PAPER • OPEN ACCESS

Bayesian error regions in quantum estimation II: region accuracy and adaptive methods

View the article online for updates and enhancements.

Recent citations

- Bayesian error regions in quantum estimation I: analytical reasonings
  Yong Siah Teo *et al*

New Journal of Physics

The open access journal at the forefront of physics

CrossMark

PAPER

Bayesian error regions in quantum estimation II: region accuracy and adaptive methods

Changhun Oh[1,3], Yong Siah Teo[1,2,3] and Hyunseok Jeong[1]

[1] Center for Macroscopic Quantum Control, Department of Physics and Astronomy, Seoul National University, 08826 Seoul, Republic of Korea
[2] BK21 Frontier Physics Research Division, Department of Physics and Astronomy, Seoul National University, 08826 Seoul, Republic of Korea
[3] Authors to whom any correspondence should be addressed.

E-mail: v55ohv@snu.ac.kr and ys_teo@snu.ac.kr

Keywords: Bayesian error regions, quantum estimation, maximum likelihood, adaptive

Abstract

Bayesian error analysis paves the way to the construction of credible and plausible error regions for a point estimator obtained from a given dataset. We introduce the concept of region accuracy for error regions (a generalization of the point-estimator mean squared error) to quantify the average statistical accuracy of all region points with respect to the unknown true parameter. We show that the increase in region accuracy is closely related to the Bayesian region dual operations in Shang *et al* (2013 *New J. Phys.* **15** 123026). Next with only the given dataset as viable evidence, we establish various adaptive methods to maximize the region accuracy relative to the true parameter subject to the type of reported Bayesian region for a given point estimator. We highlight the performance of these adaptive methods by comparing them with non-adaptive procedures in three quantum-parameter estimation examples. The results of and mechanisms behind the adaptive schemes can be understood as the region analog of adaptive approaches to achieving the quantum Cramér–Rao bound for point estimators.

1. Introduction

Error estimation for a given particular point estimator of an unknown parameter constitutes an important component in quantum estimation. The assigned error interval (or region) for the estimator conveys error information about the measured data that propagates to other physical quantities predicted with this estimator.

The preceding companion article [2] discussed asymptotic techniques for constructing Bayesian regions for the maximum-likelihood (ML) estimator. Such a Bayesian region annotates the estimator with credibility that it lies in this region of a given size. As a result, we see that this construction contains statistical elements from two principal schools of thought. If one is pedantic about labeling these elements, one might say that the concept of an unknown, but fixed, parameter is that of a frequentist, whereas terminologies like size and credibility for a distribution of parameters belong to viewpoints of a Bayesian [3–5].

In this article, rather than distinguishing between these two schools, we shall understand the underlying meaning of statistical accuracy in the construction of Bayesian regions that is based on elements from these two camps. As a means to eradicate unnecessary confusion, we note here that in relation to Bayesian statistics, an important point estimator of interest is the average of the product of both the parameter and its posterior distribution over the entire parameter space—the Bayesian mean estimator. For this point estimator, concepts of statistical accuracy exist [6, 7].

In our context however, the relation between statistical accuracy for the unknown true parameter and aspects of Bayesian regions comes in a different flavor. We are still interested in a frequentist accuracy for the true parameter of interest, just as much as an observer is interested in preparing a quantum source in a particular state, for instance. On the other hand, since we are dealing with error regions, which are sets of points, we now speak of the *region accuracy*, that is the average accuracy of *all* points in the region relative to the true parameter. In the limit of zero region size, the region accuracy becomes the usual point-estimator accuracy. After a review in section 2 on Bayesian regions, we shall

see in section 3 that this notion of region accuracy is intimately connected to the dual operations [1, 2] of fixing the region size while increasing credibility, or fixing the credibility—both actions *tend to* increase the region accuracy, and this tendency becomes *exact* in single-parameter quantum estimation.

In section 4 we will proceed to develop operational schemes to maximize region accuracy by either adaptively optimizing over all credible regions of fixed size/credibility, or over all plausible regions. The adaptive schemes require solely the collected data and parameter dimension, and are in fact region analogs of adaptively attaining the quantum Cramér–Rao bound for point estimators [8–10]. These schemes will be applied to three examples in quantum estimation that can be categorized under quantum metrology and Gaussian state characterization. All symbols and notations from [2] are carried over to this article. The prior distribution for the true parameter shall again be taken to be the uniform primitive distribution in the parameter space.

2. Brief review on Bayesian regions

For the purpose of laying out the foundations for subsequent discussion on region accuracy and adaptive quantum estimation, we state the key properties of a Bayesian credible-region $\mathcal{R} = \mathcal{R}_\lambda$ that is characterized by $0 \leqslant \lambda \leqslant 1$ with an isolikelihood boundary. The size and credibility of $\mathcal{R}_\lambda$ are defined in (1) of [2].

From hereon, we shall focus (see later section 3.5) on the situation where the true parameter $\boldsymbol{r} \notin \partial\mathcal{R}$, so that for a sufficiently large data sample size N, the error region $\mathcal{R}$ for all interesting values of λ has boundary $\partial\mathcal{R} \cap \partial\mathcal{R}_0 = \varnothing$ (Case 1 in [2]). For the case in point, we reiterate the relevant expressions

$$s_\lambda = \frac{V_d}{V_{\mathcal{R}_0}}(-2\log\lambda)^{d/2} \, \text{Det}\{\boldsymbol{F}_{\text{ML}}\}^{-1/2},$$

$$c_\lambda = 1 - \frac{\Gamma(d/2, -\log\lambda)}{(d/2 - 1)!},$$

$$\lambda_{\text{crit}} = \sqrt{\text{Det}\{2\pi\,\boldsymbol{F}_{\text{ML}}^{-1}\}} \Big/ V_{\mathcal{R}_0}, \tag{1}$$

for the size, credibility and the critical $\lambda = \lambda_{\text{crit}}$ that defines the plausible-region—the credible-region that contains *all plausible parameters and nothing else*. Based on these expressions, we can obtain the simple relation

$$s_\lambda = \frac{V_d}{V_{\mathcal{R}_0}}[2\,\Gamma_{d/2}^{-1}(1 - c_\lambda)]^{d/2} \, \text{Det}\{\boldsymbol{F}_{\text{ML}}\}^{-1/2} \tag{2}$$

between s_λ and c_λ. These analytical results apply to the uniform primitive prior with respect to the parameter $\boldsymbol{r}$. More explicitly, for $\boldsymbol{r} \cong (r_1 \ r_2 \ \dots \ r_d)^{\text{T}}$, the integral measure $(\text{d}\,\boldsymbol{r}) = (\prod_j \text{d}\,r_j)/V_{\mathcal{R}_0}$.

3. Region accuracy and its connections with the dual region operations

3.1. General formalism of the region accuracy
Suppose that after collecting the experimental data $\mathbb{D}$, the ML estimator $\hat{\boldsymbol{r}}_{\text{ML}}$ is computed over the parameter space $\mathcal{R}_0$. Then the usual mean squared error (MSE) for this point estimator relative to the true parameter $\boldsymbol{r}$,

$$\text{MSE}(\boldsymbol{r}) \equiv E\,[||\hat{\boldsymbol{r}}_{\text{ML}} - \boldsymbol{r}||^2], \tag{3}$$

measures the average statistical accuracy of $\hat{\boldsymbol{r}}_{\text{ML}}$ over all possible data $\mathbb{D}$. It is known that if Case 1 applies, then for sufficiently large N the ML estimator will ultimately be unbiased ($E\,[\hat{\boldsymbol{r}}_{\text{ML}}] = \boldsymbol{r}$) and so $\text{MSE}(\boldsymbol{r}) \to \text{Tr}\{\boldsymbol{F}^{-1}\}$ approaches the Cramér–Rao bound that is defined by the Fisher information $\boldsymbol{F}$ for $\boldsymbol{r}$.

We may generalize this description of accuracy using the language of Bayesian analysis on the ML estimator $\hat{\boldsymbol{r}}_{\text{ML}}$. Since the object in this analysis is the Bayesian region $\mathcal{R}$, it is natural to introduce the *region squared error* (RSE)

$$\text{RSE}(\boldsymbol{r}) = \frac{\int_{\mathcal{R}}(\text{d}\,\boldsymbol{r}')||\boldsymbol{r}' - \boldsymbol{r}||^2}{\int_{\mathcal{R}}(\text{d}\,\boldsymbol{r}')} \tag{4}$$

that measures the region accuracy relative to $\boldsymbol{r}$, or the average accuracy of all the points in $\mathcal{R}$, where $(\text{d}\,\boldsymbol{r}')$ is the normalized integral measure as defined in [2]. It is easy to see that when $\mathcal{R} = \mathcal{R}_{\lambda=1} = \{\hat{\boldsymbol{r}}_{\text{ML}}\}$, we return to $\text{RSE}(\boldsymbol{r}) = \text{MSE}(\boldsymbol{r})$ since for any function $f(\boldsymbol{r})$,

$$\frac{\int_{\mathcal{R}_{\lambda \to 1}} (\mathrm{d}\,\boldsymbol{r}') f(\boldsymbol{r}')}{\int_{\mathcal{R}_{\lambda \to 1}} (\mathrm{d}\,\boldsymbol{r}')} = f(\hat{\boldsymbol{r}}_{\mathrm{ML}}) + \frac{\int_{\mathcal{R}_{\lambda \to 1}} (\mathrm{d}\,\boldsymbol{r}')(\boldsymbol{r}' - \hat{\boldsymbol{r}}_{\mathrm{ML}})}{\int_{\mathcal{R}_{\lambda \to 1}} (\mathrm{d}\,\boldsymbol{r}')} \cdot \boldsymbol{\partial}_{\mathrm{ML}} f(\hat{\boldsymbol{r}}_{\mathrm{ML}}),$$

$$\left\| \frac{\int_{\mathcal{R}_{\lambda \to 1}} (\mathrm{d}\,\boldsymbol{r}')(\boldsymbol{r}' - \hat{\boldsymbol{r}}_{\mathrm{ML}})}{\int_{\mathcal{R}_{\lambda \to 1}} (\mathrm{d}\,\boldsymbol{r}')} \right\| \leqslant \max_{\boldsymbol{r}' \in \mathcal{R}_{\lambda \to 1}} \|\boldsymbol{r}' - \hat{\boldsymbol{r}}_{\mathrm{ML}}\| = 0. \tag{5}$$

To analyze the average region accuracy over all possible data for the error regions, we may adopt the *mean region squared error* (MRSE)

$$\mathrm{MRSE}(\boldsymbol{r}) \equiv E\left[\mathrm{RSE}(\boldsymbol{r})\right] = E\left[\frac{\int_{\mathcal{R}} (\mathrm{d}\,\boldsymbol{r}') \|\boldsymbol{r}' - \boldsymbol{r}\|^2}{\int_{\mathcal{R}} (\mathrm{d}\,\boldsymbol{r}')} \right]. \tag{6}$$

Statistically, the MRSE is a collective error feature of the ML point estimator $\hat{\boldsymbol{r}}_{\mathrm{ML}}$ *and* its surrounding states in $\mathcal{R}$ relative to $\boldsymbol{r}$.

To understand how the MRSE behaves with the Bayesian region properties in the asymptotic limit of N, it is necessary to calculate the MRSE in this limit. After some straightforward calculations in appendix A, it turns out that for sufficiently large N where Case 1 holds, the RSE takes the simple form

$$\mathrm{RSE}(\boldsymbol{r}) = \|\hat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r}\|^2 + 2\,\mathrm{Tr}\{\boldsymbol{F}_{\mathrm{ML}}^{-1}\} \frac{(-\log \lambda)}{d + 2}. \tag{7}$$

We observe that the RSE linearly fuses the regular 'frequentist' point-estimator accuracy measure, the squared error of $\hat{\boldsymbol{r}}_{\mathrm{ML}}$ for a fixed unknown $\boldsymbol{r}$, with 'Bayesian' elements that characterize the region $\mathcal{R}$. Evidently, we get $\mathrm{RSE}(\boldsymbol{r}) = \mathrm{MSE}(\boldsymbol{r})$
for $\lambda = 1$. With this, we may invoke the property $E\left[\hat{\boldsymbol{r}}_{\mathrm{ML}}\right] \to \boldsymbol{r}$ for sufficiently large N and arrive at the formula

$$\mathrm{MRSE}(\boldsymbol{r}) = \mathrm{Tr}\{\boldsymbol{F}^{-1}\}\left(1 - \frac{2\log \lambda}{d + 2}\right) \tag{8}$$

for the MRSE, where we have implicitly assumed that $\boldsymbol{F}_{\mathrm{ML}} \approx \boldsymbol{F} = N\boldsymbol{F}_1$ in the asymptotic limit and $\boldsymbol{F}_1$ is the Fisher information evaluated with $\boldsymbol{r}$ for a single copy $N = 1$ of datum[4]. For convenience, we shall drop the parametric variable λ hereafter.

3.2. Duality actions on credible-region accuracy
Equation (8) provides a basis for us to discuss the effects on the accuracy of credible regions depending on how an observer chooses to optimize the region qualities. We first emphasize that the action of fixing the region size while increasing credibility and that of fixing the region credibility while reducing the size are dual actions in the sense that after these actions, the credible-region is optimally defined [1]. Armed with the concept of region accuracy, we can endow the effects from these dual strategies with richer statistical meaning. To this end, we analyze the uniform-prior MRSE for different parameter dimension d values.

3.2.1. d = 1
In single-parameter estimation, the Fisher information is a numerical quantity F that is related to both the size s and credibility c by

$$F = \frac{8}{s^2 V_{\mathcal{R}_0}^2} \Gamma_{1/2}^{-1}(1 - c), \tag{9}$$

the $d = 1$ version of (2). The resulting MRSE for credible intervals is then given by

$$\mathrm{MRSE}^{(\mathrm{cred})}(r; s, c) = \frac{s^2 V_{\mathcal{R}_0}^2}{8}\left[\frac{1}{\Gamma_{1/2}^{-1}(1 - c)} + \frac{2}{3} \right] \tag{10}$$

as a function of s and c.

For $d = 1$, the influence of the region dual operations on the MRSE is clear. When the credible-region size $s = s_0$ is fixed, increasing the credibility c reduces the MRSE for a given r, as $\Gamma_{d/2}^{-1}(1 - c)$ is a (strictly) monotonically increasing function of c for any d. If $c = c_0$ is fixed instead, then reducing s would, of course, reduce the MRSE. Therefore both dual strategies *increases* the region accuracy.

---

[4] We recall that all data copies are i.i.d., such that $\boldsymbol{F}_{\mathrm{ML}} \approx \boldsymbol{F}$ is an N multiple of $\boldsymbol{F}_{\mathrm{ML},1} \approx \boldsymbol{F}_1$.

At first sight, $\text{MRSE}^{(\text{cred})}(r; s, c)$ in (10) is apparently independent of r. This thought is misleading because as a matter of fact, s and c are related to each other through F as stated in (9). If F is allowed to vary by changing the measurement setup or procedure, then s and c would behave as independent variables. Upon reviewing the dual strategies once more, increasing c for a fixed $s = s_0$ or decreasing s for a fixed $c = c_0$ both require an increase in the Fisher information F, which is really the underlying physical quantity that controls the mechanisms behind the dual actions. Hence, increasing c for a fixed $s = s_0$ is dual to decreasing s for a fixed $c = c_0$ in the sense that they both reduces the MRSE for credible intervals.

3.2.2. $d \geqslant 2$

Here, matters are slightly less straightforward, for the MRSE depends on a more complicated function of the Fisher information F that is no longer a numerical value. Both s and c are related to each other through $\text{Det}\{F\}$, while the MRSE is a function of $\text{Tr}\{F_{\text{ML}}^{-1}\}$ and $-\log \lambda$.

We may first consider the case where $s = s_0$. This sets up the constraint

$$\left(\frac{s_0 V_{\mathcal{R}_0}}{V_d}\right)^2 \text{Det}\{F\} = [2\, \Gamma_{d/2}^{-1}(1 - c)]^d \tag{11}$$

for c and the functional dependence of MRSE on F is now elucidated. To increase c (reduce λ) under a fixed $s = s_0$, it is clear that $\text{Det}\{F\}$ should increase so that λ decreases in order to maintain a fixed size. However, since $\text{Tr}\{F^{-1}\}$ is *not* a function of $\text{Det}\{F\}$, there is generally no guarantee that the MRSE will decrease with increasing $\text{Det}\{F\}$. There is however a trend that this is the case, and this statement can be made more precisely by considering the largest $\text{Tr}\{F^{-1}\}$ for a given $\text{Det}\{F\}$. If we make use of the fact that for any given physical system, the Fisher information must be trace class [11] ($\text{Tr}\{F\} \leqslant B$ for some positive constant B), then one can derive the simple inequality

$$\text{Tr}\{F^{-1}\} \leqslant \frac{d\, B^{d-1}}{\text{Det}\{F\}} \tag{12}$$

for a given $\text{Det}\{F\}$ (refer to appendix B for a short derivation). The stated upper bound is loose for $d \geqslant 2$, but is sufficient to make our case. After invoking the constraint, we then have

$$\text{MRSE}^{(\text{cred})}(r; s_0, c) \leqslant \left(\frac{s_0 V_{\mathcal{R}_0}}{V_d}\right)^2 \frac{d\, B^{d-1}}{[2\, \Gamma_{d/2}^{-1}(1 - c)]^d} \left[1 + \frac{2\, \Gamma_{d/2}^{-1}(1 - c)}{d + 2}\right]. \tag{13}$$

We see that for $d = 1$, the upper bound above reduces to the exact expression in (10). Otherwise, this upper bound decreases monotonically with increasing c for $d \geqslant 2$. The same arguments apply when $c = c_0$, only that now $\text{MRSE}^{(\text{cred})}(r; s, c_0) \leqslant \text{const.} \times s^2$ and so decreasing s by increasing $\text{Det}\{F\}$ (see (2)) reduces the upper bound quadratically.

3.3. Duality actions on the plausible-region accuracy

If we take $\lambda = \lambda_{\text{crit}}$, then this time, the dual strategies are carried out with the additional constraint imposed on the value of λ. Hence, s and c are no longer independent variables. Nonetheless, we may still choose to reduce s or increase c subject to this plausible-region condition.

3.3.1. $d = 1$

For single-parameter estimation, if we choose to increase c, then since

$$-\log(2\pi F^{-1}/V_{\mathcal{R}_0}^2) = 2\, \Gamma_{1/2}^{-1}(1 - c) \tag{14}$$

we encounter the simple formula

$$\text{MRSE}^{(\text{plaus})}(r; c) = \frac{V_{\mathcal{R}_0}^2}{2\pi}\, e^{-2\, \Gamma_{1/2}^{-1}(1-c)} \left[1 + \frac{2\, \Gamma_{1/2}^{-1}(1 - c)}{3}\right]. \tag{15}$$

It follows that increasing $c \geqslant 1 - \Gamma(1/2, 1/2)/\sqrt{\pi}$ through increasing F appropriately, reduces $\text{MRSE}^{(\text{plaus})}(r, c)$ monotonically.

On the other hand, if we choose to reduce s, then based on the one-dimensional identity

$$s = \frac{2}{V_{\mathcal{R}_0}\sqrt{F}} \left[\log\left(\frac{V_{\mathcal{R}_0}^2 F}{2\pi}\right)\right]^{1/2} \tag{16}$$

as well as the parametric form

$$\text{MRSE}^{(\text{plaus})}(r; F, s(F)) = \frac{1}{F} + \frac{1}{12}(s V_{\mathcal{R}_0})^2, \tag{17}$$

it turns out that the way to do this is, again, to increase $F \geqslant 2\pi e / V_{\mathcal{R}_0}^2$, so that $\text{MRSE}^{(\text{plaus})}(r; F, s(F))$ decreases monotonically.

3.3.2. $d \geqslant 2$

Likewise, we may carry out the same analysis for $d \geqslant 2$ by first remembering that raising $\mathrm{Det}\{\boldsymbol{F}\}$ does not guarantee a reduction in $\mathrm{Tr}\{\boldsymbol{F}^{-1}\}$. Therefore using the inequality in (12), we can instead look at the upper bound for the MRSE and find that

$$\mathrm{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r};\, c) \leqslant \frac{d\, B^{d-1} V_{\mathcal{R}_0}^2}{(2\pi)^d}\, \mathrm{e}^{-2\,\Gamma_{d/2}^{-1}(1-c)} \left[1 + \frac{2\,\Gamma_{d/2}^{-1}(1-c)}{d+2} \right]. \tag{18}$$

It can then be shown that if one increases $c \geqslant 1 - \Gamma(d/2,\, 1/2)/(d/2-1)!$ by raising $\mathrm{Det}\{\boldsymbol{F}\}$, $\mathrm{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r};\, c)$ decreases monotonically.

The same goes for the strategy of reducing s by increasing $\mathrm{Det}\{\boldsymbol{F}\} \geqslant (2\pi\mathrm{e})^d/V_{\mathcal{R}_0}^2$, the upper bound of the parametric expression

$$\mathrm{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r};\, \boldsymbol{F},\, s(\boldsymbol{F})) \leqslant \frac{d\, B^{d-1}}{\mathrm{Det}\{\boldsymbol{F}\}} \left[1 + \left(\frac{s V_{\mathcal{R}_0}}{V_d} \right)^{2/d} \frac{\mathrm{Det}\{\boldsymbol{F}\}^{1/d}}{d+2} \right] \tag{19}$$

decreases monotonically.

3.4. The short summary

We can now draw some succinct yet important conclusions, for any trace class Fisher information $\boldsymbol{F}$, regarding the statistical meaning of the dual strategies with Bayesian regions of uniform priors. For credible regions, the action of increasing c with a fixed s and its dual action either reduces the MRSE when $d = 1$, or its upper bound (set by a physical upper limit of $\mathrm{Tr}\{\boldsymbol{F}\}$) when $d \geqslant 2$. The remarks for plausible regions are highly similar. Under the constraint $\lambda = \lambda_{\mathrm{crit}}$, if an observer either increases c or decreases s for the range $\lambda_{\mathrm{crit}} \leqslant \mathrm{e}^{-1/2} \approx 0.6065$, then either the MRSE ($d = 1$) or its upper bound ($d \geqslant 2$) drops monotonically (see appendix C for the derivation of plausible-region threshold values for which these behaviors hold).

So the dual operations for credible regions, or their constrained versions for plausible regions, precisely enhance the region accuracy for $d = 1$, or produce the tendency to do so for $d \geqslant 2$. As a final note, the upper bound in (12) used to argue the general tendency in reducing the MRSE with the dual strategies for $d \geqslant 2$ may be tightened if so desired. The conclusions are then further strengthened with these tighter bounds.

3.5. Situations for Case 2 and 3 from [2]

A general MRSE formula that rigorously accounts for the occurrences of Case 2 and 3 is difficult to compute, and there is no known exact relations with the duality actions when these cases are incorporated. However under the condition of large N, we may state, with proof, the following conservativeness property for categorically assuming only Case 1 in calculating the MRSE:

Conservativeness property. For the primitive prior, if we assume that N *is large enough, so that the region boundary $\partial\mathcal{R} \cap \partial\mathcal{R}_0$ is almost flat (refer to [2] for the relevant arguments) and statistical fluctuation is small enough such that $r \approx \hat{r}_{\mathrm{ML}}$, then calculating the data average of (7) (approximated with (8)) always produces a larger value than the actual MRSE for any* d.

The main outline of the proof is to show that since the MRSE is the data average of the RSE, if we categorically insist that Case 1 happens when in fact Case 2 or 3 has actually happened, then the corresponding *as-if* RSE is always larger than the actual RSE under the large data sample condition. Indeed, this categorical RSE is precisely given by (7) evaluated with $\hat{r}_{\mathrm{ML}} \in \mathcal{R}$[5]. Once this is settled, the resulting MRSE estimate, which is *approximated* by (8), is in principle an overestimate.

The conservativeness of averaging (7) is clear for $d = 1$. The categorical RSE reads

$$\mathrm{RSE}_{\mathrm{CAT}} = \mathrm{RSE}_{d=1}(r;\, a,\, \hat{r}_{\mathrm{ML}}) = \frac{1}{3}[a^2 - 2\,\hat{r}_{\mathrm{ML}}(a + 3r) + 3r^2 + 4\,\hat{r}_{\mathrm{ML}}^2], \tag{20}$$

which is the RSE for the *as-if* Bayesian interval $[a,\, 2\,\hat{r}_{\mathrm{ML}} - a]$ centered at $\hat{r}_{\mathrm{ML}}$. Here a is data dependent. On the other hand, the RSE of the actual Bayesian interval $[a,\, b]$ is

$$\mathrm{RSE} = \mathrm{RSE}_{d=1}(r;\, a,\, b,\, \hat{r}_{\mathrm{ML}}) = \frac{1}{3}(a^2 + ab + b^2) - r(a + b) + r^2, \tag{21}$$

where $b \in \partial\mathcal{R}_0$ satisfies $a < \hat{r}_{\mathrm{ML}} \leqslant b \leqslant 2\,\hat{r}_{\mathrm{ML}} - a$, and the true parameter $a \leqslant r \leqslant b$. Consistently, we have $\mathrm{RSE} = \mathrm{RSE}_{\mathrm{CAT}}$ when $b = 2\,\hat{r}_{\mathrm{ML}} - a$. It is clear that if $r \approx \hat{r}_{\mathrm{ML}}$,

---

[5] Even when Case 3 does happen but Case 1 is assumed, under the large-N limit, the expression in equation (13) of [2] approaches the actual likelihood that peaks at the correct unrestricted maximum, so that the calculations in appendix A also asymptotically yield the categorical RSE.

$$\mathrm{RSE} - \mathrm{RSE}_{\mathrm{CAT}} \approx \frac{1}{3}(b - r)(a + b - 2r) \leqslant 0, \qquad (22)$$

which also means that the categorical RSE is an overestimate of the actual RSE.

One can also find an example for which this conservative property is lost if r is far from $\hat{r}_{\mathrm{ML}}$ or N is not large enough. We may inspect the difference

$$\mathrm{RSE} - \mathrm{RSE}_{\mathrm{CAT}} = \frac{1}{3}(a + b - 2\,\hat{r}_{\mathrm{ML}})(b - 3r + 2\,\hat{r}_{\mathrm{ML}}) \qquad (23)$$

and find that in order for it to be positive, we simply need $\hat{r}_{\mathrm{ML}} < (3r - b)/2$. This shows that when $d = 1$ using (8) is always a conservative choice for Case 2 and 3 only if $r \approx \hat{r}_{\mathrm{ML}}$, which asymptotically holds in the large-N limit. Armed with the insights from $d = 1$, appendix D separately proves the conservativeness property for arbitrary d.

4. Adaptive methods for optimizing region accuracy

4.1. Optimization of region accuracy

Our next goal is to device methods that minimizes the MRSE for any r as defined in (8). For $d = 1$, maximizing the determinant of the Fisher information directly reduces the MRSE according to the assessments in section 3, as the determinant is simply the numerical Fisher information itself. To optimize $F = N F_1$ or the MRSE, an observer may choose to either increase N for a fixed POM that defines F_1, or optimize F_1 over feasible POMs for a given N. As an example, figures 1 and 2 express what happens to the MRSE when N is increased when a fixed two-outcome POM is used to perform single-parameter estimation. In what follows, we shall address the more interesting problem of optimizing the MRSE, and its optimization for $d = 1$ is equivalent to the search for the optimal POM that approaches the well-known quantum Fisher information $F \leqslant F_{\mathrm{Q}}$ [8–10] subject to either a fixed s or c when reporting credible regions, or $\lambda = \lambda_{\mathrm{crit}}$ when reporting plausible regions.

When $d \geqslant 2$, the lesson learnt from section 3 shows that the maximization of $\mathrm{Det}\{F\}$ does *not* guarantee a minimization of the MRSE. In spite of this, an observer may still carry out POM optimization to minimize the MRSE subject to the kind of Bayesian region that he or she is interested in reporting along with the ML estimator. This is essentially the region analog of maximizing the quantum Fisher information by virtue of equation (8). The correct maximum depends on the true parameter r, which is always unknown to the observer. In view of this, we shall develop numerical adaptive protocols that require only the measured data and d to carry out the MRSE minimization. These protocols are applicable in practical experimental situations where the measurement settings for the POM are described by the variable $\boldsymbol{m} \triangleq (m_1 \ m_2 \ \ldots \ m_{d_m})^{\mathrm{T}} \in \mathcal{S}_{\boldsymbol{m}}$ with a finite dimension d_m.

4.2. Adaptive scheme for credible regions

Without knowing r, the experimental data $\mathbb{D}$, if IC in the d-dimensional vector space, can still give us a unique ML estimator $\hat{r}_{\mathrm{ML}}$. In the limit of large N, this asymptotically unbiased ML estimator is also presumed to be statistically consistent ($\hat{r}_{\mathrm{ML}} \to r$). With these good properties, $\hat{r}_{\mathrm{ML}}$ can be used as the *a posteriori* state in place of r, with which we can make educated guesses for the optimal settings that minimize the MRSE, where its asymptotic expressions in terms of F are given as

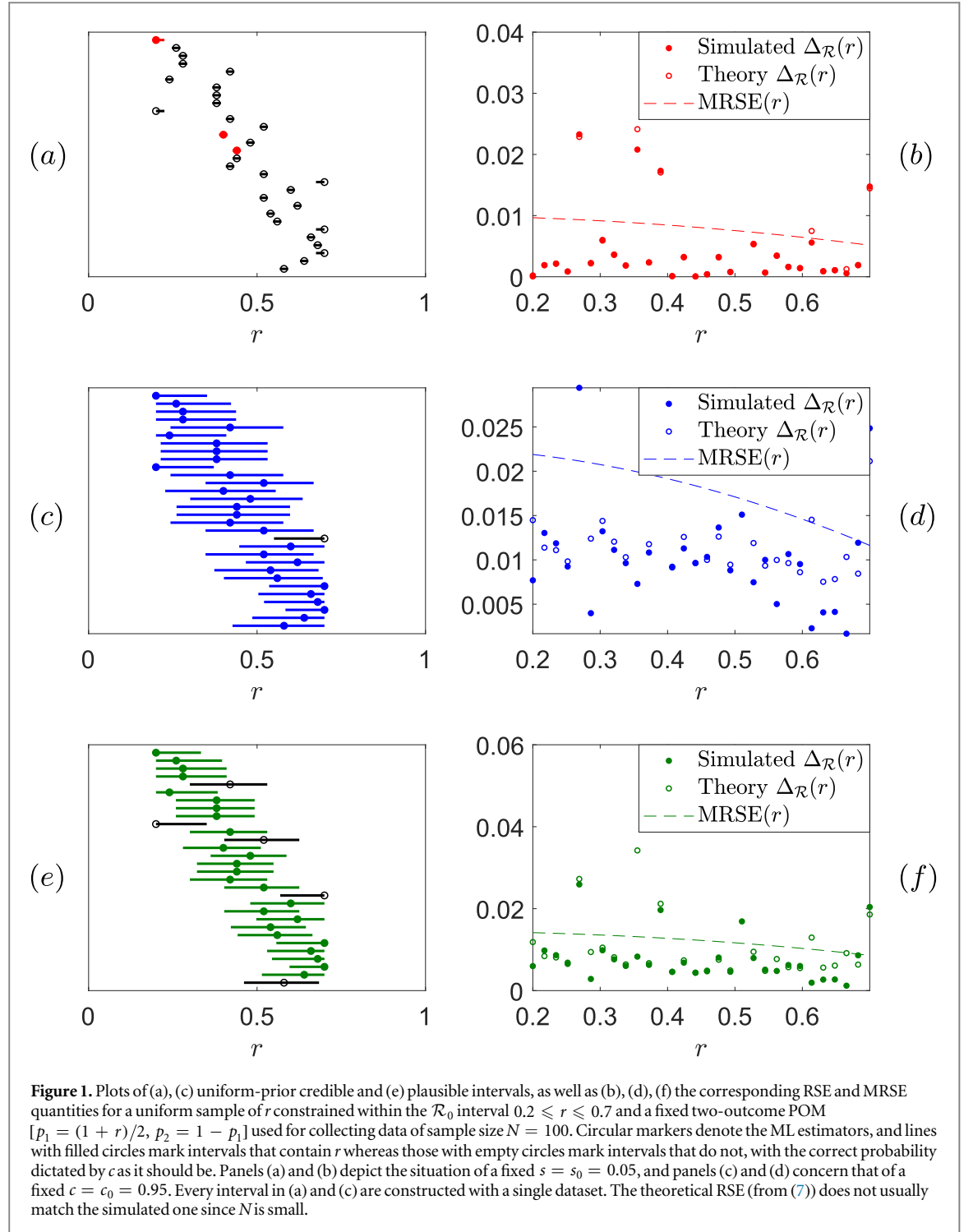$$\mathrm{MRSE}^{(\mathrm{cred})}(\boldsymbol{F}; s = s_0) = \mathrm{Tr}\{\boldsymbol{F}^{-1}\}\left[1 + \left(\frac{s_0 V_{\mathcal{R}_0}}{V_d}\right)^{2/d} \frac{\mathrm{Det}\{\boldsymbol{F}\}^{1/d}}{d + 2}\right],$$

$$\mathrm{MRSE}^{(\mathrm{cred})}(\boldsymbol{F}; c = c_0) = \mathrm{Tr}\{\boldsymbol{F}^{-1}\}\left[1 + \frac{2\Gamma_{d/2}^{-1}(1 - c_0)}{d + 2}\right]. \qquad (24)$$

As a related side note on numerically implementing the adaptive schemes, we mention that when N or d is too large for Monte Carlo numerical calculations of s and c to work, the asymptotic tools presented in [2] may be used.

Since s and c are independent, we have the following adaptive algorithm that carries out a total of K adaptive steps for a total of N measurement copies (N/K copies measured in each step) and fixed region size starting with Step $k = 1$:

MRSE minimization for credible regions

(i) Collect $\mathbb{D}_k$ with the setting $\boldsymbol{m}_k$ and compute the d-dimensional $\hat{r}_{\mathrm{ML}}$ with the accumulated dataset $\{\mathbb{D}_1, \mathbb{D}_2, \ldots, \mathbb{D}_k\}$.

(ii) Set the *a posteriori* state $\rho_0 = \hat{r}_{\mathrm{ML}}$, and generate L simulated datasets from ρ_0 for each of n_m measurement variables. Here n_m should be a reasonable number of measurement-setting variables $\{\boldsymbol{m}_j\}_{j=1}^{n_m}$ that uniformly covers $\mathcal{S}_{\boldsymbol{m}}$.

Figure 1. Plots of (a), (c) uniform-prior credible and (e) plausible intervals, as well as (b), (d), (f) the corresponding RSE and MRSE quantities for a uniform sample of r constrained within the $\mathcal{R}_0$ interval $0.2 \leqslant r \leqslant 0.7$ and a fixed two-outcome POM $[p_1 = (1 + r)/2, \, p_2 = 1 - p_1]$ used for collecting data of sample size $N = 100$. Circular markers denote the ML estimators, and lines with filled circles mark intervals that contain r whereas those with empty circles mark intervals that do not, with the correct probability dictated by c as it should be. Panels (a) and (b) depict the situation of a fixed $s = s_0 = 0.05$, and panels (c) and (d) concern that of a fixed $c = c_0 = 0.95$. Every interval in (a) and (c) are constructed with a single dataset. The theoretical RSE (from (7)) does not usually match the simulated one since N is small.

(Continued.)

(iii) With a total of Ln_m simulated and k measured datasets, obtain the projected ML estimators $\{\widetilde{\hat{r}}_{\text{ML } j,l}\}_{j=1,l=1}^{n_m,L}$ and the corresponding projected ML Fisher information $\{\widetilde{F}_{\text{ML } j,l}\}_{j=1,l=1}^{n_m,L}$.

(iv) Minimize the MRSE (evaluated with the projected Fisher information using and averaged over all L datasets) with *either* $s = s_0$ *or* $c = c_0$ and set to measure the optimal $\boldsymbol{m}_{\text{opt},k}$.

(v) Increase k by one and repeat Steps (i) through (iv) until $k = K$.

Figure 2. Similar plots as in figure 1 for $N = 10\,000$, where all other specifications remain unchanged. The RSE and MRSE are on average lower than those in figure 1 by about two orders of magnitude. This is consistent with the way the credible intervals respond to an increase in N. When $c = 0.95$, all intervals become shorter, and when $s = 0.05$, the intervals adjust their centers to increase the statistical coverage of r. All intervals eventually align to minimize the average distance from r.

4.3. Adaptive scheme for plausible regions

A similar adaptive protocol can be developed to minimize

$$\text{MRSE}^{(\text{plaus})}(\boldsymbol{F}) = \text{Tr}\{\boldsymbol{F}^{-1}\}\left[1 + \frac{\log(V_{\mathcal{R}_0}^2/(2\pi)^d)}{d+2} + \frac{\log(\text{Det}\{\boldsymbol{F}\})}{d+2}\right]. \tag{25}$$

MRSE minimization for plausible regions

 (i) Collect $\mathbb{D}_k$ with the setting $\boldsymbol{m}_k$ and compute the d-dimensional $\hat{\boldsymbol{r}}_{\text{ML}}$ with the accumulated dataset $\{\mathbb{D}_1, \mathbb{D}_2, \ldots, \mathbb{D}_k\}$.

(ii) Set the *a posteriori* state $\rho_0 = \hat{\boldsymbol{r}}_{\text{ML}}$, and generate L simulated datasets from ρ_0 for each of n_m measurement variables. Here n_m should be a reasonable number of measurement-setting variables $\{\boldsymbol{m}_j\}_{j=1}^{n_m}$ that uniformly covers $\mathcal{S}_{\boldsymbol{m}}$.

(iii) With a total of Ln_m simulated and k measured datasets, obtain the projected ML estimators $\{\widetilde{\hat{\boldsymbol{r}}}_{\text{ML}\,j,l}\}_{j=1,l=1}^{n_m,L}$ and the corresponding projected ML Fisher information $\{\widetilde{\boldsymbol{F}}_{\text{ML}\,j,l}\}_{j=1,l=1}^{n_m,L}$.

(iv) Minimize the MRSE (evaluated with the projected Fisher information and averaged over all L datasets) with $\lambda = \lambda_{\text{crit}}$ and set to measure the optimal $\boldsymbol{m}_{\text{opt},k}$.

 (v) Increase k by one and repeat Steps (i) through (iv) until $k = K$.

4.4. Differences from known Bayesian adaptive schemes

Before we proceed with the examples, it is timely to mention here that there exist adaptive schemes that choose optimal configurations for enhancing the tomographic quality of *point estimators*, for instance, in tracking drifts in quantum states and processes [12, 13]. The primary mechanism behind these adaptive schemes is to improve accuracies of point estimators measured by objective functions of the posterior Hessian that encodes geometrical properties of Bayesian region around the posterior maximum in the limit of large N.

We emphasize that these previously proposed schemes are of a different qualitative nature from that of the adaptive MRSE minimization schemes presented here. The present concern is the accuracy of an error region, as opposed to a single estimator. In this case, not only are the geometrical properties of the Bayesian region $\mathcal{R}$ around $\hat{r}_{\mathrm{ML}}$ important in our considerations, but also the quality of every state within $\mathcal{R}$ relative to the unknown true parameter $\boldsymbol{r}$. Maximizing the MRSE therefore operates on a higher hierarchical level—it is the whole error region $\mathcal{R}$, namely the point estimator $\hat{r}_{\mathrm{ML}}$ and surrounding error states, that collectively possesses the maximum (average) accuracy (minimum MRSE), not just $\hat{r}_{\mathrm{ML}}$.

That being said, the idea of region accuracy and its maximization not only forms one bridge that connects parts of frequentist and Bayesian elements, but also directly support the Bayesian spirit that surrounding states of $\hat{r}_{\mathrm{ML}}$ are just as important (according to the prior) in parameter error analysis. The adaptive methods established in sections 4.2 and 4.3 are meant for this distinct purpose.

5. Examples

5.1. Phase-shifted homodyne interferometer ($d = 1$)

An important single-parameter estimation task in quantum information is phase reconstruction for an interferometer with quantum input resources [14–16]. A very common type of interferometer is the homodyne measurement setup [17–20] that is employed in continuous-variable quantum tomography and cryptography. An interesting case arises when both the source (mode a) and local oscillator (LO) arms of the homodyne setup differ by an unknown relative phase $r = \phi$ that can be modeled by the phase-shifter described by a unitary operator $U(\phi) = \mathrm{e}^{\mathrm{i}a^{\dagger}a\phi}$. The job is to characterize the unknown phase ϕ for the interferometer setup, which is a one-dimensional problem ($d = 1$).

It is known in [21] that using a squeezed-vacuum state $|\zeta\rangle\langle\zeta|$ for mode a saturates the quantum Cramér–Rao bound in ϕ estimation, where the Born probabilities

$$p(x_{\vartheta}, \vartheta) = \frac{\mathrm{e}^{-x_{\vartheta}^{2}/2\sigma_{\vartheta,\phi}^{2}}}{\sqrt{2\pi\sigma_{\vartheta,\phi}^{2}}},$$

$$\sigma_{\vartheta,\phi}^{2} = \frac{1}{2}[\cosh(2\zeta) + \cos(2\vartheta - 2\phi)\sinh(2\zeta)], \tag{26}$$

encode the unknown phase ϕ, the (real) squeeze parameter ζ and the homodyne local oscillator (LO) phase ϑ.

The adaptive schemes in section 4 are readily applicable to this one-dimensional quantum estimation scenario. They equivalently maximize the Fisher information

$$F = N\frac{\sinh(2\zeta)^{2} - [\cosh(2\zeta) - 2\sigma_{\vartheta,\phi}^{2}]^{2}}{2\sigma_{\vartheta,\phi}^{2}} \tag{27}$$

for this problem. The optimal LO phase $m_{\mathrm{opt}} = \vartheta_{\mathrm{opt}} = \phi - \cos^{-1}(\tanh\zeta)/2$ that achieves the maximum depends on ϕ, and the adaptive schemes asymptotically select this value without this knowledge. Figures 3 and 4 demonstrate the advantage of employing adaptive schemes over non-adaptive ones in increasing region accuracies with a fixed total number of copies N and ζ by performing the relevant optimization over the space $\mathcal{S}_m$ of LO phase $m = \vartheta$. Figures 3 and 4 compare the difference in MRSE between adaptive and non-adaptive IC schemes. A more sophisticated second example in quantum metrology shall follow.

5.2. Three-path interferometer ($d = 2$)

One may generalize a typical two-arm interferometer, such as the homodyne setup discussed previously, to a three-arm interferometer (modes a, b and c) of unknown relative phases $\boldsymbol{r} \triangleq (\phi_1 \ \phi_2)^{\mathrm{T}}$ in the three arms, with ϕ_1 being the phase difference between modes a and c, and ϕ_2 between b and c. Such an interferometer poses a two-parameter estimation problem and may be modeled with the ordered sequence of first a *beam tritter* (U_3), followed by a three-arm phase shifter [$U(\phi_1, \phi_2)$], and finally another beam tritter —$U_{\mathrm{three\text{-}path}}(\phi_1, \phi_2) = U_3 U(\phi_1, \phi_2) U_3$—, where
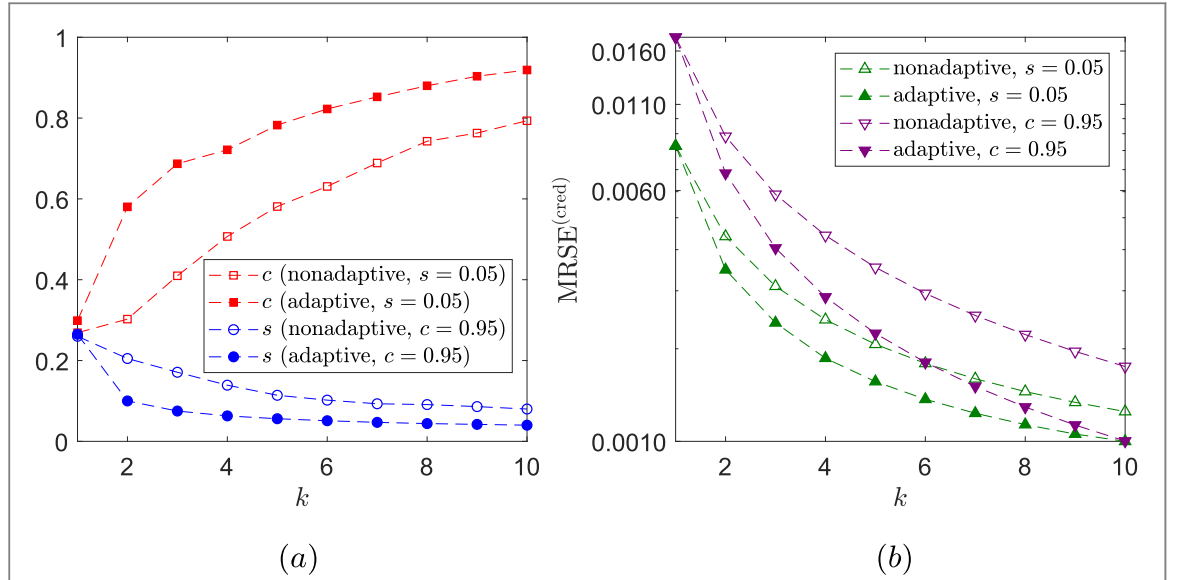
Figure 3. Plots of (a) the credible-interval properties and (b) MRSE$^{(cred)}$ for $r = 1.179$ and a fixed squeeze parameter $\zeta = 0.7$. While all numerical schemes are executed with only ML estimators, the MRSE graphs are plotted with the true parameter to show the correct accuracies, just like any analysis of the point MSE. A primitive prior that extends to the finite range $0 \leqslant \phi \leqslant \pi/2$ is assumed in the simulation as prior knowledge about the unknown relative phase r. Here $N = 1000$ copies are distributed equally to $K = 10$ adaptive steps that are carried out by each adaptive protocol. The non-adaptive versions measure the fixed LO phase $\vartheta = 1.837$ throughout the run, which is less efficient than their adaptive counterparts that begin with the same LO phase and eventually converge to the optimal LO phase subject to the constraint imposed on $\mathcal{R}$ (fixed s or c).
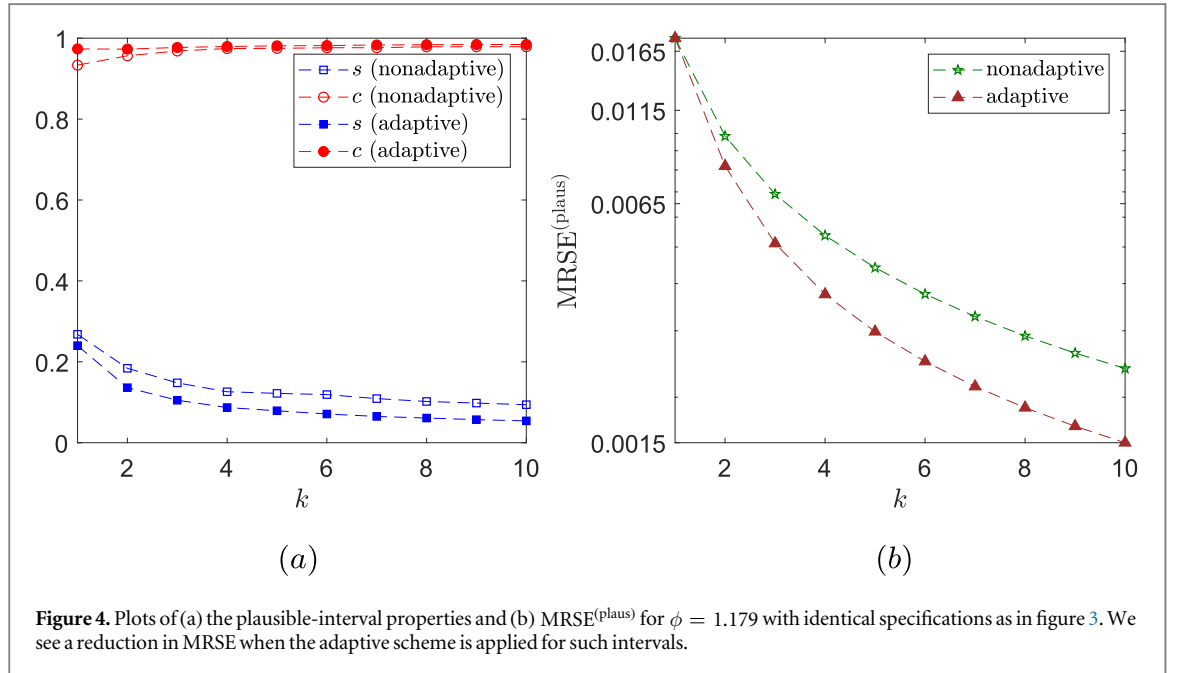

Figure 4. Plots of (a) the plausible-interval properties and (b) MRSE$^{(plaus)}$ for $\phi = 1.179$ with identical specifications as in figure 3. We see a reduction in MRSE when the adaptive scheme is applied for such intervals.

$$U(\phi_1, \phi_2) = e^{i(a^\dagger a \, \phi_1 + b^\dagger b \, \phi_2)},$$

$$U_3 \cong \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & e^{2\pi i/3} & e^{2\pi i/3} \\ e^{2\pi i/3} & 1 & e^{2\pi i/3} \\ e^{2\pi i/3} & e^{2\pi i/3} & 1 \end{pmatrix}. \tag{28}$$

To estimate a given unknown $\boldsymbol{r}$, we shall suppose that incoming photons are initially in the three-mode input state $\rho = |1, 1, 1\rangle \langle 1, 1, 1|$ to be described by a tripartite Fock state, which, after traversing the interferometer that is additionally encoded with measurement control phases $\boldsymbol{m} \cong (\psi_1 \, \psi_2)^T$ for tuning the final estimation accuracy, would then encounter idealized photon-counting detectors that result in the 10 Born probabilities $p_{n_1, n_2, n_3}(\psi_1, \psi_2; \phi_1, \phi_2) = |\langle n_1, n_2, n_3| \, U_{three-path}(\psi_1 - \phi_1, \psi_2 - \phi_2)|1, 1, 1\rangle|^2$
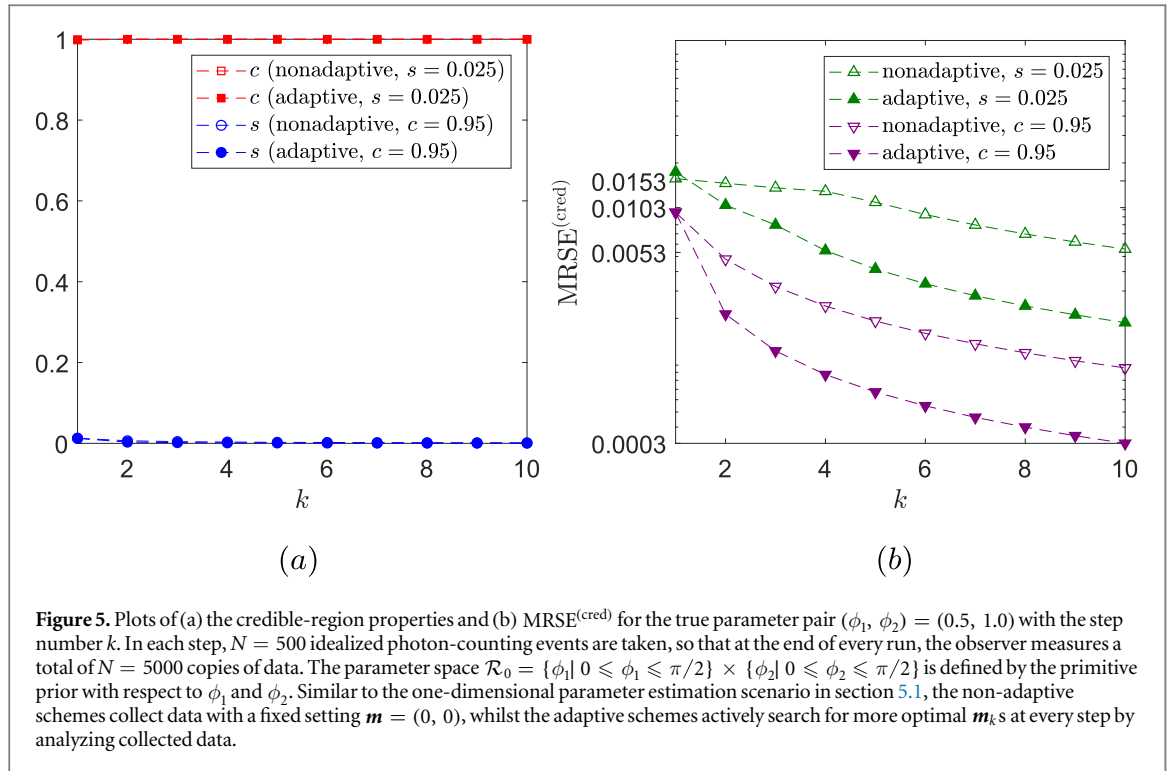
Figure 5. Plots of (a) the credible-region properties and (b) MRSE$^{(\text{cred})}$ for the true parameter pair $(\phi_1, \phi_2) = (0.5, 1.0)$ with the step number k. In each step, $N = 500$ idealized photon-counting events are taken, so that at the end of every run, the observer measures a total of $N = 5000$ copies of data. The parameter space $\mathcal{R}_0 = \{\phi_1|\, 0 \leqslant \phi_1 \leqslant \pi/2\} \times \{\phi_2|\, 0 \leqslant \phi_2 \leqslant \pi/2\}$ is defined by the primitive prior with respect to ϕ_1 and ϕ_2. Similar to the one-dimensional parameter estimation scenario in section 5.1, the non-adaptive schemes collect data with a fixed setting $\boldsymbol{m} = (0, 0)$, whilst the adaptive schemes actively search for more optimal $\boldsymbol{m}_k$s at every step by analyzing collected data.
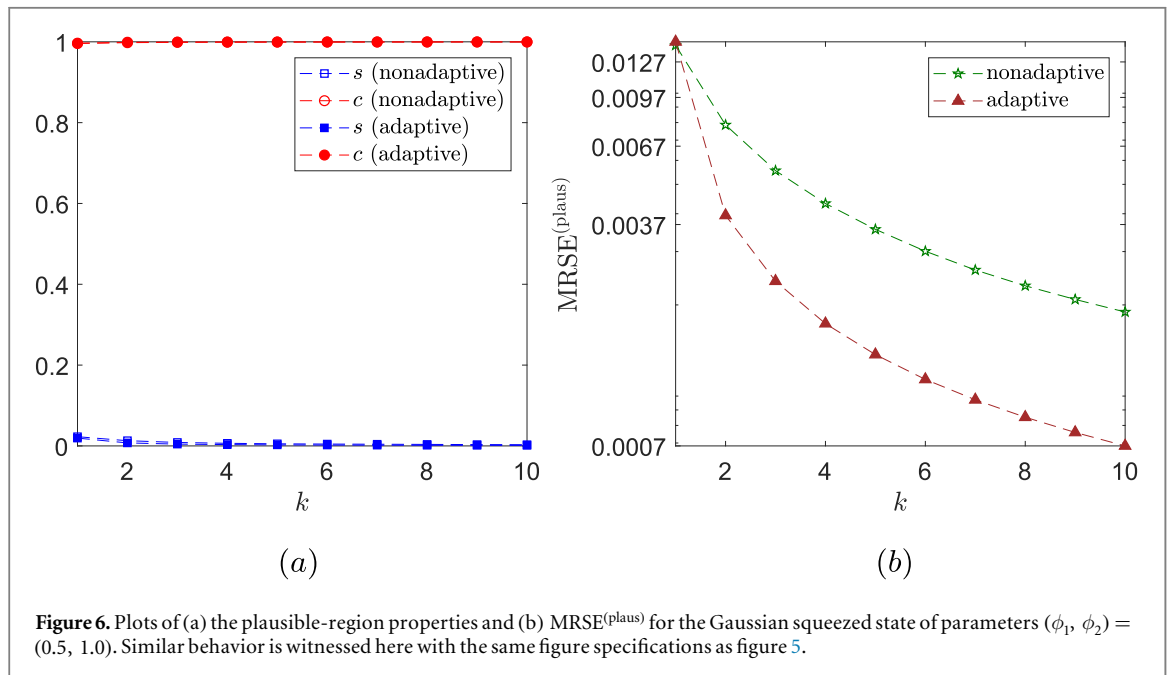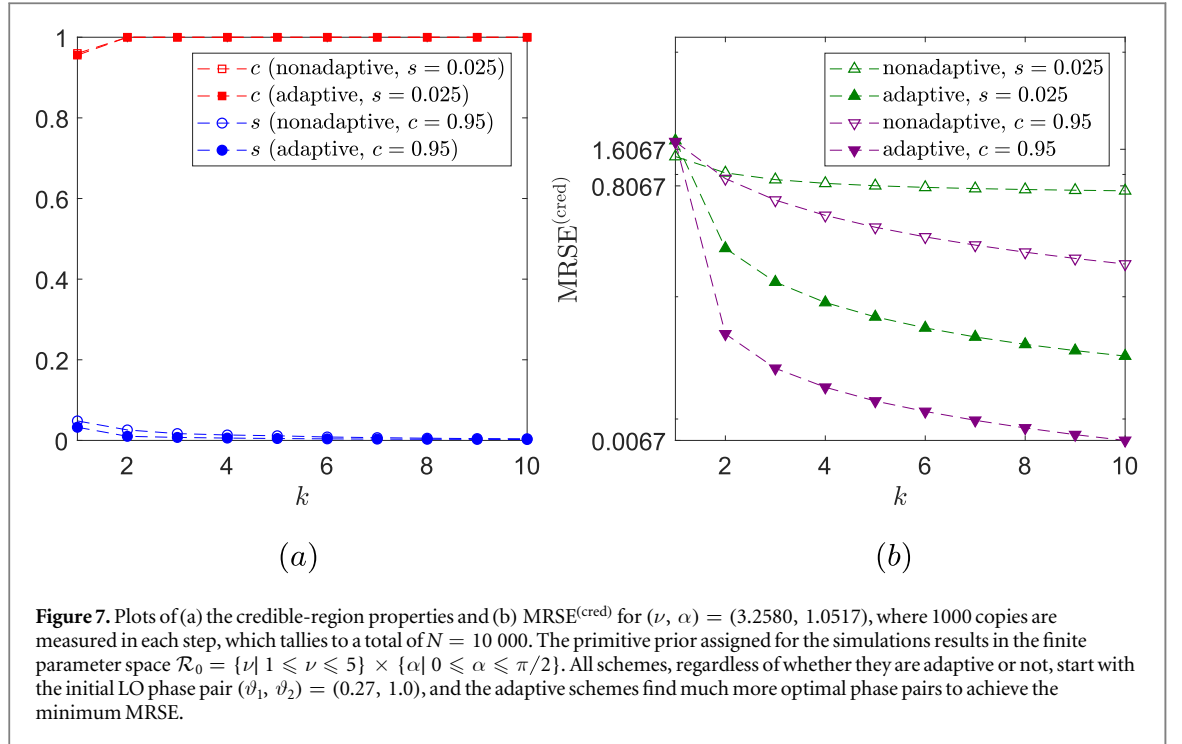

Figure 6. Plots of (a) the plausible-region properties and (b) MRSE$^{(\text{plaus})}$ for the Gaussian squeezed state of parameters $(\phi_1, \phi_2) = (0.5, 1.0)$. Similar behavior is witnessed here with the same figure specifications as figure 5.

$(n_1 + n_2 + n_3 = 3)$. We refer the interested Reader to the supplementary information of [22] for detailed calculations of $\boldsymbol{F}$ and $p_{n_1, n_2, n_3}(\psi_1, \psi_2; \phi_1, \phi_2)$, and instead provide a comparison between adaptive and non-adaptive protocols for such a two-parameter phase estimation problem with figures 5 and 6.

5.3. Squeezed state characterization ($d = 2$)

The third example that we shall investigate is related to Gaussian states, which are important resources in quantum information [23–26]. Every single-mode Gaussian state (of known mean) can be fully specified by the covariance of its Gaussian quasiprobability distribution. For simplicity, we shall again take homodyne detection as the POM for Gaussian state characterization in this section. For a given orientation angle α of its phase-space quasiprobability distribution with respect to the x phase-space ordinate is known, its temperature μ and squeeze parameter ν, the covariance of the Gaussian state is given by

Figure 7. Plots of (a) the credible-region properties and (b) MRSE$^{(\text{cred})}$ for $(\nu, \alpha) = (3.2580, 1.0517)$, where 1000 copies are measured in each step, which tallies to a total of $N = 10\,000$. The primitive prior assigned for the simulations results in the finite parameter space $\mathcal{R}_0 = \{\nu | 1 \leqslant \nu \leqslant 5\} \times \{\alpha | 0 \leqslant \alpha \leqslant \pi/2\}$. All schemes, regardless of whether they are adaptive or not, start with the initial LO phase pair $(\vartheta_1, \vartheta_2) = (0.27, 1.0)$, and the adaptive schemes find much more optimal phase pairs to achieve the minimum MRSE.

$$G = G_{\mu,\nu,\alpha} \triangleq \frac{\mu}{2\nu} \begin{pmatrix} \nu^2(\cos\alpha)^2 + (\sin\alpha)^2 & (1-\nu^2)\sin(2\alpha)/2 \\ (1-\nu^2)\sin(2\alpha)/2 & \nu^2(\sin\alpha)^2 + (\cos\alpha)^2 \end{pmatrix}. \tag{29}$$

The task of characterizing G has been studied in [27–31].

An interesting situation is when μ is preliminarily known (possibly from thermal-light calibrations) and normalized, and we are now interested in characterizing the squeezing properties $r \triangleq (\nu\ \alpha)^{\text{T}}$ of this Gaussian state [32]. It can be shown that if an IC setting $m \triangleq (\vartheta_1\ \vartheta_2)^{\text{T}}$ consisting of a pair of LO phases is measured in such a two-parameter estimation problem, the complete two-dimensional Fisher information $F = \widetilde{F}(\vartheta_1) + \widetilde{F}(\vartheta_2)$ is the sum of its independent Fisher information components, where $\widetilde{F}(\vartheta)$ contains the elements
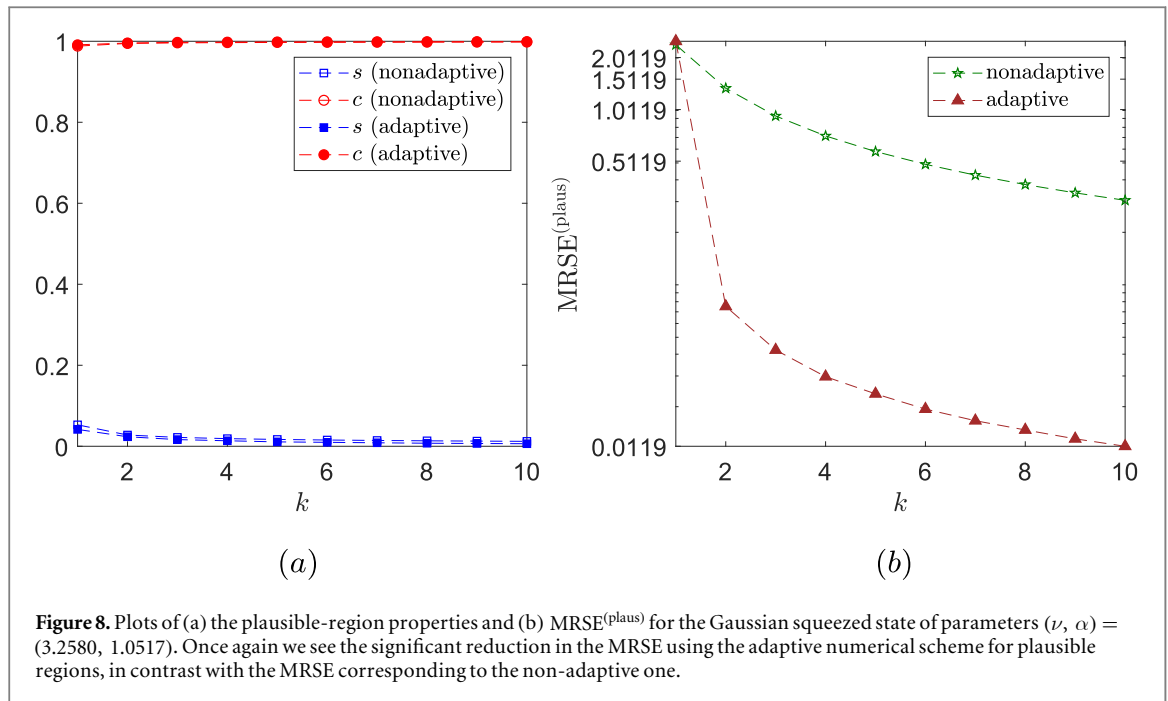
$$
\begin{aligned}
\widetilde{F}_{11}(\vartheta) &= \frac{[\nu^2 + (\nu^2 + 1)\cos(2\alpha + 2\vartheta) - 1]^2}{2\nu^2[\nu^2 + (\nu^2 - 1)\cos(2\alpha + 2\vartheta) + 1]^2}, \\
\widetilde{F}_{22}(\vartheta) &= \frac{2(\nu^2 - 1)^2[\sin(2\alpha + 2\vartheta)]^2}{[\nu^2 + (\nu^2 - 1)\cos(2\alpha + 2\vartheta) + 1]^2}, \\
\widetilde{F}_{12}(\vartheta) &= \frac{(1 - \nu^2)\sin(2\alpha + 2\vartheta)[\nu^2 + (\nu^2 + 1)\cos(2\alpha + 2\vartheta) - 1]}{\nu[\nu^2 + (\nu^2 - 1)\cos(2\alpha + 2\vartheta) + 1]^2}.
\end{aligned}
\tag{30}
$$

Figures 7 and 8 illustrate, once again, how adaptive schemes are indeed extremely useful for constructing much more accurate Bayesian regions for ML estimators whenever the observer begins the parameter estimation experiment with poorly chosen measurement-setting variables, which frequently occurs as r is unknown.

6. Conclusion

The key results of this article revolve around the definition of region accuracy, which is introduced to endow every Bayesian error region with the notion of a frequentist-flavored statistical accuracy (averaged over the entire error region) relative to the unknown true parameter of interest. The region accuracy turns out to do more than just this: it treats the point estimator and its surrounding states within the Bayesian region on equal footing (up to the prior distribution) and endows a MSE collectively. This natural concept elucidates the statistical consequences of either minimizing the credible-region size with fixed credibility, or the dual action of maximizing its credibility with fixed size—both actions increase the region accuracy consistent with our intuitive understanding of these Bayesian regions.

Efforts are then spent on establishing adaptive strategies to optimize region accuracy given only collected data, the dimension of a given estimation problem and no other assumptions about the true parameter. These adaptive procedures are applied to practically interesting examples in quantum metrology and Gaussian state characterization, all of which agree with their positive estimation performance. We believe that these adaptive

Figure 8. Plots of (a) the plausible-region properties and (b) MRSE$^{(\text{plaus})}$ for the Gaussian squeezed state of parameters $(\nu, \alpha) = (3.2580, 1.0517)$. Once again we see the significant reduction in the MRSE using the adaptive numerical scheme for plausible regions, in contrast with the MRSE corresponding to the non-adaptive one.

numerical schemes, together with the asymptotic techniques in the companion article [2], shall form a useful toolkit for Bayesian region construction in practical experimental settings where the dimension of the problem and data sample size are at least moderately large.

Acknowledgments

Appendix A. The derivation of (7)

Following appendix A in [2], we write the numerator of RSE $(\boldsymbol{r})$ as

$$g(\boldsymbol{r}) = \int \frac{\mathrm{d}t}{2\pi\mathrm{i}} \frac{\mathrm{e}^{-\mathrm{i}t\log(\lambda L_{\max})}}{t - \mathrm{i}\epsilon} \int_{\mathcal{R}_0} (\mathrm{d}\,\boldsymbol{r}')(\boldsymbol{r}' - \boldsymbol{r})^2 \, \mathrm{e}^{\mathrm{i}t\log L(\mathbb{D}|\boldsymbol{r}')} \Bigg|_{\epsilon=0}, \tag{A.1}$$

where after using the Gaussian approximation for $L(\mathbb{D}|\boldsymbol{r})$

$$\int_{\mathcal{R}_0} (\mathrm{d}\,\boldsymbol{r}')(\boldsymbol{r}' - \boldsymbol{r})^2 \, \mathrm{e}^{\mathrm{i}t\log L(\mathbb{D}|\boldsymbol{r}')}$$

$$\approx \frac{\mathrm{e}^{\mathrm{i}t\log L_{\max}}}{V_{\mathcal{R}_0}} \int \left(\prod_j \mathrm{d}r_j'' \right) \boldsymbol{r}''^2 \, \mathrm{e}^{-\frac{\mathrm{i}t}{2}\boldsymbol{r}'' \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot \boldsymbol{r}''}$$

$$+ \, \mathrm{e}^{\mathrm{i}t\log L_{\max}} \frac{(\hat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r})^2}{V_{\mathcal{R}_0}} \int \left(\prod_j \mathrm{d}r_j'' \right) \mathrm{e}^{-\frac{\mathrm{i}t}{2}\boldsymbol{r}'' \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot \boldsymbol{r}''}. \tag{A.2}$$

The first term of (A.2), which is the second moment for a multivariate Gaussian distribution, can be calculated by noting the calculus identity $\delta \, \mathrm{Det}\{\boldsymbol{A}\} / \delta \boldsymbol{A} = \mathrm{Det}\{\boldsymbol{A}\} \boldsymbol{A}^{-1}$ for any full-rank $\boldsymbol{A}$:

$$\int \left(\prod_j \mathrm{d}r_j'' \right) \boldsymbol{r}''^2 \, \mathrm{e}^{-\frac{\mathrm{i}t}{2} \boldsymbol{r}'' \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot \boldsymbol{r}''}$$

$$= \frac{2\mathrm{i}}{t} \operatorname{Tr} \left\{ \frac{\delta}{\delta \boldsymbol{F}_{\mathrm{ML}}} \int \left(\prod_j \mathrm{d}r_j'' \right) \mathrm{e}^{-\frac{\mathrm{i}t}{2} \boldsymbol{r}'' \cdot \boldsymbol{F}_{\mathrm{ML}} \cdot \boldsymbol{r}''} \right\}$$

$$= \frac{1}{\mathrm{i}t} \left(\frac{2\pi}{\mathrm{i}t} \right)^{d/2} \operatorname{Tr} \{ \operatorname{Det}\{\boldsymbol{F}_{\mathrm{ML}}\}^{-1/2} \boldsymbol{F}_{\mathrm{ML}}^{-1} \}. \tag{A.3}$$

The second term simply amounts to (A.3) in [2]. Altogether we have

$$g(\boldsymbol{r}) = \frac{\operatorname{Det}\{\boldsymbol{F}_{\mathrm{ML}}\}^{-1/2} (2\pi)^{d/2}}{V_{\mathcal{R}_0}}$$

$$\times \int \frac{\mathrm{d}t}{2\pi \mathrm{i}} \frac{\mathrm{e}^{-\mathrm{i}t \log \lambda}}{t - \mathrm{i}\epsilon} \left[\frac{\operatorname{Tr}\{\boldsymbol{F}_{\mathrm{ML}}^{-1}\}}{(\mathrm{i}t)^{d/2+1}} + \frac{(\hat{\boldsymbol{r}}_{\mathrm{ML}} - \boldsymbol{r})^2}{(\mathrm{i}t)^{d/2}} \right]. \tag{A.4}$$

The t integrals can be handled in exactly the same manner depicted in appendix A of [2], which leads to the final answer.

Appendix B. The derivation of (12)

For a d-dimensional full-rank $\boldsymbol{F}$ of fixed determinant $\operatorname{Det}\{\boldsymbol{F}\} = a$ and trace $\operatorname{Tr}\{\boldsymbol{F}\} = b$, the largest eigenvalue λ_d from the ordered sequence $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_d$ must satisfy the trivial inequality $\lambda_d \leqslant b$, and the smallest eigenvalue

$$\lambda_1 = \frac{a}{\lambda_2 \ldots \lambda_d} \geqslant \frac{a}{\lambda_d^{d-1}} \geqslant \frac{a}{b^{d-1}}. \tag{B.1}$$

This automatically bounds

$$\operatorname{Tr}\{\boldsymbol{F}^{-1}\} = \sum_{j=1}^d \frac{1}{\lambda_j} \leqslant \frac{d}{\lambda_1} \leqslant \frac{d \, b^{d-1}}{a} \tag{B.2}$$

from above. Then clearly, if $\operatorname{Tr}\{\boldsymbol{F}\} \leqslant B$ for some large constant B, a property of a trace class Fisher information, the inequality in (12) is achieved.

As a side remark, we remind the Reader that the occasional $p_j = 0$ for some POM and $\boldsymbol{r}$ does not violate the trace-class property of $\boldsymbol{F}$, since these zero-probability events are ignored when defining $L(\mathbb{D}|\boldsymbol{r})$ in the absence of experimental imperfections.

Appendix C. Threshold values for the dual strategies on plausible regions

The task is to decrease the upper bounds of $\operatorname{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r}; c)$ and $\operatorname{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r}; \boldsymbol{F}, s(\boldsymbol{F}))$. For the upper bound of $\operatorname{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r}; c)$, the relevant function of interest is $y_1(c) = \mathrm{e}^{-2\,\Gamma_{d/2}^{-1}(1-c)} \Gamma_{d/2}^{-1}(1 - c)$, which has one global maximum. This maximum stationary point can be obtain by calculating the first-order derivative

$$\frac{\mathrm{d}y_1}{\mathrm{d}c} = \mathrm{e}^{-\Gamma_{d/2}^{-1}(1-c)} \left(\frac{d}{2} - 1 \right)! [\Gamma_{d/2}^{-1}(1 - c)]^{1-d/2} [1 - 2\,\Gamma_{d/2}^{-1}(1 - c)] \tag{C.1}$$

using the derivative identity

$$\frac{\mathrm{d}}{\mathrm{d}z} \Gamma_{d/2}^{-1}(z) = -\mathrm{e}^{\Gamma_{d/2}^{-1}(z)} \left(\frac{d}{2} - 1 \right)! [\Gamma_{d/2}^{-1}(z)]^{1-d/2} \tag{C.2}$$

for the inverse incomplete Gamma function. Setting $\mathrm{d}y_1/\mathrm{d}c = 0$ then gives the solution $c_{\max} = 1 - \Gamma(d/2, 1/2)/(d/2 - 1)!$. To show that $c_{\max}$ is indeed the maximum, one can calculate the second-order derivative

$$\left(\frac{\mathrm{d}}{\mathrm{d}c_{\max}} \right)^2 y_1(c_{\max}) = -2^d [(d/2 - 1)!]^2 \tag{C.3}$$

evaluated at $c = c_{\max}$, which is clearly negative. This implies that beyond $c > c_{\max}$, $y_1(c)$, or the upper bound of $\operatorname{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r}; c)$, decreases monotonically.

To decrease the upper bound of $\operatorname{MRSE}^{(\mathrm{plaus})}(\boldsymbol{r}; \boldsymbol{F}, s(\boldsymbol{F}))$ monotonically, it suffices to obtain the threshold value for $\operatorname{Det}\{\boldsymbol{F}\}$ beyond which $s(\boldsymbol{F})$ drops monotonically. This means we need to look at

$$y_2(x) = \frac{1}{\sqrt{x}} \left[\log \left(\frac{V_{\mathcal{R}_0}^2 x}{(2\pi)^d} \right) \right]^{d/2}, \tag{C.4}$$

which also contains one global maximum. Setting its first-order derivative to zero gives $x_{\max} = (2\pi e)^d / (V_{\mathcal{R}_0}^2)$ and a negative second-order derivative

$$\left(\frac{d}{dx_{\max}} \right)^2 y_2(x_{\max}) = - \left(\frac{V_{\mathcal{R}_0}^2}{2\pi} \right)^{5/2} \frac{d^{d/2-1}}{2} e^{-5d/2}, \tag{C.5}$$

as it should be. So if $\mathrm{Det}\{\mathbf{F}\}$ increases beyond the threshold of $(2\pi e)^d / (V_{\mathcal{R}_0}^2)$, then s will decrease monotonically. These two threshold values (one for s and one for c) coincide with the common value $\lambda_{\mathrm{crit}} = e^{-1/2} \approx 0.6065$.

Appendix D. The conservativeness of averaging (7)

By invoking the asymptotic techniques in [2] used to cope with Case 2 and 3, we recall that the relevant Bayesian region $\mathcal{R}$ is essentially a truncated hyperellipsoid with $\partial \mathcal{R}_0$, and that this truncation may be approximated to a cut by a hyperplane for large N. The general integral $I_{\mathcal{R}} = \int_{\mathcal{R}} (d\mathbf{r}') f(\mathbf{r}')$ is then simply a sum of values of the function f in such an approximated truncated hyperellipsoid.

To simplify matters, we note that the truncated hyperellipsoid is mappable to a truncated hypersphere of some λ-dependent radius $R = R_\lambda$ in the diagonal basis of the covariance for the hyperellipsoid, so that we essentially have $I_{\mathcal{R}} \approx I_{S_{d-1} \setminus \mathrm{cap}_h}$, the truncated hyperspherical integral with a cap of height $0 \leqslant h = h_\lambda \leqslant R$ removed. By invoking the hyperspherical coordinates and taking $f(\mathbf{r}') = \mathbf{r}'^2$ (the squared error with the center of the hypersphere assuming small statistical fluctuation $\mathbf{r} \approx \hat{\mathbf{r}}_{\mathrm{ML}} = \mathbf{0}$), we may write

$$I_{S_{d-1} \setminus \mathrm{cap}_h} = \frac{2 \pi^{(d-1)/2} (1/2)!}{(d/2-1)!} \int_0^{\cos^{-1}(1-h/R)} d\vartheta (\sin \vartheta)^{d-2}$$
$$\times \left[\frac{R^{d+2}}{d+2} - \frac{1}{d+2} \left(\frac{R-h}{\cos \vartheta} \right)^{d+2} \right] \tag{D.1}$$

after some simple geometry. The truncated volume ($f(\mathbf{r}') = 1$) is also given by

$$V_{S_{d-1} \setminus \mathrm{cap}_h} = \frac{2 \pi^{(d-1)/2} (1/2)!}{(d/2-1)!} \int_0^{\cos^{-1}(1-h/R)} d\vartheta (\sin \vartheta)^{d-2}$$
$$\times \left[\frac{R^d}{d} - \frac{1}{d} \left(\frac{R-h}{\cos \vartheta} \right)^d \right]. \tag{D.2}$$

Altogether, we have $\mathrm{RSE} = \mathrm{RSE}_d(\mathbf{r}; h, R) = I_{S_{d-1} \setminus \mathrm{cap}_h} / V_{S_{d-1} \setminus \mathrm{cap}_h}$.

The simple symmetry fact that $\mathrm{RSE}_d(\mathbf{r}; 0, R) = \mathrm{RSE}_d(\mathbf{r}; R, R)$ forms the first key condition for the conservativeness proof of (8). The second key condition for the proof is obtained by observing that the fraction of the integrand for $I_{S_{d-1} \setminus \mathrm{cap}_h}$ to that for $V_{S_{d-1} \setminus \mathrm{cap}_h}$ strictly increases with ϑ. It therefore suffices to prove the following mathematical lemma for discrete summations and carry it over to integrations, which are also limited summations:

Lemma. *Let $\{a_j \geqslant 0\}_{j=0}^N$ and $\{b_j \geqslant 0\}_{j=0}^N$ for which $a_0/b_0 = \sum_{j=0}^N a_j / \sum_{j=0}^N b_j$ and $a_j/b_j < a_{j+1}/b_{j+1}$ (strictly increasing fractions) for $1 \leqslant j \leqslant N$. Then $\sum_{j=0}^k a_j / \sum_{j=0}^k b_j < a_0/b_0$ for $1 \leqslant k \leqslant N - 1$ and there is exactly one unique minimum value at $k = k^*$.*

Proof. Define $t(k) \equiv \sum_{j=0}^k a_j / \sum_{j=0}^k b_j$ such that $a_0/b_0 = t(0) = t(N)$. We first note that $a_0/b_0 \geqslant (a_0 + a_1)/(b_0 + b_1)$. If not, $a_0/b_0 < a_1/b_1 < \cdots < a_N/b_N$ and we have eventually the inequality $t(k) < t(k+1)$ for $0 \leqslant k \leqslant N - 1$. This contradicts the initial condition $t(0) = t(N)$. Next, the fact that $\exists k = k^* \in [1, N-1] \mid t(k^* + 1) \geqslant t(k^*)$ is obvious, and what remains is to show that k^* is a unique point. This is straightforward since $t(k^*) \leqslant t(k^* + 1) \implies t(k^*) \leqslant a_{k^*+1}/b_{k^*+1}$, and so the strictly increasing fraction chain then tells us that $t(k^*) \leqslant t(m)$ for $k^* < m \leqslant N$. $\square$

The above lemma implies that $\mathrm{RSE} \leqslant \mathrm{RSE}_{\mathrm{CAT}} = \mathrm{RSE}_d(\mathbf{r}; 0, R) = \mathrm{RSE}_d(\mathbf{r}; R, R)$, which implies that the categorical RSE is an overestimate of the actual RSE for any d. The average sum of all these overestimates over all possible data then gives an overestimated MRSE. This is precisely the conservativeness property of (8).

ORCID iDs

Yong Siah Teo ● https://orcid.org/0000-0002-1766-6402

References

[1] Shang J, Ng H K, Sehrawat A, Li X and Englert B G 2013 *New J. Phys.* **15** 123026
[2] Teo Y S, Oh C and Jeong H 2018 *New J. Phys.* **20** 093009
[3] Efron B 1986 *Am. Stat.* **40** 1
[4] Cousins R D 1995 *Am. J. Phys.* **63** 398
[5] Suess D, Rudnicki Ł, Maciel O T and Gross D 2017 *New J. Phys.* **19** 093013
[6] Blume-Kohout R 2010 *New J. Phys.* **12** 043034
[7] Efron B 2015 *J. R. Stat. Soc.* B **77** 617
[8] Yuen H P and Lax M 2010 *IEEE Trans. Inf. Theory* **19** 740
[9] Helstrom C W and Kennedy R S 2010 *IEEE Trans. Inf. Theory* **20** 16
[10] Braunstein S and Caves C M 1994 *Phys. Rev. Lett.* **72** 3439
[11] Nordebo S, Gustafsson M, Khrennikov A, Nilsson B and Toft J 2012 *J. Math. Phys.* **53** 123503
[12] Granade C, Ferrie C, Wiebe N and Cory D G 2012 *New J. Phys.* **14** 103013
[13] Granade C, Combes J and Cory D G 2016 *New J. Phys.* **18** 033024
[14] Caves C M 1981 *Phys. Rev.* D **23** 1693
[15] Dorner U, Demkowicz-Dobrzanski R, Smith B J, Lundeen J S, Wasilewski W, Banaszek K and Walmsley I A 2009 *Phys. Rev. Lett.* **102** 040403
[16] Paesani S, Gentile A, Santagati R, Wang J, Wiebe N, Tew D, O'Brien J and Thompson M 2017 *Phys. Rev. Lett.* **118** 100503
[17] Yuen H P and Chan V W S 1983 *Opt. Lett.* **8** 177
[18] Abbas G L, Chan V W S and Yee T K 1983 *Opt. Lett.* **8** 419
[19] Vogel K and Risken H 1989 *Phys. Rev.* A **40** 2847
[20] Banaszek K and Wódkiewicz K 1997 *Phys. Rev.* A **55** 3117
[21] Olivares S and Paris M G A 2009 *J. Phys. B: At. Mol. Opt. Phys.* **42** 055506
[22] Ciampini M A, Spagnolo N, Vitelli C, Pezzè L, Smerzi A and Sciarrino F 2016 *Sci. Rep.* **6** 28881
[23] Lorenz S, Korolkova N and Leuchs G 2004 *Appl. Phys.* B **79** 273
[24] Lance A M, Symul T, Sharma V, Weedbrook C, Ralph T C and Lam P K 2005 *Phys. Rev. Lett.* **95** 180503
[25] Scarani V, Bechmann-Pasquinucci H, Cerf N J, Dušek M, Lütkenhaus N and Peev M 2009 *Rev. Mod. Phys.* **81** 1301
[26] Weedbrook C, Pirandola S, García-Patrón R, Cerf N J, Ralph T C, Shapiro J H and Lloyd S 2012 *Rev. Mod. Phys.* **84** 621
[27] Řeháček J, Olivares S, Mogilevtsev D, Hradil Z, Paris M G A, Fornaro S, D'Auria V, Porzio A and Solimeno S 2009 *Phys. Rev.* A **79** 032111
[28] Řeháček J, Teo Y S, Hradil Z and Wallentowitz S 2015 *Sci. Rep.* **5** 12289
[29] Müller C R *et al* 2016 *Phys. Rev. Lett.* **117** 070801
[30] Teo Y S, Müller C R, Jeong H, Hradil Z, Řeháček J and Sánchez-Soto L L 2017 *Phys. Rev.* A **95** 042322
[31] Teo Y S, Jeong H and Sánchez-Soto L L 2017 *Phys. Rev.* A **96** 042333
[32] Müller C R, Stoklasa B, Peuntinger C, Gabriel C, Řeháček J, Hradil Z, Klimov A B, Leuchs G, Marquardt C and Sánchez-Soto L L 2012 *New J. Phys.* **14** 085002