi, S. Masis, O. Shtempluck, V. Kochetok, and E. Buks, AIP Advances **6**, 075311 (2016).
[298] P. Kómár, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye, and M. D. Lukin, Nature **10**, 582 (2014).
[299] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov, Phys. Rev. A **100**, 042304 (2019).

[300] J. A. Gross, "Estimating functions of multiple parameters," (2019), blog post.
[301] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, Phys. Rev. A **97**, 042337 (2018).