Editors' Suggestion

Physics-inspired forms of the Bayesian Cramér-Rao bound

Mankei Tsang *

Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583 and Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551

(Received 30 August 2020; revised 2 December 2020; accepted 2 December 2020; published 23 December 2020)

Using differential geometry, I derive a form of the Bayesian Cramér-Rao bound that remains invariant under reparametrization. With the invariant formulation at hand, I find the optimal and naturally invariant bound among the Gill-Levit family of bounds. By assuming that the prior probability density is the square of a wave function, I also express the bounds in terms of functionals that are quadratic with respect to the wave function and its gradient. The problem of finding an unfavorable prior to tighten the bound for minimax estimation is shown, in a special case, to be equivalent to finding the ground state of a Schrödinger equation, with the Fisher information playing the role of the potential. To illustrate the theory, two quantum estimation problems, namely, optomechanical waveform estimation and subdiffraction incoherent optical imaging, are discussed.

DOI: 10.1103/PhysRevA.102.062217

I. INTRODUCTION

Differential geometry has been useful in the study of statistical divergence measures, Cramér-Rao bounds, and asymptotic statistics [1–3], but its usefulness for Bayesian and minimax statistics is less clear. The Bayesian Cramér-Rao bounds [4,5], pioneered by Schützenberger [6] and Van Trees [7], may serve as a bridge.

To set the stage, consider a p-dimensional parameter $\theta = (\theta^1, \ldots, \theta^p) \in \Theta \subseteq \mathbb{R}^p$, a scalar parameter of interest $\beta(\theta) \in \mathbb{R}$ that is a function of θ, and an estimator $\check{\beta}(X)$, where X is a set of n independent and identically distributed (i.i.d.) observation random variables with a family of probability densities $\{f^{(n)}(x|\theta) = \prod_{j=1}^n f(x^j|\theta) : \theta \in \Theta\}$ and a reference measure μ that gives $d\mu^{(n)}(x) = \prod_{j=1}^n d\mu(x_j)$. Generalization of the theory for a vectoral β is straightforward but tedious and deferred to the Appendix. Define the mean-square risk as

$$\mathsf{R}(\theta) \equiv \int [\check{\beta}(x) - \beta(\theta)]^2 f^{(n)}(x|\theta) d\mu^{(n)}(x). \tag{1.1}$$

The Cramér-Rao bound for any unbiased estimator is given by

$$\mathsf{R}(\theta) \geqslant \frac{\mathsf{C}(\theta)}{n}, \tag{1.2}$$

$$\mathsf{C}(\theta) \equiv u_a(\theta)[F(\theta)^{-1}]^{ab} u_b(\theta), \tag{1.3}$$

where

$$u_a \equiv \partial_a \beta, \quad \partial_a \equiv \frac{\partial}{\partial \theta^a}. \tag{1.4}$$

Einstein summation is assumed, F is the Fisher information matrix defined as

$$F_{ab} \equiv \int (\partial_a \ln f)(\partial_b \ln f) f d\mu, \tag{1.5}$$

F^{-1} is its inverse such that $F_{ab}(F^{-1})^{bc} = \delta_a^c$, and δ is the Kronecker delta. For simplicity, hereafter I call Eqs. (1.2) and (1.3) the local bound and the theory concerning $\mathsf{C}(\theta)$ the local theory, as $\mathsf{C}(\theta)$ depends only on the local properties of the statistical model in the neighborhood of θ.

The restriction to unbiased estimators is one of the biggest shortcomings of the local bound. A fruitful remedy is to consider bounds on the Bayesian risk,

$$\langle \mathsf{R} \rangle = \mathbb{E}[(\check{\beta} - \beta)^2] = \int \mathsf{R}(\theta)\pi(\theta) d^p\theta, \tag{1.6}$$

where $\mathbb{E}$ denotes the expectation over both the observation and the parameter as random variables and π is a prior probability density [4]. In particular, Gill and Levit proposed a general family of Bayesian Cramér-Rao bounds, valid for any biased or unbiased estimator, given by [5]

$$\langle \mathsf{R} \rangle \geqslant \mathsf{B} \equiv \frac{\langle \mathsf{A} \rangle^2}{n\langle \mathsf{F} \rangle + \langle \mathsf{P} \rangle}, \tag{1.7}$$

$$\mathsf{A} \equiv v^a u_a, \tag{1.8}$$

$$\mathsf{F} \equiv v^a F_{ab} v^b, \tag{1.9}$$

$$\mathsf{P} \equiv \left[\frac{1}{\pi}\partial_a(\pi v^a)\right]^2, \tag{1.10}$$

where v, A, F, and P are all functions of θ, πv is assumed to vanish on the boundary of Θ, and $\langle\cdot\rangle$ denotes the prior expectation, as in Eq. (1.6).

This work studies only the bound B; the attainability of the bound is outside the scope of this work. Some recently proposed Bayesian Cramér-Rao bounds [8] may not fall under the Gill-Levit family and are also outside the scope of this work. There also exist many other types of Bayesian bounds that may be tighter, such as the Ziv-Zakai bounds and the Weiss-Weinstein bounds [4], but the Cramér-Rao bounds are

*mankei@nus.edu.sg; https://blog.nus.edu.sg/mankei/

 ©2020 American Physical Society

often much easier to compute because they are based on the Fisher information, a well-studied quantity.

In Eqs. (1.8)–(1.10), $v : \mathbb{R}^p \to \mathbb{R}^p$ is a free term, and by choosing it judiciously, many useful forms of B can be obtained [5]. An arbitrarily chosen v, however, may lead to a B that varies if the parametrization of the underlying model with respect to θ is changed. To give a simple example, suppose that $p = 1$, θ is a scalar, and $\beta = \theta$. Consider the Gill-Levit bound for $v^1 = 1$. If the parametrization of the underlying statistical model is changed, say, via the relation $\theta = \tilde{\theta}^{1/3}$ and $\tilde{\theta} = \theta^3$, then the Gill-Levit bound for $\beta = \theta = \tilde{\theta}^{1/3}$ and $v^1 = 1$ would usually be different when computed with respect to the new parameter $\tilde{\theta}$, even if the statistical problem remains the same. This property is unpleasant, as there can be infinitely many parametrizations for the same model and it is not clear which parametrization leads to the tightest bound for a given problem. Note that the local bound given by Eq. (1.3) does not suffer from such a problem, as it is well known to be invariant upon reparametrization [9]. In Sec. II, I propose a condition on v that makes B invariant. I also derive an invariant form of B using the language of differential geometry [10]. With the invariant form, B is guaranteed to give the same value for a model, regardless of the parametrization.

A related question is how v should be chosen. Although Gill and Levit suggested a few options based on prior works or convenience, it is unclear which is better or whether there exists an optimal choice. In Sec. III, I show that there is indeed an optimal choice, and it agrees with a couple of popular options in special cases. The inspiration comes from the geometric picture of v as a vector field, which generalizes the role of a tangent vector in the local theory [2,9]. By virtue of the invariant formalism, the resultant bound is naturally invariant.

Bayesian bounds are also useful for minimax statistics [11] by providing lower bounds on the worst-case risk via

$$\sup_{\theta\in\Theta} \mathsf{R}(\theta) \geqslant \langle \mathsf{R} \rangle \tag{1.11}$$

for any prior. In this context, one should no longer choose the prior according to Bayesian principles. Instead, one should choose an unfavorable prior as a mathematical device to tighten a lower bound. Given Eqs. (1.7)–(1.10), it is unclear how the prior should be chosen, as $\langle \mathsf{P} \rangle$ is highly nonlinear with respect to π. To help with this problem, in Sec. IV I rewrite Eqs. (1.7)–(1.10) in a form that looks more familiar, at least to physicists. To be specific, I identify the prior density with the square of a wave function, such that $\langle \mathsf{A} \rangle$, $\langle \mathsf{F} \rangle$, and, most importantly, $\langle \mathsf{P} \rangle$ all become quadratic functionals of the wave function and its gradient. In a special case, $n\langle \mathsf{F} \rangle + \langle \mathsf{P} \rangle$ becomes the average energy of a wave that obeys a Schrödinger equation. Finding the tightest bound for minimax estimation then becomes equivalent to finding the ground-state energy of the wave, and insights from quantum mechanics turn out to be handy.

In terms of other prior works, Refs. [12,13] also study Bayesian Cramér-Rao bounds in geometric terms but do not discuss the question of invariance or find the optimal Gill-Levit bound. References [14] derive the asymptotically optimal form of the Gill-Levit bounds but do not find the exact optimal form. Example 4.2 in Ref. [15] studies the optimization of a Bayesian Cramér-Rao bound for a special problem, but not in the generality considered here. Regarding the wave picture, the fact that F is quadratic with respect to $\partial_a(f^{1/2})$ is well known in statistics [1], and Frieden even claimed that it serves as a fundamental principle for physics [16]. He assumed that f is the square of a wave function and derived wave equations from this fact but had to introduce further creative assumptions. He also did not consider Bayesian bounds. To my knowledge, the wave picture of a Bayesian Cramér-Rao bound is first proposed in Ref. [17], which considers the special case $\beta = \theta$ with a scalar θ and uses the wave picture as a trick to solve a parameter-estimation problem in optical imaging. Here, as before [17], I do not claim that my results have any foundational implications for physics, merely that the correspondence is interesting and useful for statistics problems.

Section V comes full circle and applies the statistical theory to quantum estimation [18,19], where actual quantum systems are considered. I consider two important problems in quantum optics, namely, optomechanical waveform estimation [20,21] and subdiffraction incoherent imaging [18,22]. The first problem is relevant to gravitational-wave detectors, where quantum noise is now playing a major role [23]; I show the importance of including prior information in deriving a meaningful quantum limit in terms of spectral quantities, following Ref. [21]. The second problem is, of course, a fundamental one in optics and relevant to both fluorescence microscopy and observational astronomy. Recent studies, based on quantum estimation theory, have shown that judicious measurements can substantially improve the imaging of subdiffraction objects [22], although most prior works are based on the local bound, which is valid for unbiased estimators only. By considering the minimax perspective, the Bayesian bound, and the wave picture, I discuss the implication of a zero information for the estimator convergence rate for the multisource localization problem studied in Refs. [22,24,25].

II. INVARIANCE

To model reparametrization, consider a bijective differentiable map $\tilde{\theta}(\theta)$. The transformation laws are

$$\partial_a = J_a^b \tilde{\partial}_b, \quad \tilde{\partial}_a \equiv \frac{\partial}{\partial \tilde{\theta}^a}, \tag{2.1}$$

$$d^p\theta = \frac{d^p\tilde{\theta}}{\|J\|}, \quad \pi = \|J\|\tilde{\pi}, \tag{2.2}$$

$$u_a = J_a^b \tilde{u}_b, \quad F_{ab} = J_a^c \tilde{F}_{cd} J_b^d, \tag{2.3}$$

where

$$J_a^b \equiv \partial_a \tilde{\theta}^b \tag{2.4}$$

is the Jacobian matrix, $|J|$ denotes its determinant, and $\|J\|$ denotes the absolute value of the determinant. Equations (2.3) imply that the components of u are covariant and F is a (0,2) tensor. On the other hand, β, $\check{\beta}$, f, μ, R, and $\langle\cdot\rangle$ remain invariant in the sense that $\beta(\theta) = \tilde{\beta}(\tilde{\theta}(\theta))$, $f(x|\theta) = \tilde{f}(x|\tilde{\theta}(\theta))$, $\mathsf{R}(\theta) = \tilde{\mathsf{R}}(\tilde{\theta}(\theta))$, etc., as these quantities depend on the statistical problem and should not depend on the parametrization of the underlying model.

It is well known that the local bound is invariant under reparametrization [9], in the sense of

$$u_a\big(F^{-1}\big)^{ab}u_b = \tilde{u}_a\big(\tilde{F}^{-1}\big)^{ab}\tilde{u}_b. \tag{2.5}$$

The Gill-Levit bounds can also be made invariant.

Proposition 1. B is invariant under reparametrization if v obeys the transformation law

$$v^a J^b_a = \tilde{v}^b. \tag{2.6}$$

Proof. Given Eq. (2.6), it is obvious that

$$\mathsf{A} = \tilde{v}^a\tilde{u}_a, \quad \mathsf{F} = \tilde{v}^a\tilde{F}_{ab}\tilde{v}^b \tag{2.7}$$

remain invariant upon reparametrization. To deal with P, define the inverse Jacobian matrix as

$$\tilde{J}^b_a \equiv \tilde{\partial}_a\theta^b, \tag{2.8}$$

which obey

$$J^b_a\tilde{J}^c_b = \tilde{J}^b_a J^c_b = \delta^c_a, \quad |\tilde{J}| = \frac{1}{|J|}. \tag{2.9}$$

Consider

$$\frac{1}{\pi}\partial_a(\pi v^a) = \frac{J^b_a}{|J|\tilde{\pi}}\tilde{\partial}_b\big(|J|\tilde{\pi}\tilde{v}^c\tilde{J}^a_c\big) \tag{2.10}$$

$$= \tilde{v}^c\frac{J^b_a}{|J|}\tilde{\partial}_b\big(|J|\tilde{J}^a_c\big) + \frac{1}{\tilde{\pi}}\tilde{\partial}_b(\tilde{\pi}\tilde{v}^b). \tag{2.11}$$

The first term can be shown to vanish as follows:

$$\frac{J^b_a}{|J|}\tilde{\partial}_b\big(|J|\tilde{J}^a_c\big) = \tilde{\partial}_c\ln|J| + J^b_a\tilde{\partial}_b\tilde{J}^a_c \tag{2.12}$$

$$= -\tilde{\partial}_c\ln|\tilde{J}| + J^b_a\tilde{\partial}_b\tilde{J}^a_c \tag{2.13}$$

$$= -J^b_a\tilde{\partial}_c\tilde{J}^a_b + J^b_a\tilde{\partial}_b\tilde{J}^a_c \tag{2.14}$$

$$= -J^b_a(\tilde{\partial}_c\tilde{\partial}_b\theta^a - \tilde{\partial}_b\tilde{\partial}_c\theta^a) = 0, \tag{2.15}$$

where Eq. (2.14) uses Jacobi's formula to simplify $\tilde{\partial}_c\ln|\tilde{J}|$. Hence

$$\frac{1}{\pi}\partial_a(\pi v^a) = \frac{1}{\tilde{\pi}}\tilde{\partial}_b\big(\tilde{\pi}\tilde{v}^b\big), \tag{2.16}$$

and P is invariant. As the prior expectation $\langle\cdot\rangle$ is also invariant, B is invariant. ∎

In the language of differential geometry, Eq. (2.6) means that the components of v are contravariant. In other words, v defines a vector field in the parameter space Θ, with components $(v^1,\ldots,v^p)$ with respect to a parametrization. If one does not transform the components as per Eq. (2.6) upon reparametrization, B changes—the reason, from the geometric perspective, is that it has become a bound for a different vector field. For someone familiar with differential geometry, Proposition 1 may seem trivial in hindsight, but this triviality should be regarded as a virtue—it is evidence that differential geometry is useful in simplifying the problem here.

A "natural" choice of the v components according to Gill and Levit is [5]

$$v^a = (F^{-1})^{ab}u_b. \tag{2.17}$$

This form is contravariant, in the sense that Eq. (2.17) for one parametrization and $\tilde{v}^a = (\tilde{F}^{-1})^{ab}\tilde{u}_b$ for another parametrization obey Eq. (2.6) and must give the same bound for a given problem. This choice also leads to the simplification

$$\mathsf{A} = \mathsf{F} = u_a(F^{-1})^{ab}u_b = \mathsf{C}, \tag{2.18}$$

which coincides with the local bound given by Eq. (1.3). The resultant Bayesian bound is

$$\mathsf{B} = \frac{\langle\mathsf{C}\rangle^2}{n\langle\mathsf{C}\rangle + \langle\mathsf{P}\rangle}. \tag{2.19}$$

For a scalar θ, this becomes an inequality of Borovkov and Sakhanenko [26]; see also Ref. [27]. Most importantly, Eq. (2.19) agrees with some classic theorems in the asymptotic local theory by Hájek and Le Cam that generalize the Cramér-Rao bound but are much more sophisticated [5,28]. Equation (2.17) is not the only contravariant choice, however. It does not even exist if u is not in the range of the F matrix [29]. It is also not the optimal choice for the Gill-Levit bounds in general, as Sec. III shows.

Another useful choice of the v components is

$$v^a = [(n\langle F\rangle + \langle G\rangle)^{-1}]^{ab}\langle u_b\rangle, \tag{2.20}$$

$$G_{ab} \equiv \frac{1}{\pi}(\partial_a\pi)\frac{1}{\pi}(\partial_b\pi), \tag{2.21}$$

leading to

$$\mathsf{B} = \langle u_a\rangle[(n\langle F\rangle + \langle G\rangle)^{-1}]^{ab}\langle u_b\rangle. \tag{2.22}$$

If u is θ independent, Eq. (2.22) coincides with the original version by Schützenberger and Van Trees [6,7]. $\langle G\rangle$ plays the role of prior information and can regularize the inverse when $\langle F\rangle$ is ill conditioned. The regularization is especially important for waveform-estimation problems [7,21]. The form of Eq. (2.20) is usually not contravariant, however, in the sense that, except for special cases, Eq. (2.20) for one parametrization and $\tilde{v}^a = [(n\langle\tilde{F}\rangle + \langle\tilde{G}\rangle)^{-1}]^{ab}\langle\tilde{u}_b\rangle$ for another parametrization do not obey Eq. (2.6), and the resultant bounds may be different for a given problem.

In the following, I generalize Θ, the parameter space, to a p-dimensional manifold and assume that v is a vector field on the manifold. The formalism can then be made more elegant by defining the invariant quantities

$$\epsilon \equiv \sqrt{|g|}d^p\theta, \quad \rho \equiv \frac{\pi}{\sqrt{|g|}}, \quad \pi d^p\theta = \rho\epsilon, \tag{2.23}$$

where $|g|$ is the determinant of a Riemannian (positive-definite) metric g_{ab}. It should be emphasized that the metric here is merely a mathematical tool to keep track of parametrization invariance and deal with more general manifolds for Θ, and this work is not concerned with the concept of statistical manifolds and distances between probability measures in information geometry [1]. Although many have argued that the Fisher information is a natural metric in information geometry [1], there is no particular reason to pick the Fisher information as the metric here. That choice may also cause problems if $|F| = 0$, so I keep the metric unspecified here for generality. The divergence term in Eq. (1.10) becomes

$$\frac{1}{\pi}\partial_a(\pi v^a) = \frac{1}{\sqrt{|g|}\rho}\partial_a(\sqrt{|g|}\rho v^a) = \frac{1}{\rho}\nabla_a(\rho v^a), \tag{2.24}$$

where ∇_a is the Riemannian covariant derivative. With these suggestive expressions at hand, I propose the following.

Proposition 2 (invariant Gill-Levit bounds). If ρv vanishes on any boundary of the parameter manifold Θ, the Bayesian mean-square risk has a lower bound given by Eq. (1.7), where

$$\langle \mathsf{A} \rangle = \int (v^a u_a) \rho \epsilon, \tag{2.25}$$

$$\langle \mathsf{F} \rangle = \int (v^a F_{ab} v^b) \rho \epsilon, \tag{2.26}$$

$$\langle \mathsf{P} \rangle = \int \left[\frac{1}{\rho} \nabla_a (\rho v^a) \right]^2 \rho \epsilon. \tag{2.27}$$

Proof. For completeness, I provide a proof that proceeds in a manifestly invariant way, so that the proposition is proved also for a curved metric. Define the bias as

$$\mathsf{b} \equiv \int (\check{\beta} - \beta) f^{(n)} d\mu^{(n)} \tag{2.28}$$

and write, via the Leibniz rule for the covariant derivative,

$$\int \nabla_a (\mathsf{b} \rho v^a) \epsilon = \iint (\check{\beta} - \beta) \nabla_a (f^{(n)} \rho v^a) d\mu^{(n)} \epsilon - \int (v^a \nabla_a \beta) \rho \epsilon. \tag{2.29}$$

It can be shown that the left-hand side of Eq. (2.29) is 0 by applying the Stokes theorem [10,30] and requiring that ρv vanishes on the boundary of Θ if there is a boundary. With $\nabla_a \beta = \partial_a \beta$ when ∇_a acts on a scalar, the last term in Eq. (2.29) is precisely $\langle \mathsf{A} \rangle$ in Eq. (2.25). I obtain

$$\langle \mathsf{A} \rangle = \iint (\check{\beta} - \beta) \nabla_a (f^{(n)} \rho v^a) d\mu^{(n)} \epsilon \tag{2.30}$$

$$= \mathbb{E}[(\check{\beta} - \beta) s], \tag{2.31}$$

where s is a generalized score function given by

$$s \equiv \frac{1}{f^{(n)} \rho} \nabla_a (f^{(n)} \rho v^a) \tag{2.32}$$

$$= \frac{1}{f^{(n)}} v^a \nabla_a f^{(n)} + \frac{1}{\rho} \nabla_a (\rho v^a). \tag{2.33}$$

The expectation can be regarded as an inner product. The Cauchy-Schwarz inequality then gives

$$\langle \mathsf{A} \rangle^2 \leqslant \mathbb{E}[(\check{\beta} - \beta)^2] \mathbb{E}(s^2). \tag{2.34}$$

With the usual premise

$$\int \nabla_a f d\mu = \int \partial_a f(x|\theta) d\mu(x) = \partial_a \int f d\mu = 0, \tag{2.35}$$

it can be shown that

$$\mathbb{E}(s^2) = n \langle \mathsf{F} \rangle + \langle \mathsf{P} \rangle, \tag{2.36}$$

with $\langle \mathsf{F} \rangle$ given by Eq. (2.26) and $\langle \mathsf{P} \rangle$ given by Eq. (2.27). Hence, Eq. (2.34) leads to Eq. (1.7), together with Eqs. (2.25)–(2.27). ∎

The original Gill-Levit bounds given by Eqs. (1.7)–(1.10) may be viewed as a special case of Proposition 2 if one can pick a parametrization (coordinate system) with $g_{ab} = \delta_{ab}$ everywhere in Θ. If the Riemann curvature tensor with respect to the metric is 0 everywhere, then one can always find a parametrization for which $g_{ab} = \delta_{ab}$ [10], and the two formulations are equivalent in essence. But if not, the metric is said to be curved, and Proposition 2 is more general. Proposition 2 may also be regarded as a special case of Theorem 2.1 in Ref. [12], although the latter is so general that the bound there may depend on the estimator.

While it is unclear whether curved metrics are useful for the kind of problems considered here, one immediate advantage of the invariant formulation is that all the ensuing results are guaranteed to be invariant.

III. OPTIMAL GILL-LEVIT BOUND

To derive the optimal Gill-Levit bound, it is illuminating to first recall the concept of least favorable submodels in the local theory, as outlined in Ref. [9]; see also Ref. [31]. Pick a curve in the parameter space that passes through the true value and denote a tangent vector there as v. The local bound for the one-dimensional submodel is given by

$$\mathsf{C}(v) = \frac{(v^a u_a)^2}{v^a F_{ab} v^b}. \tag{3.1}$$

Define an inner product between two vectors as

$$\langle v, w \rangle_g \equiv v^a w_a = v^a g_{ab} w^a, \tag{3.2}$$

where the usual convention of index lowering and raising via g_{ab} and its inverse g^{ab} in differential geometry is assumed. Let F be an operator that obeys $(Fv)_a = F_{ab} v^b$. If F is positive-definite, F^{-1} and the square roots $F^{1/2}$ and $F^{-1/2}$ exist [32]. The Cauchy-Schwarz inequality gives

$$\mathsf{C}(v) = \frac{\langle v, u \rangle_g^2}{\langle v, Fv \rangle_g} = \frac{\langle F^{1/2} v, F^{-1/2} u \rangle_g^2}{\langle v, Fv \rangle_g} \tag{3.3}$$

$$\leqslant \langle u, F^{-1} u \rangle_g = u_a (F^{-1})^{ab} u_b, \tag{3.4}$$

which coincides with Eq. (1.3) for the full model. A least favorable tangent vector that attains the equality must satisfy

$$v^a \propto (F^{-1})^{ab} u_b. \tag{3.5}$$

Thus, Eq. (1.3) can be evaluated by considering the tangent space at the true parameter and picking the worst direction.

For the Gill-Levit bounds, the "natural" choice of v given by Eq. (2.17) is a least favorable choice in the local theory. Thus, one may intuit that v plays an analogous role of picking out directions in the Bayesian bound, except that v should now be considered as a vector field, as depicted in Fig. 1. In differential geometry, a vector field can generate a family of integral curves, called a flow, in the manifold, and vice versa [30]. In the context of statistics, each curve corresponds to a one-dimensional submodel, so the concept of locally least favorable submodels may be generalized to a concept of least favorable flows. Following this intuition, I can generalize the strategy of optimizing over v to obtain the tightest bound, as follows.

Theorem 1 (optimal Gill-Levit bound).

$$\max_v \mathsf{B} = \langle u, L^{-1} u \rangle_\rho \equiv \mathsf{B}_{\max}, \tag{3.6}$$

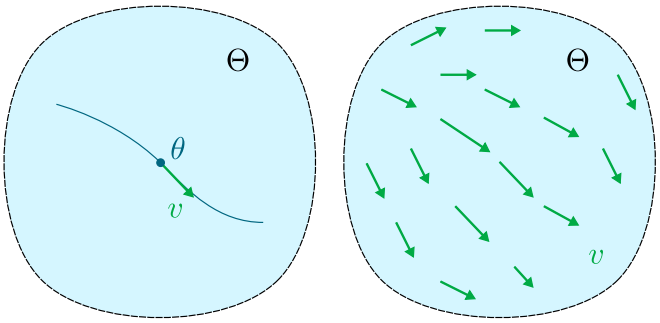

FIG. 1. Left: A geometric picture of a one-dimensional submodel as a curve in the manifold and a tangent vector v at the true parameter value θ in the local theory. Right: A picture of v as a vector field in the Bayesian theory.

where the inner product between two vector fields is defined as

$$\langle v, u\rangle_\rho \equiv \int v^a u_a \rho \epsilon, \tag{3.7}$$

the linear, self-adjoint, and positive-semidefinite operator L is defined as

$$(Lv)_a \equiv nF_{ab}v^b - \nabla_a\left[\frac{1}{\rho}\nabla_b(\rho v^b)\right], \tag{3.8}$$

and u is assumed to be in the range of L, such that $L^{-1}u$ exists. A least favorable vector field, defined as a v that maximizes B, must satisfy

$$v \propto L^{-1}u. \tag{3.9}$$

Proof. In terms of the inner product given by Eq. (3.7), Eqs. (2.25)–(2.27) can be expressed as

$$\langle \mathsf{A}\rangle = \langle v, u\rangle_\rho, \tag{3.10}$$

$$\langle \mathsf{F}\rangle = \langle v, Fv\rangle_\rho, \tag{3.11}$$

$$\langle \mathsf{P}\rangle = \int [\nabla_a(\rho v^a)]\frac{1}{\rho}\nabla_b(\rho v^b)\epsilon \tag{3.12}$$

$$= -\int \rho v^a \nabla_a\left[\frac{1}{\rho}\nabla_b(\rho v^b)\right]\epsilon \tag{3.13}$$

$$= \langle v, Pv\rangle_\rho, \tag{3.14}$$

$$(Pv)_a \equiv -\nabla_a\left[\frac{1}{\rho}\nabla_b(\rho v^b)\right], \tag{3.15}$$

where Eq. (3.13) comes from integration by parts, as enabled by the Leibniz rule and the Stokes theorem, and the assumption that ρv vanishes on any boundary of Θ. One can check that F and P are linear, self-adjoint, and positive-semidefinite operators. Furthermore,

$$n\langle \mathsf{F}\rangle + \langle \mathsf{P}\rangle = \langle v, Lv\rangle_\rho, \quad L = nF + P. \tag{3.16}$$

As $L^{-1}u$ is assumed to exist, the Cauchy-Schwarz inequality yields

$$\mathsf{B} = \frac{\langle v, u\rangle_\rho^2}{\langle v, Lv\rangle_\rho} = \frac{\langle L^{1/2}v, L^{-1/2}u\rangle_\rho^2}{\langle v, Lv\rangle_\rho} \leqslant \langle u, L^{-1}u\rangle_\rho, \tag{3.17}$$

and the equality is attained if and only if v obeys Eq. (3.9). ■

Within the Gill-Levit family, $\mathsf{B}_{\max}$ is not only the maximum but also the closest in spirit to the local bound given by Eq. (1.3), with the L^{-1} operator generalizing the role of F^{-1}. Moreover, note that $\mathsf{B}_{\max}$ is naturally invariant. Although it is also possible to derive $\mathsf{B}_{\max}$ starting from Eqs. (1.7)–(1.10) without the invariant formalism, at least for a flat metric, the invariance of $\mathsf{B}_{\max}$ would have been much more tedious to prove, with a proliferation of Jacobians.

The most difficult part of computing $\mathsf{B}_{\max}$ is solving for $L^{-1}u$. Let $v = L^{-1}u$, which is a least favorable field. It obeys the second-order field equation

$$(Lv)_a = nF_{ab}v^b - \nabla_a\left[\frac{1}{\rho}\nabla_b(\rho v^b)\right] = u_a. \tag{3.18}$$

The solution, expressible in terms of an impulse-response (Green) function, can be substituted into Eq. (3.6) to give $\mathsf{B}_{\max}$. For large n, Eq. (3.18) can be simplified to

$$nF_{ab}v^b \approx u_a, \quad \mathsf{B}_{\max} \approx \frac{\langle \mathsf{C}\rangle}{n}, \tag{3.19}$$

so Eq. (3.5) is asymptotically least favorable to the Gill-Levit family, in nice agreement with the local theory [28] and earlier results [14]. Note, however, that the exact optimal choice according to Eq. (3.18) also depends on the prior and some derivatives. The correction to the local theory becomes especially important if u is not in the range of F and Eq. (3.5) has no solution. The question of what to do when u is not even in the range of L, and Eq. (3.18) has no solution, remains open.

Another special case is when a parametrization with $g_{ab} = \delta_{ab}$ is assumed, u and F are θ independent, and π is Gaussian with covariance matrix G^{-1}. Then the solution to Eq. (3.18) is

$$v^a = [(nF + G)^{-1}]^{ab}u_b, \tag{3.20}$$

and $\mathsf{B}_{\max}$ becomes

$$\mathsf{B}_{\max} = u_a[(nF + G)^{-1}]^{ab}u_b, \tag{3.21}$$

which coincides with the Schützenberger–Van Trees version given by Eq. (2.22), since $\langle u\rangle = u$, $\langle F\rangle = F$, and $\langle G\rangle = G$ in this case. Furthermore, if $f(x|\theta) = f(x-\theta)$ and $f(x-\theta)$ is also Gaussian, such that F is the inverse of the covariance matrix of f, then it is well known that the minimum Bayes risk $\min_{\check{\beta}}\langle \mathsf{R}\rangle$ is also given by the right-hand side of Eq. (3.21) [7], and $\mathsf{B}_{\max}$ is a tight bound.

IV. WAVE PICTURE

I now switch gears and make the substitution

$$\rho = \psi^2, \tag{4.1}$$

where ψ is a real function of the parameter. I call ψ a wave function. All the functionals in Eqs. (2.25)–(2.27) turn out to be quadratic with respect to ψ and $\nabla_a\psi$, given by

$$\langle \mathsf{A}\rangle = \int (v^a u_a)\psi^2\epsilon, \tag{4.2}$$

$$\langle \mathsf{F}\rangle = \int (v^a F_{ab}v^b)\psi^2\epsilon, \tag{4.3}$$

$$\langle \mathsf{P} \rangle = \int (\mathsf{D}\psi)^2 \epsilon, \tag{4.4}$$

$$\mathsf{D}\psi \equiv (\nabla_a v^a)\psi + 2v^a \nabla_a \psi. \tag{4.5}$$

The problem of choosing an unfavorable prior to tighten the bound for minimax estimation now becomes a problem of finding the wavefunction that maximizes B. To simplify, I define yet another inner product as

$$\langle \psi, \phi \rangle \equiv \int \psi \phi \epsilon. \tag{4.6}$$

The normalization condition for the prior density becomes

$$\int \rho \epsilon = \langle \psi, \psi \rangle = 1. \tag{4.7}$$

It can be shown that

$$\langle \mathsf{A} \rangle = \langle \psi, \mathsf{A}\psi \rangle, \tag{4.8}$$

$$\langle \mathsf{F} \rangle = \langle \psi, \mathsf{F}\psi \rangle, \tag{4.9}$$

$$\langle \mathsf{P} \rangle = \langle \mathsf{D}\psi, \mathsf{D}\psi \rangle = \langle \psi, \mathsf{D}^\dagger \mathsf{D}\psi \rangle, \tag{4.10}$$

$$\mathsf{D}^\dagger \psi = (\nabla_a v^a)\psi - 2v^a \nabla_a \psi, \tag{4.11}$$

$$\mathsf{B} = \frac{\langle \psi, \mathsf{A}\psi \rangle^2}{\langle \psi, \mathsf{H}\psi \rangle}, \tag{4.12}$$

$$\mathsf{H} \equiv n\mathsf{F} + \mathsf{D}^\dagger \mathsf{D}. \tag{4.13}$$

Note that D may be a nonlinear operator, if the choice of v, such as Eq. (3.18), depends on the prior. To proceed, I assume that v does not depend on ψ and D is linear. Then I can follow the approach in Sec. III to obtain

$$\mathsf{B} = \frac{\langle \mathsf{H}^{1/2}\psi, \mathsf{H}^{-1/2}\mathsf{A}\psi \rangle^2}{\langle \psi, \mathsf{H}\psi \rangle} \leqslant \langle \mathsf{A}\psi, \mathsf{H}^{-1}\mathsf{A}\psi \rangle. \tag{4.14}$$

The equality is attained if and only if

$$\mathsf{H}\psi = (n\mathsf{F} + \mathsf{D}^\dagger \mathsf{D})\psi = \lambda \mathsf{A}\psi, \tag{4.15}$$

where λ is an arbitrary nonzero real number. Let ψ_λ be a solution of Eq. (4.15) as a function of λ, subject to the normalization constraint given by Eq. (4.7). Then

$$\mathsf{B} = \frac{1}{\lambda} \langle \psi_\lambda, \mathsf{A}\psi_\lambda \rangle, \tag{4.16}$$

and this expression should be maximized with respect to λ to obtain the tightest lower bound on $\sup_\theta \mathsf{R}(\theta)$.

A substantial simplification can be made if a parametrization with $g_{ab} = \delta_{ab}$ can be assumed and u, v, and therefore A are θ independent. Equation (4.15) becomes

$$[n\mathsf{F}(\theta) - 4(v^a \partial_a)^2]\psi(\theta) = \lambda \mathsf{A}\psi(\theta), \tag{4.17}$$

which is a time-independent Schrödinger equation. The Fisher information $\mathsf{F} = v^a F_{ab} v^b$, evaluated in the direction of v, plays the role of the potential, while $-(v^a \partial_a)^2$, in terms of the directional derivative $v^a \partial_a$, plays the role of the kinetic-energy operator. The bound becomes

$$\mathsf{B} = \frac{\mathsf{A}^2}{\langle \psi, \mathsf{H}\psi \rangle}. \tag{4.18}$$

To maximize B, one should therefore solve for

$$\mathsf{B}_{\text{worst}} \equiv \sup_{\psi:\langle \psi,\psi\rangle=1} \mathsf{B} = \frac{\mathsf{A}^2}{\mathsf{E}_{\min}}, \tag{4.19}$$

$$\mathsf{E}_{\min} \equiv \inf_{\psi:\langle \psi,\psi\rangle=1} \langle \psi, \mathsf{H}\psi \rangle, \tag{4.20}$$

that is, the ground-state energy. The infimum is used here in case a normalizable ground state does not exist. Adding a phase to the wave function cannot reduce the energy, so the consideration of only real wave functions is justified here.

The wave correspondence makes sense, as intuition suggests that an unfavorable prior should be concentrated near the minimum of the Fisher information, just as the ground state should be concentrated near the bottom of the potential. If the prior density is made too sharp, however, the prior information $\langle \mathsf{P} \rangle$ would become large, and therefore a balance between $n\langle \mathsf{F} \rangle$ and $\langle \mathsf{P} \rangle$ should be struck to minimize their sum, just as the ground state achieves the optimal balance between the potential and the kinetic energies.

In the limit $n \to \infty$, the ground-state energy is the classical-mechanics limit given by

$$\mathsf{E}_{\min} = n \inf_{\theta \in \Theta} \mathsf{F}(\theta) + o(n), \tag{4.21}$$

where $o(g(n))$ denotes a term of a smaller order than $g(n)$ as $n \to \infty$. Other asymptotic notations [33], including $\Theta(g(n))$ [same order as $g(n)$] and $\Omega(g(n))$ [order at least as large as $g(n)$], are also used in the following. If the infimum of $\mathsf{F}(\theta)$ is strictly positive, $\mathsf{B}_{\text{worst}}$ obeys the parametric rate $\Theta(n^{-1})$. A more interesting case is when the infimum is 0, $\mathsf{E}_{\min} = o(n)$, and the bound mandates a convergence rate slower than the parametric rate. A concrete special case is as follows.

Theorem 2. Suppose that u and v are θ independent and obey $v^a u_a \neq 0$. Suppose also that there exists a one-dimensional submodel with parametrization

$$\theta^a(\tau) = \theta^a(0) + v^a \tau, \tag{4.22}$$

$\tau \in (\tau_1, \tau_2) \subseteq \mathbb{R}$, $\tau_1 \leqslant 0 \leqslant \tau_2$, $\tau_1 \neq \tau_2$, and Fisher information bounded by

$$\mathsf{F}(\tau) = v^a F_{ab}(\theta(\tau)) v^b \leqslant A|\tau|^m, \tag{4.23}$$

where A is a positive constant and $m \geqslant 0$. Then

$$\sup_{\theta \in \Theta} \mathsf{R}(\theta) \geqslant \mathsf{B}_{\text{worst}} = \Omega(n^{-2/(m+2)}). \tag{4.24}$$

Proof. With the given conditions, the average energy for the submodel is

$$\langle \psi, \mathsf{H}\psi \rangle = \int_{\tau_1}^{\tau_2} \left\{ n\mathsf{F}(\tau)\psi(\tau)^2 + \left[\frac{\partial \psi(\tau)}{\partial \tau} \right]^2 \right\} d\tau. \tag{4.25}$$

Let $\psi(\tau) = \phi(\tau/W)/\sqrt{W}$, where ϕ is a trial function and $0 < W \leqslant 1$ scales the width of ψ. Then

$$\langle \psi, \mathsf{H}\psi \rangle \leqslant nAW^m \int_{\tau_1/W}^{\tau_2/W} \phi(y)^2 |y|^m dy + \frac{4}{W^2} \int_{\tau_1/W}^{\tau_2/W} \left[\frac{\partial \phi(y)}{\partial y} \right]^2 dy \tag{4.26}$$

$$= nAW^m \int_{\tau_1}^{\tau_2} \phi(y)^2 |y|^m dy + \frac{4}{W^2} \int_{\tau_1}^{\tau_2} \left[\frac{\partial \phi(y)}{\partial y}\right]^2 dy, \tag{4.27}$$

where the last step uses the fact that $\tau_1/W \leqslant \tau_1$ and $\tau_2/W \geqslant \tau_2$, since $\tau_1 \leqslant 0 \leqslant \tau_2$ and $0 < W \leqslant 1$, and $\phi(y)$ vanishes outside (τ_1, τ_2). It is not difficult to show that, regardless of τ_1 and τ_2, there always exists a trial function that makes both integrals in Eq. (4.27) converge. Minimizing Eq. (4.27) with respect to W, I obtain

$$W = A_1 n^{-1/(m+2)}, \tag{4.28}$$

$$\langle \psi, \mathsf{H}\psi \rangle \leqslant A_2 n^{2/(m+2)}, \tag{4.29}$$

where A_1 and A_2 are positive constants. For a large enough n, the assumption $W \leqslant 1$ can be satisfied. The theorem then follows from Eqs. (1.7), (1.11), (4.19), (4.20), and (4.29). ■

A concrete example is $\mathsf{F}(\tau) \leqslant A\tau^2$, in which case we can borrow from the theory of quantum harmonic oscillators to find that the ground-state energy for a potential $nA\tau^2$ is $\Theta(n^{1/2})$, leading to $\mathsf{B}_{\text{worst}} = \Omega(n^{-1/2})$.

V. QUANTUM ESTIMATION THEORY

A. Basics

Assume $n = 1$ without loss of generality. Let $\{\varrho(\theta) : \theta \in \Theta\}$ be a family of density operators that model a quantum system. The generalized Born's rule states that the statistics of any measurement of the system can be modeled by a positive operator-valued measure (POVM) E [19] via

$$f(x|\theta) d\mu(x) = \text{tr}\,[dE(x)\varrho(\theta)], \tag{5.1}$$

where tr denotes the operator trace. For any POVM, an upper bound on the Fisher information is given by [19,34]

$$\mathsf{F} = v^a F_{ab} v^b \leqslant v^a K_{ab} v^b \equiv \mathsf{K} \tag{5.2}$$

for any vector v, where K is the Helstrom information matrix [18], defined as

$$K_{ab}(\theta) \equiv \text{tr}\,[\varrho(\theta)\mathcal{S}_a(\theta) \circ \mathcal{S}_b(\theta)], \tag{5.3}$$

$A \circ B \equiv (AB + BA)2$ denotes the Jordan product, and $\mathcal{S}_a$, a score operator, is a solution to

$$\partial_a \varrho(\theta) = \varrho(\theta) \circ \mathcal{S}_a(\theta). \tag{5.4}$$

There exist other quantum versions of the Fisher information and the Cramér-Rao bound that are of interest when β is vectoral [19,35–37], but they are outside the scope of this work, and I focus on the Helstrom information hereafter.

With Eq. (5.2), a quantum lower bound on B for any POVM can be obtained simply by replacing F with K. To be explicit,

$$\langle \mathsf{R} \rangle \geqslant \mathsf{B} \geqslant \mathsf{Q} \equiv \frac{\langle \mathsf{A} \rangle^2}{\langle \mathsf{K} \rangle + \langle \mathsf{P} \rangle}. \tag{5.5}$$

For B to attain Q, the equality in Eq. (5.2) must hold for all $\theta \in \Theta$, and that is usually not possible.

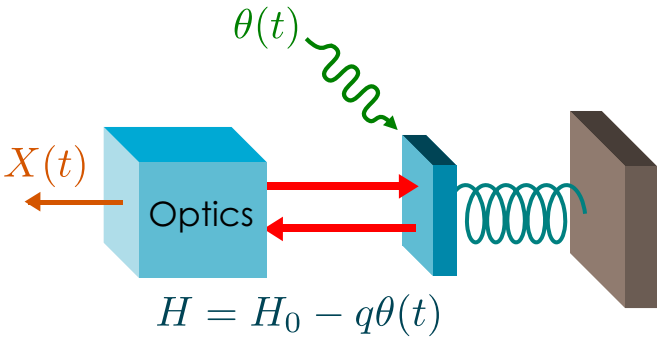

FIG. 2. An optomechanical force sensor under continuous optical measurements. $\theta(t)$ is the unknown classical force, H is the system Hamiltonian, and $X(t)$ is the observation process.

As K is also a positive-semidefinite (0,2) tensor, all the results in the previous sections apply to the quantum bound as well. In particular, following Theorem 1, the optimal Q is

$$\mathsf{Q}_{\max} \equiv \max_v \mathsf{Q} = \langle u, R^{-1} u \rangle_\rho, \tag{5.6}$$

$$(Rv)_a \equiv K_{ab} v^b - \nabla_a \left[\frac{1}{\rho} \nabla_b (\rho v^b)\right]. \tag{5.7}$$

It is not difficult to prove that

$$\mathsf{Q}_{\max} \leqslant \mathsf{B}_{\max} \tag{5.8}$$

for any POVM. A simple example is the quantum Gaussian shift model, where $\varrho(\theta)$ is the quantum state of m harmonic oscillators with a Gaussian Wigner representation and $\theta \in \mathbb{R}^p$, with $p = 2m$, is its displacement in phase space [36,38]. Assuming the standard parametrization with $g_{ab} = \delta_{ab}$, K is the inverse of the covariance matrix of the Wigner function and θ independent. By measuring the object together with an auxiliary in a Gaussian state with the same covariance matrix, it is possible to produce classical Gaussian shift statistics that achieves $F = K/2$ [39]. If π is also Gaussian with covariance matrix G^{-1} and u is θ independent, then, by the same rationale that gives Eq. (3.21), it can be shown that

$$\mathsf{Q}_{\max} = u^\top (K + G)^{-1} u, \tag{5.9}$$

and for the measurement just mentioned,

$$\min_{\check{B}} \langle \mathsf{R} \rangle = \mathsf{B}_{\max} = u^\top (K/2 + G)^{-1} u, \tag{5.10}$$

$$\mathsf{Q}_{\max} \leqslant \min_{\check{B}} \langle \mathsf{R} \rangle \leqslant 2\mathsf{Q}_{\max}. \tag{5.11}$$

A further optimization of the measurement for a given u may be possible, but the optimization problem becomes more difficult in general, especially when u is θ dependent or π is non-Gaussian.

B. Waveform estimation

Consider a quantum dynamical system, such as the optomechanical force sensor depicted in Fig. 2, under the influence of a classical waveform $\theta(t)$. Using the principles of a larger Hilbert space and deferred measurements [40], the statistics of a sequentially measured quantum system can be

modeled by a POVM at the final time and a density-operator family given by

$$\varrho(\theta) = U(\theta)\,|\Psi\rangle\,\langle\Psi|\,U(\theta)^\dagger, \tag{5.12}$$

$$U(\theta) = \mathcal{T}\exp\left\{\frac{1}{i\hbar}\int_{-T/2}^{T/2}[H_0(t) - q\theta(t)]dt\right\}, \tag{5.13}$$

where $|\Psi\rangle$ is the initial state of the quantum system, q is a position operator, $H_0(t)$ is the rest of the Hamiltonian, T is the total observation time, and $\mathcal{T}$ denotes time ordering of the operator exponential.

Let the parameter of interest be defined in terms of a weight function $h(t)$ as

$$\beta = \int_{-T/2}^{T/2} h(t)\theta(t)dt. \tag{5.14}$$

For example, if $\beta = \theta(\tau)$ at an instant of time τ is of interest, then $h(t) = \delta(t-\tau)$. To derive analytic results, I follow Ref. [21] and discretize time as

$$t_a = -\frac{T}{2} + a\delta t, \quad T = p\delta t. \tag{5.15}$$

Assuming

$$\theta(t_a) = \theta^a, \quad h(t_a) = h_a, \tag{5.16}$$

and

$$\beta \approx h_a\theta^a\delta t, \tag{5.17}$$

$$U \approx U(t_p, t_1) \equiv \exp\left[\frac{H_0(t_p)\delta t}{i\hbar}\right]\exp\left(\frac{iq\theta^p\delta t}{\hbar}\right)\cdots \exp\left[\frac{H_0(t_1)\delta t}{i\hbar}\right]\exp\left(\frac{iq\theta^1\delta t}{\hbar}\right), \tag{5.18}$$

it can be shown that

$$u_a \approx \partial_a(h_a\theta^a\delta t) = h_a\delta t, \tag{5.19}$$

$$K_{ab} \approx \frac{4\delta t^2}{\hbar^2}C_q(t_a, t_b), \tag{5.20}$$

$$C_q(t_a, t_b) \equiv \langle\Psi|\,\hat{q}(t_a)\circ\hat{q}(t_b)\,|\Psi\rangle - \langle\Psi|\,\hat{q}(t_a)\,|\Psi\rangle\,\langle\Psi|\,\hat{q}(t_b)\,|\Psi\rangle\,, \tag{5.21}$$

where

$$\hat{q}(t_a) \equiv U(t_{a-1}, t_1)^\dagger q U(t_{a-1}, t_1) \tag{5.22}$$

is the Heisenberg picture of q, C_q is its covariance function, and the right-hand side of Eq. (5.20) is the exact Helstrom information for $\varrho(\theta) = U(t_p, t_1)\,|\Psi\rangle\,\langle\Psi|\,U(t_p, t_1)^\dagger$. If $\hat{q}(t)$ is stationary, the covariance can be written in terms of a power spectral density $S_q(\omega)$ as [20]

$$C_q(t_a, t_b) = \int_{-\infty}^{\infty} S_q(\omega)\exp\left[i\omega(t_b - t_a)\right]\frac{d\omega}{2\pi}. \tag{5.23}$$

With the assumption of stationary processes and a long observation time (SPLOT) [7], K can be approximated as a circulant matrix [41] and expressed as

$$K_{ab} \approx \frac{\delta t}{p}\sum_{j=0}^{p-1}\frac{4S_q(\omega_j)}{\hbar^2}\exp\left[i\omega_j(t_b - t_a)\right], \tag{5.24}$$

where $\omega_j = \omega_0 + 2\pi j/T$ and $\omega_0 = -\pi/\delta t$. Similarly, if $\theta(t)$ is a stationary Gaussian random process with power spectral density $S_\theta(\omega)$,

$$G_{ab} \approx \frac{\delta t}{p}\sum_{j=0}^{p-1}\frac{1}{S_\theta(\omega_j)}\exp\left[i\omega_j(t_b - t_a)\right]. \tag{5.25}$$

As $V_{ja} \equiv \exp(-i\omega_j t_a)/\sqrt{p}$ is a unitary matrix, the inverse of $K + G$ can be computed analytically to give

$$[(K+G)^{-1}]^{ab} \approx \frac{1}{T}\sum_{j=0}^{p-1}\frac{\exp\left[i\omega_j(t_a - t_b)\right]}{4S_q(\omega_j)/\hbar^2 + 1/S_\theta(\omega_j)}. \tag{5.26}$$

u, as given by Eq. (5.19), does not depend on θ. If the dynamics of the system is linear [20], K also does not depend on θ. Thus, the same argument that leads to Eq. (3.21) can be used to give

$$\mathsf{Q}_{\max} = u_a[(K+G)^{-1}]^{ab}u_b, \tag{5.27}$$

$$\approx \frac{1}{T}\sum_{j=0}^{p-1}\frac{\delta t^2 h_a h_b\exp\left[i\omega_j(t_a - t_b)\right]}{4S_q(\omega_j)/\hbar^2 + 1/S_\theta(\omega_j)}. \tag{5.28}$$

Taking the continuous and long-time limit with $\delta t \to 0$, $T \to \infty$, and $d\omega = 2\pi/T$ hence results in

$$\mathsf{Q}_{\max} \to \int_{-\infty}^{\infty}\frac{|\tilde{h}(\omega)|^2}{4S_q(\omega)/\hbar^2 + 1/S_\theta(\omega)}\frac{d\omega}{2\pi}, \tag{5.29}$$

$$\tilde{h}(\omega) \equiv \int_{-\infty}^{\infty} h(t)\exp(-i\omega t)dt. \tag{5.30}$$

If $\beta = \theta(\tau)$ with $h(t) = \delta(t-\tau)$ and $|\tilde{h}(\omega)| = 1$, Eq. (5.29) agrees with the result in Ref. [21]. Compared with Ref. [21], which derives a quantum bound on $\langle\mathsf{R}\rangle$ directly, the derivation here clarifies the relation of Eq. (5.29) to the Helstrom information and the Gill-Levit formalism. The new insight implied by the theory here is that the bound remains invariant upon any reparametrization and cannot be further improved by picking a different v.

While Eq. (5.29) holds for any measurement, it can say something more about measurements in the linear form of

$$X(t) = \int_{-\infty}^{\infty} h_X(t - t')\theta(t')dt' + Z(t), \tag{5.31}$$

where h_X is an impulse-response function of the system and Z is a stationary noise process that is uncorrelated with θ. In optomechanics, such a process can be obtained by homodyne detection of the output light. Let the estimator be

$$\check{\beta} = \int_{-\infty}^{\infty}\check{h}(t)X(t)dt, \tag{5.32}$$

where $\check{h}(t)$ is a linear filter, or more precisely a smoother in control-theoretic terminology, as it is applied to the whole observation record to estimate the waveform at an intermediate time [42]. By standard Wiener filtering theory [7], the minimum mean-square risk in the SPLOT limit is

$$\langle\mathsf{R}\rangle \to \int_{-\infty}^{\infty}\frac{|\tilde{h}(\omega)|^2}{|\tilde{h}_X(\omega)|^2/S_Z(\omega) + 1/S_\theta(\omega)}\frac{d\omega}{2\pi}, \tag{5.33}$$

$$\tilde{h}_X(\omega) \equiv \int_{-\infty}^{\infty} h_X(t)\exp(-i\omega t)dt, \tag{5.34}$$

where S_Z is the power spectral density of Z. Comparing Eqs. (5.29) and (5.33), one sees that $\langle \mathsf{R} \rangle \geqslant \mathsf{Q}_{\max}$ implies

$$\frac{S_Z(\omega)}{|\tilde{h}_X(\omega)|^2} \geqslant \frac{\hbar^2}{4S_q(\omega)}, \tag{5.35}$$

which serves as a fundamental quantum limit on the noise floor. To reach this limit for an optomechanical system, backaction evasion and quantum-limited measurements are necessary [21]. It is possible to derive alternative quantum limits in terms of the optics by appealing to the interaction picture and tighter limits that account for loss by choosing the purification of the quantum state judiciously [43]. Reference [44] reports an experimental demonstration of mirror-motion estimation close to such quantum limits.

It is noteworthy that, prior to Ref. [21], Braginsky and coworkers derived an expression similar to Eq. (5.20) by optimizing a signal-to-noise ratio (SNR) in terms of an observable [20]. A spectral form of their optimal SNR, derived from a heuristic energy-time uncertainty relation, can be found in Ref. [45]. They called their results the energetic quantum limit. The similarity is not a coincidence, as the Helstrom information can also be expressed as the solution to the optimization problem

$$\mathsf{K} = \max_Y \frac{(v^a \partial_a \bar{Y})^2}{\mathrm{tr}(Y-\bar{Y})^2 \varrho}, \tag{5.36}$$

$$\bar{Y} \equiv \mathrm{tr}\, Y \varrho, \tag{5.37}$$

where Y is any observable and the right-hand side of Eq. (5.36) is similar to the SNR studied in Ref. [20]. Equation (5.36) can be proved by applying the Cauchy-Schwarz inequality to $(v^a\partial_a \bar{Y})^2 = (\mathrm{tr}\, Y v^a \partial_a \varrho)^2 = [\mathrm{tr}(Y-\bar{Y}) v^a \partial_a \varrho]^2 = \{\mathrm{tr}[(Y-\bar{Y}) \circ (v^a \mathcal{S}_a)]\varrho\}^2 \leqslant [\mathrm{tr}(Y-\bar{Y})^2\varrho][\mathrm{tr}(v^a\mathcal{S}_a)^2\varrho]$. While their results are seminal and capture the basic physics, the results here and in Ref. [21] are more precise in terms of meaning. The SNR does not have a direct operational meaning in statistics, whereas here the statistical problem is clearly defined in terms of a mean-square risk, and the bound is proven to hold for any POVM and any biased or unbiased estimator, not just observables. The clear definition of a risk is important, as different problems have different types of risk and different optimal measurements, and no single SNR-based treatment can deal with all of them. For example, while a linear measurement in the form of Eq. (5.31) can achieve the optimal SNR and also optimal waveform estimation, more careful studies reveal that it is suboptimal with respect to the quantum limits for waveform detection [46] and spectrum parameter estimation [47], and photon-counting measurements can perform much better for those problems.

Equation (5.29) demonstrates the importance of prior information in the form of $1/S_\theta(\omega)$, as the integral may not converge without it; see Ref. [48] for an example in optical phase estimation. If $\beta = \theta(\tau)$, Eqs. (5.29) and (5.33) are steady-state values that do not scale with T. This is an extreme example where the i.i.d. condition does not hold, the standard asymptotic theory [19,28] fails, the convergence rate is slower than the parametric rate, and prior information is indispensable. The information that can be acquired in one time slot with duration δt is infinitesimal, but a finite risk can still be achieved because there exist prior correlations in $\theta(t)$ across different times before and after $t = \tau$, meaning that information over multiple time slots can contribute to the estimation of each $\theta(\tau)$. This intuition explains why the optimal estimator is a smoother.

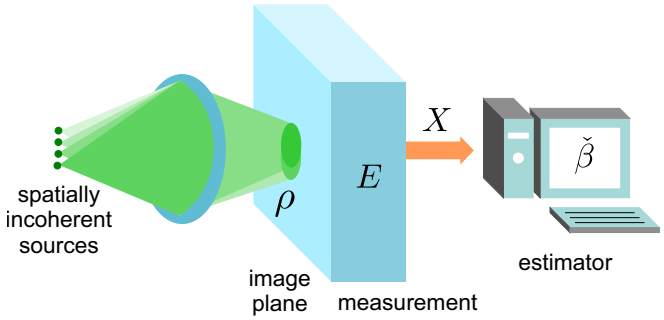

FIG. 3. Basic setup of the optical imaging problem.

C. Subdiffraction incoherent optical imaging

For another application of quantum estimation theory, consider the far-field paraxial imaging of p spatially incoherent and equally bright point sources [49], as depicted in Fig. 3. On the image plane, the density operator of each photon can be modeled as [22]

$$\varrho(\theta) = \frac{1}{p}\sum_{a=1}^{p} \exp(-ik\theta^a)\,|\Psi\rangle\langle\Psi|\exp(ik\theta^a), \tag{5.38}$$

$$|\Psi\rangle = \int_{-\infty}^{\infty} dx \Psi(x)\,|x\rangle\,, \tag{5.39}$$

where θ is a vector of the unknown source positions on the object plane that is assumed to be one-dimensional for simplicity, $|x\rangle$ is the Dirac eigenket for the image-plane photon position that obeys $\langle x|x'\rangle = \delta(x-x')$, with an image-plane coordinate x that is normalized with respect to the magnification factor, Ψ is the point-spread function of the imaging system for the optical field, and k is the momentum operator.

Direct imaging can be modeled as a measurement of each photon in the position basis [22]. The probability density of each observed position is then

$$f(x|\theta) = \langle x|\,\varrho(\theta)\,|x\rangle = \frac{1}{p}\sum_{a=1}^{p} h(x-\theta^a), \tag{5.40}$$

$$h(x) \equiv |\Psi(x)|^2. \tag{5.41}$$

The Fisher information is

$$v^a F_{ab}(\theta) v^b = \int_{-\infty}^{\infty} \frac{[v^a \partial_a h(x-\theta^a)]^2}{p^2 f(x|\theta)} dx. \tag{5.42}$$

In particular, at $\theta = 0$,

$$v^a F_{ab}(0) v^b = (v^a w_a)^2 \int_{-\infty}^{\infty} \frac{1}{h(x)}\left[\frac{\partial h(x)}{\partial x}\right]^2 dx, \tag{5.43}$$

$$w_a = \frac{1}{p}, \quad a = 1, \ldots, p. \tag{5.44}$$

The kernel of $F(0)$ is then the $(p-1)$–dimensional space

$$\ker[F(0)] = \{v \in \mathbb{R}^p : v^a w_a = 0\}, \tag{5.45}$$

while the range is the one-dimensional space

$$\text{range}[F(0)] = \{cw : c \in \mathbb{R}\}. \tag{5.46}$$

Assume hereafter that β is a linear function of θ, such that u is θ independent. For any β with $u \notin \text{range}[F(0)]$, a $v \in \ker[F(0)]$ can always be found such that $v^a u_a \neq 0$ but $v^a F_{ab}(0) v^b = 0$. Section IV then implies that, from the minimax perspective, any estimator of this β must have a convergence rate slower than the parametric rate with respect to n detected photons. Only a β with $u \in \text{range}[F(0)]$ has a nonzero information at $\theta = 0$ for any v with $v^a u_a \neq 0$. This β is proportional to the object centroid $w_a \theta^a = (\sum_a \theta^a)/p$, and the parametric rate is indeed possible by taking the sample mean of the photon positions, provided that h has a finite variance [50].

For $p = 2$, other than the centroid, the second parameter may be taken as the separation $|\theta^2 - \theta^1|$ between the two sources. Reference [17] uses a special case of Theorem 2 to prove that, since the exponent of the Fisher information is $m = 2$ for $u \propto v \propto (1, -1)$, a limit on the convergence rate is $\mathsf{B}_{\text{worst}} = \Omega(n^{-1/2})$. This rate is also observed numerically in Refs. [17,51]. Paúr and coworkers showed that the exponent can be improved to $m = 1$ if the point-spread function has zeros [24], and the limit becomes $\mathsf{B}_{\text{worst}} = \Omega(n^{-2/3})$ according to Theorem 2.

The Helstrom information turns out to be much higher [22,25]. For n detected photons and i.i.d. quantum states, the Helstrom information is simply n times that for one photon [19]. For $p = 2$, $K(\theta)$ turns out to be full-rank [22], and separation estimation at the parametric rate is also possible via spatial-mode demultiplexing [17]. For $p \geqslant 2$, Bisketzi and coworkers found that $K(\theta)$ has a rank of 2 as $\theta \to 0$ [25]. Then Sec. IV implies that any β with a $u \notin \text{range}[K(0)]$ cannot be estimated at the parametric rate by any measurement, and only a β with u in the two-dimensional range may be estimated at the parametric rate.

VI. CONCLUSION

Compared with the local theory, the use of Bayesian Cramér-Rao bounds has been less systematic in the literature and often relied on the ingenuity of the researcher to pick the appropriate form. This work resolves some of the ambiguities and hopefully inspires further progress via the physics connections.

The formalism here looks ripe for a generalization for infinite-dimensional parameter spaces in a manner similar to the local theory [2,3,9]. An important application would be to derive semiparametric bounds with slow convergence rates [5] in a more systematic fashion.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation (NRF) Singapore under its Quantum Engineering Programme (Grant No. QEP-P7).

APPENDIX: VECTORAL PARAMETER OF INTEREST

Here I generalize the fundamental results in Secs. II–V for a vector parameter of interest $\beta(\theta) = (\beta^1, \ldots, \beta^q) \in \mathbb{R}^q$ with $1 \leqslant q \leqslant p$. Define the mean-square risk as

$$\mathsf{R}(\theta) \equiv \int [\check{\beta}^j(x) - \beta^j(\theta)] \gamma_{jk}(\theta) [\check{\beta}^k(x) - \beta^k(\theta)] \times f^{(n)}(x|\theta) d\mu^{(n)}(x), \tag{A1}$$

where γ is a positive-definite weight matrix. For clarity, indices starting from j are used to label the components of β, to be distinguished from indices that start from a for the components of θ. The Bayesian risk is

$$\langle \mathsf{R} \rangle = \int \mathsf{R}(\theta) \pi(\theta) d^p\theta. \tag{A2}$$

Define

$$u^j_a \equiv \partial_a \beta^j. \tag{A3}$$

The Gill-Levit bounds B still have the form of Eq. (1.7), but now [5]

$$\mathsf{A} \equiv v^a_j u^j_a, \tag{A4}$$

$$\mathsf{F} \equiv \gamma^{jk} v^a_j F_{ab} v^b_k, \tag{A5}$$

$$\mathsf{P} \equiv \gamma^{jk} \left[\frac{1}{\pi} \partial_a (\pi v^a_j)\right] \left[\frac{1}{\pi} \partial_b (\pi v^b_k)\right], \tag{A6}$$

where v now has $q \times p$ entries and

$$\gamma^{jk} \equiv (\gamma^{-1})^{jk}. \tag{A7}$$

$\{\gamma^{jk}\}$ are the entries of the B matrix in Ref. [5], while $\{v^a_j\}$ are the entries of the C matrix in Ref. [5].

Upon reparametrization of θ, γ should remain invariant, in the sense of $\gamma(\theta) = \tilde{\gamma}(\tilde{\theta}(\theta))$, so that the statistical problem remains unchanged.

It is straightforward to generalize Proposition 1.

Proposition 3. B is invariant under reparametrization if v_j for each j obeys the transformation law

$$v^a_j J^b_a = \tilde{v}^b_j. \tag{A8}$$

Proof. The proof is almost identical to that of Proposition 1 and omitted for brevity. ∎

For a manifold Θ, a generalization of Proposition 2 is as follows.

Proposition 4. If ρv vanishes on any boundary of Θ, the Bayesian mean-square risk has a lower bound given by Eq. (1.7), where

$$\langle \mathsf{A} \rangle \equiv \int (v^a_j u^j_a) \rho \epsilon, \tag{A9}$$

$$\langle \mathsf{F} \rangle \equiv \int (\gamma^{jk} v^a_j F_{ab} v^b_k) \rho \epsilon, \tag{A10}$$

$$\langle \mathsf{P} \rangle \equiv \int \gamma^{jk} \left[\frac{1}{\rho} \nabla_a (\rho v^a_j)\right] \left[\frac{1}{\rho} \nabla_b (\rho v^b_k)\right] \rho \epsilon. \tag{A11}$$

Proof. Let

$$\mathsf{b}^j \equiv \int (\check{\beta}^j - \beta^j) f^{(n)} d\mu^{(n)}. \tag{A12}$$

By the Leibniz rule,

$$\int \nabla_a\big(\mathsf{b}^j \rho v_j^a\big)\epsilon = \iint (\check{\beta}^j - \beta^j)\nabla_a\big(f^{(n)}\rho v_j^a\big)d\mu^{(n)}\epsilon - \int \big(v_j^a \nabla_a \beta^j\big)\rho\epsilon. \tag{A13}$$

The left-hand side is 0 by the Stokes theorem, if ρv vanishes on any boundary of Θ. Then

$$\langle \mathsf{A}\rangle = \mathbb{E}[(\check{\beta}^j - \beta^j)\gamma_{jk}s^k], \tag{A14}$$

$$s^j \equiv \frac{\gamma^{jk}}{f^{(n)}\rho}\nabla_a\big(f^{(n)}\rho v_k^a\big). \tag{A15}$$

Considering the right-hand side of Eq. (A14) as an inner product between $(\check{\beta} - \beta)$ and s that is weighted by γ and applying the Cauchy-Schwarz inequality, I obtain

$$\langle \mathsf{A}\rangle^2 \leqslant \langle \mathsf{R}\rangle \mathbb{E}\big(s^j\gamma_{jk}s^k\big). \tag{A16}$$

Standard procedures then lead to Eqs. (1.7) and (A9)–(A11). ■

A generalization of Theorem 1 is as follows.

Theorem 3.

$$\max_v \mathsf{B} = \langle u, L^{-1}u\rangle_\rho \equiv \mathsf{B}_{\max}, \tag{A17}$$

where the inner product is defined as

$$\langle v, u\rangle_\rho \equiv \int \big(v_j^a u_a^j\big)\rho\epsilon, \tag{A18}$$

the linear, self-adjoint, and positive-semidefinite operator L is defined as

$$(Lv)_a^j \equiv n\gamma^{jk}F_{ab}v_k^b - \nabla_a\left[\frac{\gamma^{jk}}{\rho}\nabla_b\big(\rho v_k^b\big)\right], \tag{A19}$$

and u is assumed to be in the range of L, such that $L^{-1}u$ exists. A least favorable v that maximizes B must satisfy

$$v \propto L^{-1}u. \tag{A20}$$

Proof. The proof is similar to that of Theorem 1 and omitted for brevity. ■

With the substitution $\rho = \psi^2$, Eqs. (4.2)–(4.5) can be generalized to

$$\langle \mathsf{A}\rangle = \int \big(v_j^a u_a^j\big)\psi^2\epsilon, \tag{A21}$$

$$\langle \mathsf{F}\rangle = \int \big(\gamma^{jk}v_j^a F_{ab}v_k^b\big)\psi^2\epsilon, \tag{A22}$$

$$\langle \mathsf{P}\rangle = \int \gamma^{jk}(\mathsf{D}_j\psi)(\mathsf{D}_k\psi)\epsilon, \tag{A23}$$

$$\mathsf{D}_j\psi \equiv \big(\nabla_a v_j^a\big)\psi + 2v_j^a\nabla_a\psi, \tag{A24}$$

while Eqs. (4.10), (4.11), and (4.13) can be generalized to

$$\langle \mathsf{P}\rangle = \langle \psi, \mathsf{D}_j^\dagger(\gamma^{jk}\mathsf{D}_k\psi)\rangle, \tag{A25}$$

$$\mathsf{D}_j^\dagger\psi = \big(\nabla_a v_j^a\big)\psi - 2v_j^a\nabla_a\psi, \tag{A26}$$

$$\mathsf{H}\psi \equiv n\mathsf{F}\psi + \mathsf{D}_j^\dagger(\gamma^{jk}\mathsf{D}_k\psi). \tag{A27}$$

If a parametrization with $g_{ab} = \delta_{ab}$ can be assumed and v is θ independent, a further simplification is

$$\mathsf{H}\psi = n\mathsf{F}\psi - 4v_j^a\partial_a\big(\gamma^{jk}v_k^b\partial_b\psi\big). \tag{A28}$$

The last term becomes the Laplacian $\partial_a\partial^a\psi$ if $q = p$ and v and γ are assumed to be identity matrices.

To apply the preceding results to quantum problems, Eq. (5.2) can be generalized to

$$\mathsf{F} = \gamma^{jk}v_j^a F_{ab}v_k^b \leqslant \gamma^{jk}v_j^a K_{ab}v_k^b, \tag{A29}$$

as both $K_{ab} - F_{ab}$ and $\gamma^{jk}v_j^a v_k^b$ are positive-semidefinite. Equations (5.6) and (5.7) can then be generalized by redefining R as

$$(Rv)_a^j \equiv \gamma^{jk}K_{ab}v_k^b - \nabla_a\left[\frac{\gamma^{jk}}{\rho}\nabla_b\big(\rho v_k^b\big)\right]. \tag{A30}$$

[1] S. Amari and H. Nagaoka, *Methods of Information Geometry* (American Mathematical Society/Oxford University Press, Providence, RI, 2000); S. Amari, *Information Geometry and Its Applications* (Springer Japan, Tokyo, 2016).

[2] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models* (Springer, New York, 1993).

[3] M. Tsang, F. Albarelli, and A. Datta, Quantum Semiparametric Estimation, Phys. Rev. X **10**, 031023 (2020).

[4] H. L. Van Trees and K. L. Bell (eds.), *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (Wiley-IEEE, Piscataway, NJ, 2007).

[5] R. D. Gill and B. Y. Levit, Applications of the Van Trees inequality: A Bayesian Cramér-Rao bound, Bernoulli **1**, 59 (1995).

[6] M. P. Schützenberger, A generalization of the Fréchet-Cramér inequality to the case of Bayes estimation, Bull. Am. Math. Soc. **63**, 142 (1957).

[7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (John Wiley & Sons, New York, 2001).

[8] L. Bacharach, C. Fritsche, U. Orguner, and E. Chaumette, Some results on tighter Bayesian lower bounds on the mean-square error, arXiv:1907.09509 [cs, eess, math].

[9] C. Stein, Efficient nonparametric testing and estimation, in *Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability*, Vol. 1: Contributions to the Theory of Statistics (University of California Press, Berkeley, 1956), pp. 187–195.

[10] S. M. Carroll, *Spacetime and Geometry: An Introduction to General Relativity* (Cambridge University Press, Cambridge, UK, 2019).

[11] A. B. Tsybakov, *Introduction to Nonparametric Estimation* (Springer, New York, 2009).

[12] P. E. Jupp, A van Trees inequality for estimators on manifolds, J. Multivar. Anal. **101**, 1814 (2010).

[13] M. A. Kumar and K. V. Mishra, Information geometric approach to Bayesian lower error bounds, in *2018 IEEE In-*

ternational Symposium on Information Theory (ISIT) (IEEE, Piscataway, NJ, 2018), pp. 746–750.

[14] R. Abu-Shanab and A. Veretennikov, On asymptotic Borovkov â Sakhanenko inequality with unbounded parameter set, Theory Prob. Math. Stat. **90**, 1 (2015); K. Koike, Asymptotic comparison of some Bayesian information bounds, Commu. Stat. Theory Methods (2020), doi:10.1080/03610926.2020.1752722.

[15] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, Some classes of global Cramér-Rao bounds, Ann. Stat. **15**, 1421 (1987).

[16] B. R. Frieden, *Physics from Fisher Information: A Unification* (Cambridge University Press, Cambridge, UK, 1998).

[17] M. Tsang, Conservative classical and quantum resolution limits for incoherent imaging, J. Mod. Opt. **65**, 1385 (2018).

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

[19] M. Hayashi, *Quantum Information Theory: Mathematical Foundation* (Springer, Berlin, 2017).

[20] V. B. Braginsky and F. Ya. Khalili, *Quantum Measurement* (Cambridge University Press, Cambridge, UK, 1992).

[21] M. Tsang, H. M. Wiseman, and C. M. Caves, Fundamental Quantum Limit to Waveform Estimation, Phys. Rev. Lett. **106**, 090401 (2011).

[22] M. Tsang, R. Nair, and X.-M. Lu, Quantum Theory of Superresolution for Two Incoherent Optical Point Sources, Phys. Rev. X **6**, 031033 (2016); M. Tsang, Resolving starlight: A quantum perspective, Contemp. Phys. **60**, 279 (2019).

[23] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW150914: The Advanced LIGO Detectors in the Era of First Discoveries, Phys. Rev. Lett. **116**, 131103 (2016); H. Miao, R. X. Adhikari, Y. Ma, B. Pang, and Y. Chen, Towards the Fundamental Quantum Limit of Linear Measurements of Classical Signals, *ibid.* **119**, 050801 (2017); M. Tse *et al.*, Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy, *ibid.* **123**, 231107 (2019); F. Acernese *et al.* (Virgo Collaboration), Increasing the Astrophysical Reach of the Advanced Virgo Detector Via the Application of Squeezed Vacuum States of Light, *ibid.* **123**, 231108 (2019); H. Yu *et al.* (LIGO Scientific Collaboration), Quantum correlations between light and the kilogram-mass mirrors of LIGO, Nature **583**, 43 (2020).

[24] M. Paúr, B. Stoklasa, J. Grover, A. Krzic, L. L. Sánchez-Soto, Z. Hradil, and J. Řeháček, Tempering Rayleigh's curse with PSF shaping, Optica **5**, 1177 (2018); M. Paúr, B. Stoklasa, D. Koutný, J. Řeháček, Z. Hradil, J. Grover, A. Krzic, and L. L. Sánchez-Soto, Reading out Fisher information from the zeros of the point spread function, Opt. Lett. **44**, 3114 (2019).

[25] E. Bisketzi, D. Branford, and A. Datta, Quantum limits of localisation microscopy, New J. Phys. **21**, 123032 (2019).

[26] A. A. Borovkov and A. I. Sakhanenko, Estimates for averaged quadratic risk, Probab. Math. Statist. **1**, 185 (1980).

[27] A. A. Borovkov, *Mathematical Statistics* (Gordon and Breach, Amsterdam, 1998).

[28] A. W. van der Vaart, *Asymptotic Statistics* (Cambridge University Press, Cambridge, UK, 1998).

[29] P. Stoica and T. L. Marzetta, Parameter estimation problems with singular information matrices, IEEE Trans. Signal Process. **49**, 87 (2001).

[30] J. M. Lee, *Introduction to Smooth Manifolds* (Springer-Verlag, New York, 2003).

[31] J. A. Gross and C. M. Caves, One from many: Estimating a function of many parameters, J. Phys. A: Math. Theor. (2020), doi: 10.1088/1751-8121/abb9ed.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1985).

[33] D. E. Knuth, Big Omicron and big Omega and big Theta, ACM SIGACT News **8**, 18 (1976).

[34] T. Y. Young, Asymptotically efficient approaches to quantum-mechanical parameter estimation, Info. Sci. **9**, 25 (1975); H. Nagaoka, On Fisher information of quantum statistical models, in *Asymptotic Theory of Quantum Statistical Inference*, edited by M. Hayashi (World Scientific, Singapore, 2005), pp. 113–124.

[35] R. D. Gill and M. Guţă, On asymptotic quantum statistical inference, in *From Probability to Statistics and Back: High-Dimensional Models and Processes—A Festschrift in Honor of Jon A. Wellner*. Collections, Vol. 9, edited by M. Banerjee, F. Bunea, J. Huang, V. Koltchinskii, and M. H. Maathuis (Institute of Mathematical Statistics, Beachwood, OH, 2013), pp. 105–127.

[36] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă, Multi-parameter estimation beyond quantum Fisher information, J. Phys. A: Math. Theor. **53**, 363001 (2020).

[37] J. Suzuki, Y. Yang, and M. Hayashi, Quantum state estimation with nuisance parameters, J. Phys. A: Math. Theor. **53**, 453001 (2020).

[38] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, Pisa, Italy, 2011).

[39] F. Albarelli, M. Tsang, and A. Datta, Upper bounds on the Holevo Cramér-Rao bound for multiparameter quantum parametric and semiparametric estimation, arXiv:1911.11036 [quant-ph].

[40] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2011).

[41] R. M. Gray, Toeplitz and circulant matrices: A review, Commun. Info. Theory **2**, 155 (2006).

[42] M. Tsang, Time-Symmetric Quantum Theory of Smoothing, Phys. Rev. Lett. **102**, 250403 (2009); Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing, Phys. Rev. A **80**, 033840 (2009); Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. II. Applications to atomic magnetometry and Hardy's paradox, **81**, 013824 (2010).

[43] M. Tsang, Quantum metrology with open dynamical systems, New J. Phys. **15**, 073005 (2013).

[44] K. Iwasawa, K. Makino, H. Yonezawa, M. Tsang, A. Davidovic, E. Huntington, and A. Furusawa, Quantum-Limited Mirror-Motion Estimation, Phys. Rev. Lett. **111**, 163602 (2013).

[45] V. B. Braginsky, M. L. Gorodetsky, F. Ya. Khalili, and K. S. Thorne, Energetic quantum limit in large-scale interferometers, AIP Conf. Proc. **523**, 180 (2000).

[46] M. Tsang and R. Nair, Fundamental quantum limits to waveform detection, Phys. Rev. A **86**, 042115 (2012).

[47] S. Ng, S. Z. Ang, T. A. Wheatley, H. Yonezawa, A. Furusawa, E. H. Huntington, and M. Tsang, Spectrum analysis with quantum dynamical systems, Phys. Rev. A **93**, 042121 (2016).

[48] D. W. Berry, M. J. W. Hall, and H. M. Wiseman, Stochastic Heisenberg Limit: Optimal Estimation of a Fluctuating Phase, Phys. Rev. Lett. **111**, 113601 (2013).

[49] J. W. Goodman, *Introduction to Fourier Optics* (McGraw–Hill, New York, 2004).

[50] M. Tsang, Semiparametric estimation for incoherent optical imaging, Phys. Rev. Research **1**, 033006 (2019).

[51] W.-K. Tham, H. Ferretti, and A. M. Steinberg, Beating Rayleigh's Curse by Imaging Using Phase Information, Phys. Rev. Lett. **118**, 070801 (2017).