数理解析研究所講究録
第1834巻 2013年 109-117

量子推定における有限データの下での推定誤差解析

杉山太香典,[1,*] Peter S. Turner,[1] and 村尾美緒 [1,2]

[1] *Department of Physics, Graduate School of Science, The University of Tokyo.*
[2] *Institute for Nano Quantum Information Electronics, The University of Tokyo.*

We analyze the behavior of estimation errors evaluated by two loss functions, the Hilbert-Schmidt distance and infidelity, in one-qubit state tomography with finite data, and improve these estimation errors by using an adaptive design of experiment. First, we derive an explicit form of a function reproducing the behavior of the estimation errors for finite data by introducing two approximations: a Gaussian approximation of the multinomial distributions of outcomes, and linearizing the boundary. Second, in order to reduce estimation errors, we consider an estimation scheme adaptively updating measurements according to previously obtained outcomes and measurement settings. Updates are determined by the average-variance-optimality (A-optimality) criterion, known in the classical theory of experimental design and applied here to quantum state estimation. We compare numerically two adaptive and two nonadaptive schemes for finite data sets and show that the A-optimality criterion gives more precise estimates than standard quantum tomography.

I. INTRODUCTION

Quantum tomography has become a standard measurement technique in quantum physics. It is especially important in the field of quantum information as it is used for the confirmation of successful experimental implementation of quantum protocols. For example, it can be used to confirm that the quantum states required in a quantum information protocol are sufficiently close to their theoretical targets [1]. In practice, experimental data obtained from tomographic measurements are used to assign a mathematical description to an unknown quantum state or operation, called an estimate. Statistically, this is a constrained multi-parameter estimation problem – the quantum estimation problem – where we assume we are given a finite number of identical copies of a quantum state or process, we perform measurements whose mathematical description is assumed to be known, and from the outcome statistics we make our estimate. Due to the probabilistic behavior of the measurement outcomes and the finiteness of the number of measurement trials, there always exist statistical errors in any quantum estimate. The size of the error depends on the choice of measurements and the estimation procedure. In statistics, the former is called an *experimental design*, while the latter is called an *estimator*. It is, therefore, a key aim of quantum estimation theory to evaluate precisely the size of the estimation error for a given combination of experimental design and estimator and to find a combination of experimental design and estimator which gives us more precise estimation results using fewer measurement trials.

* sugiyama@eve.phys.s.u-tokyo.ac.jp

A. Evaluation of estimation errors

For evaluating the size of the estimation error, we introduce a distance-like function, called a loss function, between the estimate and the true operator. One way to evaluate estimation errors using a loss function is an expected loss, which is the statistical expectation value of the loss function over all possible data sets. In quantum information experiments, the infidelity (one minus the fidelity) and the trace distance are often used as loss functions for state estimation. These evaluations are often performed in the theoretical limit of infinite data, called the asymptotic regime. The asymptotic behavior of these expected losses for this combination has been well studied [2, 3]. Using the asymptotic theory of parameter estimation, we can show that for a sufficiently large number of measurement trials, N, there is a lower bound of the expected losses, called the Cramér-Rao bound. It is known that a maximum likelihood estimator achieves the Cramér-Rao bound asymptotically, and that those expected losses decrease as $O(1/N)$.

In practice of course, no experiment produces infinitely many data, and there are problems in applying the asymptotic theory of expected losses to finite data sets. First of all, the Cramér-Rao inequality holds only for a specific class of estimators, namely those that are unbiased. A maximum likelihood estimator is asymptotically unbiased, but is not unbiased for finite N, so the expected losses can be smaller than the bound for finite N. Particularly, when the purity of the true density matrix becomes high, the bias becomes larger. This is due to the boundary in the parameter space imposed by the condition that density matrices be positive semidefinite, and the expected losses can deviate significantly from the asymptotic behavior [4, 5]. A natural question is then to ask at what value of N the expected losses begin to behave asymptotically. If N is large enough for the effect of the bias to be negligible, we can safely apply the asymptotic theory for evaluating the estimation error in an experiment. However, in general, determining the effects of

the bias is a difficult problem.

In this material, as the first step towards solving the problem, we clarify the effect of the bias on the estimation errors for one-qubit state tomography with finite data sets. By introducing two simple approximations, we are able to qualitatively reproduce the behavior of estimate errors for one-qubit state estimation.

B. Improvement of estimation errors

A standard combination in quantum information experiments is that of quantum tomography and maximum likelihood estimator. Although the term "quantum tomography" can be used in several different contexts, we use it to mean an experimental design in which an independently and identically prepared set of measurements are used throughout the entire experiment [1]. The performance of different choices for the set of tomographic measurements have been studied, in, for example, [4, 6]. This of course raises the question of the performance of *adaptive* experimental designs, in which the measurements performed from trial to trial are not independent, and are chosen according to previous measurement settings and the outcomes obtained. Clearly, adaptive experimental designs are a superset of the nonadaptive ones, and as such can potentially achieve higher performance.

Adaptive designs are characterized by the way in which measurements are related from trial to trial, referred to as an *update criterion*. Previously proposed update criteria include those based on asymptotic statistical estimation theory (Fisher information) [7–9], direct calculations of the estimates expected to be obtained in the next measurement [10, 11], mutually unbiased basis [12], as well as Bayesian estimators and Shannon entropy [10, 13, 14]. Theoretical investigations report that some of the proposed update criteria give more precise estimates than nonadaptive quantum tomography. Experimental implementations of the update criteria proposed in [10] and in [9] have been performed in an ion trap system [15] and in an optical system [16], respectively. If N denotes the number of measurement trials and N is sufficiently large, it is known in 1-qubit state estimation that the expectation value of infidelity averaged over states, a measure of the estimation error, can decrease at best as $O(N^{-3/4})$ in a nonadaptive experiment [17], compared to $O(N^{-1})$ in adaptive experiments [18]. Most of the proposed update criteria, however, have high computational cost that makes real experiments infeasible.

In this material, we propose an adaptive experimental design whose average expected infidelity decreases as $O(N^{-1})$ and whose update criterion, known as average-variance optimality (A-optimality) in classical statistics, has low computational cost for one-qubit state estimation.

II. SUMMARY OF RESULTS

The following is a summary of our results. The details are explained in the Appendix.

A. Evaluation of expected losses

We analyzed the nonasymptotic (finite data) behavior of the expected losses using a maximum likelihood estimator [19]. We derived a simple function which approximates the expected squared Hilbert-Schmidt distance and the expected infidelity between a tomographic maximum likelihood estimate and the true state under two approximations: a Gaussian distribution matched to the moments of the asymptotic multinomial distribution, and a linearization of the parameter space boundary imposed by the positivity of quantum states. The form of this function indicates that the boundary effect decreases exponentially as the number of measurement trials N increases, and we were able to obtain a typical number of measurement trials N^* which can be used for judging whether the expected losses start to converge to the asymptotic behavior.

We performed Monte Carlo simulations of one-qubit state tomography and evaluated the accuracy of the approximation formulas by comparing them to the numerical results. Panels (EHS) and (EIF) in Figure 1 show the pointwise expected squared Hilbert-Schmidt distance and the expected infidelity, respectively. In these two panels, the line styles are as follows: a solid black line for the numerically simulated expected loss, a dashed red line for the approximate expected loss, a chain green line for the Cramér-Rao bound, and a dotted black vertical line for the typical number of measurement trials N^*. The numerical comparison shows that our approximation reproduces the behavior in the nonasymptotic regime much better than the asymptotic theory, and the typical number of measurement trials derived from the approximation is a reasonable threshold after which the expected loss starts to converge to the asymptotic behavior.

B. Improvement of expected losses

In order to improve the estimation error, we considered adaptive experimental design and applied a measurement update method known in statistics as the A-optimality criterion to one-qubit mixed state estimation using arbitrary rank-1 projective measurements [5]. We derived an analytic solution of the A-optimality update procedure in this case, reducing the complexity of measurement updates considerably. Our analytic solution is applicable to any case in which the loss function can be approximated by a quadratic function to least order.

We performed Monte Carlo simulation of this and several nonadaptive schemes in order to compare the behavior of estimation errors for a finite number of mea-

surement trials. We compared the average and pointwise expected squared Hilbert-Schmidt distance and infidelity of the following four measurement update criteria. Panel (Aopt) in Figure 1 shows the pointwise expected infidelity. In the panel, the line style is as follows: a solid black line for the numerically simulated expected infidelity of standard quantum state tomography (repetition of three orthogonal projective measurements), and a dashed blue line for that of the A-optimality update scheme for the infidelity. The numerical results show that A-optimality gives more precise estimates than standard quantum state tomography with respect to the expected infidelity.

ACKNOWLEDGMENTS

T. S. would like to thank R. Blume-Kohout and C. Ferrie for their correspondence, as well as F. Tanaka for helpful discussion on mathematical statistics and Terumasa Tadano for useful advice on numerical simulation. Additionally, P. S. T. would like to thank D. Mahler, L. Rozema and A. Steinberg for discussions pointing out the importance of this problem. This work was supported by JSPS Research Fellowships for Young Scientists (22-7564) and Project for Developing Innovation Systems of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

APPENDIX

Appendix A: Evaluation of estimation errors with finite data

1. Preliminaries

In this subsection, we give a brief review of known results in quantum state tomography and asymptotic estimation theory. The purpose of quantum state tomography is to identify the density matrix characterizing the state of a quantum system of interest. Here we only consider states of a single qubit. Let $\mathcal{H}$ be the 2-dimensional Hilbert space $\mathbb{C}^2$ and $\mathcal{S}(\mathbb{C}^2)$ be the set of all positive semidefinite density matrices acting on $\mathcal{H}$. Such a density matrix ρ can be parametrized as

$$\rho(\boldsymbol{s}) = \frac{1}{2}(\mathbb{1} + \boldsymbol{s}\cdot\boldsymbol{\sigma}), \tag{A1}$$

where $\mathbb{1}$ is the identity matrix on $\mathbb{C}^2$, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)^{\mathrm{T}}$ is the vector of Pauli matrices, and $\boldsymbol{s} \in \mathbb{R}^3$, $\|\boldsymbol{s}\| \leq 1$, is called the Bloch vector. Let us define the parameter space $S := \{\boldsymbol{s} |\ \rho(\boldsymbol{s}) \in \mathcal{S}(\mathbb{C}^2)\}$. Identifying the true density matrix $\rho \in \mathcal{S}(\mathbb{C}^2)$ is equivalent to identifying the true parameter $\boldsymbol{s} \in S$. Let $\boldsymbol{\Pi} = \{\Pi_x\}_{x\in\mathcal{X}}$ denote the POVM characterizing the measurement apparatus used in the tomographic experiment, where $\mathcal{X}$ is the set of measurement outcomes. Like a density matrix, a POVM can be parametrized as

$$\Pi_x = v_x\mathbb{1} + \boldsymbol{w}_x\cdot\boldsymbol{\sigma}, \tag{A2}$$

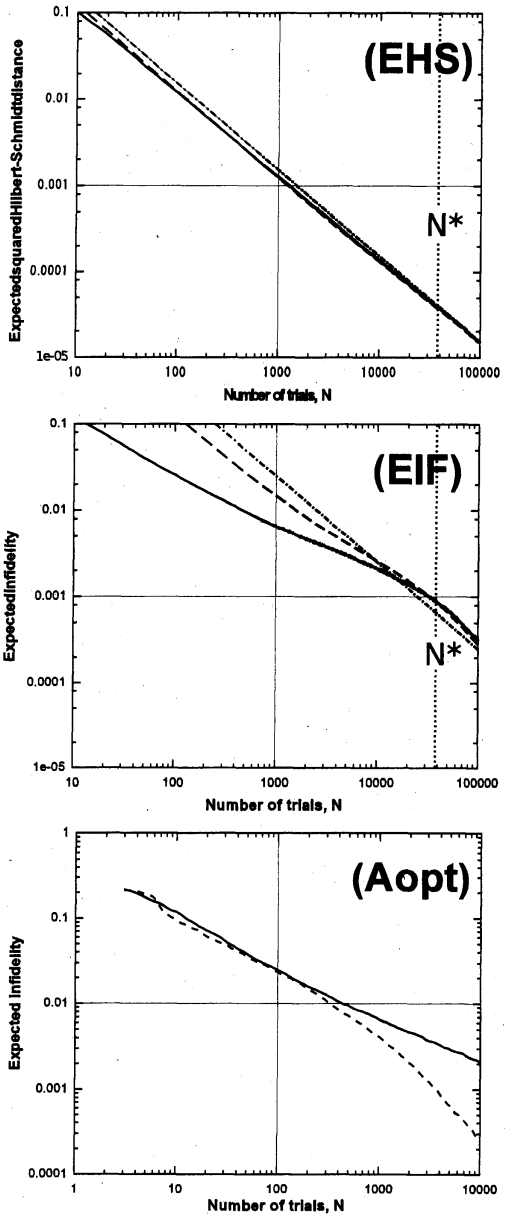

FIG. 1. Pointwise expected losses plotted against the number of measurement trials N for the true Bloch vector $\boldsymbol{s}$ given by $(r, \theta, \phi) = (0.99, \pi/4, \pi/4)$. The other plots are shown in [5, 19]

where $(v_x, \boldsymbol{w}_x) \in \mathbb{R}^4$. When the true density matrix is $\rho(\boldsymbol{s})$, Born's Rule tells us that the probability distribution describing the tomographic experiment is given by

$$p(x|\boldsymbol{s}) = \mathrm{Tr}[\rho(\boldsymbol{s})\Pi_x] \quad \text{(A3)}$$

$$= v_x + \boldsymbol{w}_x \cdot \boldsymbol{s}, \quad \text{(A4)}$$

where Tr denotes the trace operation with respect to $\mathbb{C}^2$. We assume that in the experiment we prepare identical copies of an unknown state $\rho(\boldsymbol{s})$. We perform N measurement trials and obtain a data set $\boldsymbol{x}^N = (x_1, \ldots, x_N)$, where $x_i \in \mathcal{X}$ is the outcome observed in the i-th trial. Let N_x denote the number of times that outcome x occurs in $\boldsymbol{x}^N$, then $f_N(x) := N_x/N$ is the relative frequency of x for the data set $\boldsymbol{x}^N$. In the limit of $N \to \infty$, the relative frequency converges to the true probability $p(x|\boldsymbol{s})$. A POVM is called *informationally complete* if $\mathrm{Tr}[\rho\Pi_x] = \mathrm{Tr}[\rho'\Pi_x]$ has a unique solution ρ' for arbitrary $\rho \in \mathcal{S}(\mathcal{H})$ [20]. This condition is equivalent to that of the POVM $\boldsymbol{\Pi}$ being a basis for the set of all Hermitian matrices on $\mathcal{H}$. For finite N, the relative frequency and true probability are generally not the same, *i.e.*, there is unavoidable statistical error, and we need to choose an estimation procedure that takes the experimental result $\boldsymbol{x}^N$ to a density matrix, that is, we need an estimator.

It is natural to consider a linear estimator, which demands that we find a 2×2 matrix ρ_N^{li} satisfying

$$\mathrm{Tr}[\rho_N^{\mathrm{li}}\Pi_x] = f_N(x), \ x \in \mathcal{X}. \quad \text{(A5)}$$

However, Eq.(A5) does not always have a solution, and even when it does, although the solution is Hermitian and normalized, it is not guaranteed that ρ_N^{li} is positive semidefinite. Let us explore this point further in the one qubit case. The positive semidefinite condition restricts the physically permitted parameter region to the ball $B := \{\boldsymbol{s} \in \mathbb{R}^3 | \|\boldsymbol{s}\| \leq 1\}$. On the other hand, a linear estimate is a random variable that can take values anywhere in the cube $C := \{\boldsymbol{s} \in \mathbb{R}^3 | -1 \leq s_\alpha \leq 1, \alpha = 1,2,3\}$. There is therefore a 'gap' between B and C, consisting of unphysical linear estimates. When the true Bloch parameter $\boldsymbol{s}$ is in the interior of B and N becomes sufficiently large, the probability that linear estimates are out of B becomes negligibly small. However, when the Bloch vector is on the boundary of B, or when N is not sufficiently large, the effect of unphysical linear estimates cannot be ignored. A maximum likelihood estimator ρ^{ml} is one way to address these problems. The estimated density matrix and the Bloch vector are defined as

$$\rho_N^{\mathrm{ml}} := \mathrm{argmax}_{\rho \in \mathcal{S}(\mathcal{H})} \prod_{i=1}^N \mathrm{Tr}[\rho\Pi_{x_i}], \quad \text{(A6)}$$

$$\boldsymbol{s}_N^{\mathrm{ml}} := \mathrm{argmax}_{\boldsymbol{s} \in B} \prod_{i=1}^N \mathrm{Tr}[\rho(\boldsymbol{s})\Pi_{x_i}]. \quad \text{(A7)}$$

It can be shown that when $\rho_N^{\mathrm{li}} \in \mathcal{S}(\mathcal{H})$, $\rho_N^{\mathrm{li}} = \rho_N^{\mathrm{ml}}$ holds.

In order to evaluate the precision of estimates, we introduce a loss function. A loss function Δ is a map from $\mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H})$ to $\mathbb{R}$ such that (i) $\forall \rho, \sigma \in \mathcal{S}(\mathcal{H})$, $\Delta(\rho, \sigma) \geq 0$, and (ii) $\forall \rho \in \mathcal{O}, \Delta(\rho, \rho) = 0$. For example, the trace-distance and the infidelity (one minus the fidelity) are loss functions for density matrices. For our loss functions, we use both the squared Hilbert-Schmidt distance Δ^{HS} and the infidelity Δ^{IF} [17] defined as

$$\Delta^{\mathrm{HS}}(\boldsymbol{s}, \boldsymbol{s}') := \frac{1}{2}\mathrm{Tr}\left[\left(\rho(\boldsymbol{s}) - \rho(\boldsymbol{s}')\right)^2\right] \quad \text{(A8)}$$

$$= \frac{1}{4}(\boldsymbol{s} - \boldsymbol{s}')^2, \quad \text{(A9)}$$

$$\Delta^{\mathrm{IF}}(\boldsymbol{s}, \boldsymbol{s}') := 1 - \mathrm{Tr}\left[\sqrt{\sqrt{\rho(\boldsymbol{s})}\rho(\boldsymbol{s}')\sqrt{\rho(\boldsymbol{s})}}\right]^2 \quad \text{(A10)}$$

$$= \frac{1}{2}(1 - \boldsymbol{s} \cdot \boldsymbol{s}' - \sqrt{1 - \|\boldsymbol{s}\|^2}\sqrt{1 - \|\boldsymbol{s}'\|^2}). \quad \text{(A11)}$$

The Hilbert-Schmidt distance is a normalized Euclidean distance in the parameter space, and the infidelity is a conventional loss function used in experiments. We note that the Hilbert-Schmidt distance coincides with the trace distance in one-qubit systems, but it does not in general.

The outcomes of quantum measurements are random variables, and the value of the loss function between an estimate and the true density matrix is also a random variable. Thus, in order to evaluate the precision of a general estimator ρ^{est} (not the estimate) for the true density matrix, we use the statistical expectation value of the loss function, called an *expected loss* (sometimes called a risk function)[21]. The explicit form is given by

$$\bar{\Delta}_N(\rho^{\mathrm{est}}|\rho) := \sum_{x^N \in \mathcal{X}^N} p(x^N|\rho)\Delta(\rho_N^{\mathrm{est}}(x^N), \rho). \quad \text{(A12)}$$

The value of the expected loss depends on the choice of the estimator as well as the true density matrix. The latter is of course unknown in an experiment, and one way to eliminate its dependence is to average over all possible true states

$$\bar{\Delta}_N^{\mathrm{ave}}(\rho^{\mathrm{est}}) := \int_{\rho \in \mathcal{S}} d\mu(\rho)\bar{\Delta}_N(\rho^{\mathrm{est}}|\rho), \quad \text{(A13)}$$

where μ is a probability measure on $\mathcal{S}$.

Let us assume that $\|\boldsymbol{s}\| < 1$. For any unbiased estimator $\boldsymbol{s}^{\mathrm{est}}$ and any positive semidefinite matrix $H_{\boldsymbol{s}}$, the inequality

$$\bar{\Delta}_N(\boldsymbol{s}^{\mathrm{est}}|\boldsymbol{s}) := \sum_{x^N \in \mathcal{X}^N} p(x^N|\boldsymbol{s})[\boldsymbol{s}_N^{\mathrm{est}}(x^N) - \boldsymbol{s}]^T H_{\boldsymbol{s}}[\boldsymbol{s}_N^{\mathrm{est}}(x^N) - \boldsymbol{s}]$$

$$\geq \frac{1}{N}\mathrm{tr}[H_{\boldsymbol{s}}F_{\boldsymbol{s}}^{-1}] \quad \text{(A14)}$$

holds, where

$$F_{\boldsymbol{s}} := \sum_{x \in \mathcal{X}} \frac{\nabla_{\boldsymbol{s}} p(x|\boldsymbol{s}) \nabla_{\boldsymbol{s}}^T p(x|\boldsymbol{s})}{p(x|\boldsymbol{s})}, \quad \text{(A15)}$$

$$= \sum_{x \in \mathcal{X}} \frac{\boldsymbol{w}_x \boldsymbol{w}_x^T}{v_x + \boldsymbol{w}_x \cdot \boldsymbol{s}} \quad \text{(A16)}$$

is called the Fisher matrix and tr denotes the trace operation with respect to the parameter space $\mathbb{R}^3$. Equation

(B1) is called the Cramér-Rao inequality, and it holds not only for one-qubit state tomography, but also for arbitrary finite dimensional parameter estimation problems under some regularity condition [22]. The matrix $F_{\boldsymbol{s}}$ is a 3×3 positive semidefinite matrix for $\boldsymbol{s} \in \mathbb{R}^3$. It is known that a maximum likelihood estimator asymptotically achieves the equality of Eq.(B1) [22]. From the explicit formulas for the squared Hilbert-Schmidt distance and infidelity in Eqs. (A9) and (A11), we have

$$\Delta^{\mathrm{HS}}(\boldsymbol{s}, \boldsymbol{s}') = (\boldsymbol{s}' - \boldsymbol{s})^T \frac{1}{4} I (\boldsymbol{s}' - \boldsymbol{s}), \tag{A17}$$

$$\Delta^{\mathrm{IF}}(\boldsymbol{s}, \boldsymbol{s}') = (\boldsymbol{s}' - \boldsymbol{s})^T \frac{1}{4} \Big(I + \frac{\boldsymbol{s}\boldsymbol{s}^T}{1 - \|\boldsymbol{s}\|^2} \Big) (\boldsymbol{s}' - \boldsymbol{s}) + O(\|\boldsymbol{s}' - \boldsymbol{s}\|^3), \tag{A18}$$

where I is the identity matrix on $\mathbb{R}^3$. Therefore when we use the Hilbert-Schmidt distance as our loss function, we substitute $H_{\boldsymbol{s}}$ in Eq. (B1) by $H_{\boldsymbol{s}}^{\mathrm{HS}} := \frac{1}{4} I$. On the other hand, when our loss function is the infidelity, we must use $H_{\boldsymbol{s}}^{\mathrm{IF}} := \frac{1}{4}\Big(I + \frac{\boldsymbol{s}\boldsymbol{s}^T}{1 - \|\boldsymbol{s}\|^2} \Big)$. These two matrices $H_{\boldsymbol{s}}^{\mathrm{HS}}$ and $H_{\boldsymbol{s}}^{\mathrm{IF}}$ are half of the Hesse matrices for Δ^{HS} and Δ^{IF}, respectively.

2. Theoretical analysis

In this subsection, we derive a function which approximates the expected losses of the squared Hilbert-Schmidt distance and infidelity for finite data sets.

a. *Two approximations*

In general, the explicit form of expected losses with finite data sets is extremely complicated. In this presentation, we try to derive not the exact form but a simpler function which reproduces the behavior of the true function accurately enough to help us understand the boundary effect. In order to accomplish this, we introduce two approximations. First, we approximate the multinomial distribution generated by successive trials by a Gaussian distribution. Second, we approximate the spherical boundary by a plane tangent to its boundary.

From the central limit theorem, we can readily prove that the distribution of a linear estimator $\boldsymbol{s}^{\mathrm{li}}$ converges to a Gaussian distribution with mean $\boldsymbol{s}$ and covariance matrix $F_{\boldsymbol{s}}^{-1}$. For finite N, we approximate the true probability distribution by the Gaussian distribution

$$p_G(\boldsymbol{s}_N^{\mathrm{li}} | \boldsymbol{s}) := \frac{N^{3/2}}{(2\pi)^{3/2} \sqrt{\det F_{\boldsymbol{s}}^{-1}}} \times \exp\Big[-\frac{N}{2} (\boldsymbol{s}_N^{\mathrm{li}} - \boldsymbol{s}) \cdot F_{\boldsymbol{s}} (\boldsymbol{s}_N^{\mathrm{li}} - \boldsymbol{s}) \Big] \tag{A19}$$

We will refer to this as the Gaussian distribution approximation (GDA).

For a one-qubit system, the boundary between the physical and unphysical regions of the state space is a sphere with unit radius. Despite its simplicity, it is difficult to derive the explicit formula of a maximum likelihood estimator even in this case. Indeed, this is a major contributor to the general complexity of the expected loss behavior in quantum tomography. We therefore choose the simplest possible way to approximate the boundary, namely by replacing it with a plane in the state space. Suppose that the true Bloch vector is $\boldsymbol{s} \in B$. The boundary of the Bloch ball, ∂B, is represented as

$$\partial B := \{ \boldsymbol{s}' \in \mathbb{R}^3 | \ \|\boldsymbol{s}'\| = 1 \}. \tag{A20}$$

We approximate this by the tangent plane to the sphere at the point $\boldsymbol{e}_{\boldsymbol{s}} := \boldsymbol{s}/\|\boldsymbol{s}\|$, represented as

$$\partial D_{\boldsymbol{s}} := \{ \boldsymbol{s}' \in \mathbb{R}^3 | \ \boldsymbol{s} \cdot (\boldsymbol{s}' - \boldsymbol{e}_{\boldsymbol{s}}) = 0 \}, \tag{A21}$$

and so the approximated parameter space is represented as

$$D_{\boldsymbol{s}} = \{ \boldsymbol{s}' \in \mathbb{R}^3 | \ \boldsymbol{s} \cdot (\boldsymbol{s}' - \boldsymbol{e}_{\boldsymbol{s}}) \leq 0 \}. \tag{A22}$$

We will refer to this as the linear boundary approximation (LBA).

b. *Approximated maximum likelihood estimator*

In [23], it is proved that the distribution of a maximum likelihood estimator in a constrained parameter estimation problem converges to the distribution of the following vector

$$\tilde{\boldsymbol{s}}_N^{\mathrm{ml}} := \mathrm{argmin}_{\boldsymbol{s}' \in D_{\boldsymbol{s}}} (\boldsymbol{s}_N^{\mathrm{li}} - \boldsymbol{s}') \cdot F_{\boldsymbol{s}} (\boldsymbol{s}_N^{\mathrm{li}} - \boldsymbol{s}'). \tag{A23}$$

By using the Lagrange multiplier method, we can derive the approximated maximum likelihood estimates as

$$\tilde{\boldsymbol{s}}_N^{\mathrm{ml}} = \begin{cases} \boldsymbol{s}_N^{\mathrm{li}} & (\boldsymbol{s}_N^{\mathrm{li}} \in D_{\boldsymbol{s}}) \\ \boldsymbol{s}_N^{\mathrm{li}} - \frac{\boldsymbol{e}_{\boldsymbol{s}} \cdot \boldsymbol{s}_N^{\mathrm{li}} - 1}{\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-1} \boldsymbol{e}_{\boldsymbol{s}}} F_{\boldsymbol{s}}^{-1} \boldsymbol{e}_{\boldsymbol{s}} & (\boldsymbol{s}_N^{\mathrm{li}} \notin D_{\boldsymbol{s}}) \end{cases}. \tag{A24}$$

c. *Expected squared Hilbert-Schmidt distance*

From a straightforward calculation using formulas for Gaussian integrals, we can derive the approximate expected squared Hilbert-Schmidt distance.

$$\bar{\Delta}_N^{\mathrm{HS}}(\tilde{\boldsymbol{s}}^{\mathrm{ml}} | \boldsymbol{s}) = \frac{1}{4N} \Big(\mathrm{tr}[F_{\boldsymbol{s}}^{-1}] - \frac{1}{2} \frac{\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-2} \boldsymbol{e}_{\boldsymbol{s}}}{\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-1} \boldsymbol{e}_{\boldsymbol{s}}} \mathrm{erfc}\Big[\sqrt{\frac{N}{N^*}} \Big] \Big) - \frac{1}{4} \frac{1 - \|\boldsymbol{s}\|}{\sqrt{2\pi \boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-1} \boldsymbol{e}_{\boldsymbol{s}}}} \frac{\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-2} \boldsymbol{e}_{\boldsymbol{s}}}{\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-1} \boldsymbol{e}_{\boldsymbol{s}}} \frac{e^{-N/N^*}}{\sqrt{N}} + \frac{1}{8} (1 - \|\boldsymbol{s}\|^2) \frac{\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-2} \boldsymbol{e}_{\boldsymbol{s}}}{(\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-1} \boldsymbol{e}_{\boldsymbol{s}})^2} \mathrm{erfc}\Big[\sqrt{\frac{N}{N^*}} \Big], \tag{A25}$$

where

$$\mathrm{erfc}[a] := \frac{2}{\sqrt{\pi}} \int_a^\infty dt \ e^{-t^2} \tag{A26}$$

is the complementary error function and

$$N^* := 2\frac{\boldsymbol{e}_{\boldsymbol{s}} \cdot F_{\boldsymbol{s}}^{-1}\boldsymbol{e}_{\boldsymbol{s}}}{(1-\|\boldsymbol{s}\|)^2} \tag{A27}$$

is a typical scale for the number of trials.

d. *Expected infidelity*

In order to analyze the expected infidelity, we take the Taylor expansion of the infidelity around the true Bloch vector $\boldsymbol{s}$ up to the second order. The explicit form is in Eq. (A18). Again, using formulas for Gaussian integrals we can derive the approximate expected infidelity. When $\|\boldsymbol{s}\| < 1$,

$$\begin{aligned}\bar{\Delta}_N^{\mathrm{IF}}(\hat{\boldsymbol{s}}^{\mathrm{ml}}|\boldsymbol{s}) = \frac{1}{4}\Big(\mathrm{tr}[F_{\boldsymbol{s}}^{-1}] + \frac{\boldsymbol{s}\cdot F_{\boldsymbol{s}}^{-1}\boldsymbol{s}}{1-\|\boldsymbol{s}\|^2}\Big)\frac{1}{N}\Big(1-\frac{1}{2}\mathrm{erfc}\Big[\sqrt{\frac{N}{N^*}}\Big]\Big)\\ -\frac{1}{4}\frac{1-\|\boldsymbol{s}\|}{\sqrt{2\pi\boldsymbol{e}_{\boldsymbol{s}}\cdot F_{\boldsymbol{s}}^{-1}\boldsymbol{e}_{\boldsymbol{s}}}}\Big(\mathrm{tr}[F_{\boldsymbol{s}}^{-1}] - \mathrm{tr}[(Q_{\boldsymbol{s}}F_{\boldsymbol{s}}Q_{\boldsymbol{s}})^{-}]\\ +\frac{\boldsymbol{s}\cdot F_{\boldsymbol{s}}^{-1}\boldsymbol{s}}{1-\|\boldsymbol{s}\|^2}\Big)\frac{e^{-N/N^*}}{\sqrt{N}}\\ +\frac{1}{4}(1-\|\boldsymbol{s}\|)\mathrm{erfc}\Big[\sqrt{\frac{N}{N^*}}\Big],\end{aligned} \tag{A28}$$

where

$$Q_{\boldsymbol{s}} := I - \boldsymbol{e}_{\boldsymbol{s}}\boldsymbol{e}_{\boldsymbol{s}}^T \tag{A29}$$

is the projection matrix onto the subspace orthogonal to $\boldsymbol{s}$, and A^- is the Moore-Penrose generalized inverse of a matrix A. From the argument above, we can see that the approximate expected infidelity converges to the Cramér-Rao bound in the limit of large N.

Appendix B: Improvement of estimation errors by adaptive design of experiments

1. Preliminaries

a. *Experimental design*

We consider sequential measurements on copies of ρ. We will index measurement trials using subscripts $n \in \{1,2,\ldots,N\}$, and sequences using superscripts. Thus, for some symbol A, A_n is its value taken at the n-th trial, while A^n is the sequence $\{A_1, A_2, \ldots, A_n\}$. We will also try to use calligraphic fonts for supersets. Adaptivity in our sense means that the POVM performed at $(n+1)$-th trial can depend on all the previous n trials' outcomes and POVMs.

The measurement class $\mathcal{M}_n$ is the set of POVMs which are available at the n-th trial. We choose the n-th POVM, $\boldsymbol{\Pi}_n = \{\Pi_{n,x}\}_{x\in\mathcal{X}_n}$ from $\mathcal{M}_n$, where $\mathcal{X}_n$ denotes the set of measurement outcomes for the n-th trial. When it is independent of the trial, as is usually the case, we omit the index, using $\mathcal{M}$ for the measurement class and $\mathcal{X}$ for the outcome set. Let $x^n = \{x_1, \ldots, x_n\}$ denote the sequence of outcomes obtained up to the n-th trial, where $x_i \in \mathcal{X}_i$. We will denote the pair of measurement performed and outcome obtained by $D_n = (\boldsymbol{\Pi}_n, x_n) \in \mathcal{D}_n := \mathcal{M}_n \times \mathcal{X}_n$, and refer to it as the data for trial n. The sequence of data up to trial n is thus $D^n = \{D_1, \ldots, D_n\} \in \mathcal{D}^n := \times_{i=1}^n \mathcal{D}_i$. After the n-th measurement, we choose the next, $(n+1)$-th, POVM $\boldsymbol{\Pi}_{n+1} = \{\Pi_{n+1,x}\}_{x\in\mathcal{X}_{n+1}}$ according to the previously obtained data. Let u_n denote the map from the data to the next measurement, that is, $u_n : \mathcal{D}^{n-1} \to \mathcal{M}_n$, $\boldsymbol{\Pi}_n = u_n(D^{n-1})$. We call u_n the measurement update criterion for the n-th trial and $u^N := \{u_1, u_2, \ldots, u_N\}$ the measurement update rule. Note that u_1 is a map from $\emptyset$ to $\mathcal{M}_1$ and corresponds to the choice of the first measurement.

b. *A generalized Cramér-Rao inequality*

The A-optimality criterion is a measurement update criterion based on the asymptotic theory of statistical parameter estimation [24, 25]. In this subsection we introduce a few basic results of the asymptotic theory. First let us parametrize the state space $\mathcal{S}(\mathcal{H})$. Any density matrix on d-dimensional Hilbert space can be parametrized by d^2-1 real numbers, $\boldsymbol{s} \in \mathbb{R}^{d^2-1}$, i.e. $\rho = \rho(\boldsymbol{s})$. In the $d=2$ case, we take $\rho(\boldsymbol{s}) = \frac{1}{2}(\mathbb{1} + \boldsymbol{s}\cdot\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$, σ_α $(\alpha = 1,2,3)$ are the Pauli matrices, and $\boldsymbol{s} \in \mathbb{R}^3$, $\|\boldsymbol{s}\| \le 1$, is called the Bloch vector. The estimation of ρ is equivalent to the estimation of $\boldsymbol{s}$, and we let $\boldsymbol{s}^{\mathrm{est}}$ denote the estimator. Estimates of a density matrix and of a Bloch vector are related as $\rho_n^{\mathrm{est}}(D^n) = \rho(\boldsymbol{s}_n^{\mathrm{est}}(D^n))$.

For any estimator $\boldsymbol{s}^{\mathrm{est}}$, any number of measurement trials N, and any positive semidefinite matrix $H(\boldsymbol{s})$, the inequality

$$\begin{aligned}&\sum_{D^N\in\mathcal{D}^N} p(D^N|\boldsymbol{s})[\boldsymbol{s}_N^{\mathrm{est}}(D^N) - \boldsymbol{s}]^T H(\boldsymbol{s})[\boldsymbol{s}_N^{\mathrm{est}}(D^N) - \boldsymbol{s}]\\ &\ge \mathrm{tr}[H(\boldsymbol{s})G_N(u^N, \boldsymbol{s}^{\mathrm{est}}, \boldsymbol{s})^T F_N(u^N, \boldsymbol{s})^{-1} G_N(u^N, \boldsymbol{s}^{\mathrm{est}}, \boldsymbol{s})]\end{aligned} \tag{B1}$$

holds, where

$$p(D^N|\boldsymbol{s}) := p(D^N|\rho(\boldsymbol{s})), \tag{B2}$$

$$G_N(u^N, \boldsymbol{s}^{\mathrm{est}}, \boldsymbol{s}) := \nabla_{\boldsymbol{s}} \sum_{D^N\in\mathcal{D}^N} p(D^N|\boldsymbol{s})\boldsymbol{s}_N^{\mathrm{est}T}(D^N), \tag{B3}$$

$$F_N(u^N, \boldsymbol{s}) := \sum_{D^N\in\mathcal{D}^N} \frac{\nabla_{\boldsymbol{s}} p(D^N|\boldsymbol{s}) \nabla_{\boldsymbol{s}}^T p(D^N|\boldsymbol{s})}{p(D^N|\boldsymbol{s})}, \tag{B4}$$

and tr denotes the trace operation with respect to the parameter space. Eq.(B1) is a known generalization of the Cramér-Rao inequality [22]. $F_N(\boldsymbol{s})$ is a $(d^2-1)\times(d^2-1)$ positive semidefinite matrix called the Fisher matrix of the probability distribution $\{p(D^N|\boldsymbol{s})\}_{D^N\in\mathcal{D}^N}$.

If the estimate converges to the true parameter, i.e., $\boldsymbol{s}_N^{\rm est}(D^N) \to \boldsymbol{s}$ as $N \to \infty$ with probability 1, the LHS of Eq.(B1) converges to 0 and therefore the RHS should converge to 0. In this case, if we assume the exchangeability of the limit and derivative, the matrix $G_N(u^N, \boldsymbol{s}^{\rm est}, \boldsymbol{s})$ converges to the identity matrix I, and the quantity $K_N(u^N, \boldsymbol{s})$ defined as

$$K_N(u^N, \boldsymbol{s}) := \mathrm{tr}[H(\boldsymbol{s}) F_N(u^N, \boldsymbol{s})^{-1}] \tag{B5}$$

converges to 0. This $K_N(u^N, \boldsymbol{s})$ can be interpreted as a lower bound of the weighted (by $H(\boldsymbol{s})$) mean squared error when N is sufficiently large. It is known that under certain regularity conditions, a maximum likelihood estimator achieves the equality of Eq.(B1) asymptotically. For a given $\boldsymbol{s}$, it would be wise to choose a measurement update rule which makes the value of $K_N(u^N, \boldsymbol{s})$ as small as possible. This is the guiding principle of the A-optimality criterion.

c. *A-optimality criteria*

We move on to the explanation of the procedure of A-optimality. The "A" stands for "average-variance" [25]. According to the asymptotic theory of statistical parameter estimation described in the previous subsection, we wish to minimize the value of $K_N(u^N, \boldsymbol{s})$. Suppose that we perform n trials and obtained the data sequence D^n. We would like to choose the POVM minimizing $K_{n+1}(u^N, \boldsymbol{s})$ in $\mathcal{M}_{n+1}$ as the next, $(n+1)$-th, measurement. When we consider minimizing this function, there are two problems. In order to avoid them, we introduce two approximations. The first problem is that the minimized function depends on the true parameter $\boldsymbol{s}$. Of course the true parameter is unknown in parameter estimation problems, and we must use an estimate in the update criterion, $\hat{\boldsymbol{s}}_n^{\rm est}(D^n)$, instead. The mesurement update estimator $\hat{\boldsymbol{s}}^{\rm est}$ is not necessarily the same as $\boldsymbol{s}^{\rm est}$. The second problem is that unlike the independent and identically distributed (i.i.d.) measurement case, calculation of the Fisher matrix in the adaptive case requires summing over an exponential amount of data, and is computationally intensive. To avoid this problem, we approximate the sum over all possible measurements by that over only those measurements that have been performed:

$$F_{n+1}(u^{n+1}, \boldsymbol{s}) \approx \tilde{F}_{n+1}(u^{n+1}, \boldsymbol{s}|D^n) \tag{B6}$$

$$:= \sum_{i=1}^{n+1} F(\boldsymbol{\Pi}_i, \boldsymbol{s}), \tag{B7}$$

where

$$F(\boldsymbol{\Pi}_i, \boldsymbol{s}) := \sum_{x_i \in \mathcal{X}_i} \frac{\nabla_{\boldsymbol{s}} p(x_i; \boldsymbol{\Pi}_i|\boldsymbol{s}) \nabla_{\boldsymbol{s}}^T p(x_i; \boldsymbol{\Pi}_i|\boldsymbol{s})}{p(x_i; \boldsymbol{\Pi}_i|\boldsymbol{s})}, \tag{B8}$$

$$\boldsymbol{\Pi}_i = u_i(D^{i-1}),\ i = 1, \cdots, n+1. \tag{B9}$$

The matrix $F(\boldsymbol{\Pi}_i, \boldsymbol{s})$ is the Fisher matrix for the i-th measurement probability distribution $\{p(x_i; \boldsymbol{\Pi}_i|\boldsymbol{s})\}_{x_i \in \mathcal{X}_i}$, and $\tilde{F}_{n+1}(u^{n+1}, \boldsymbol{s}|D^n)$ is the sum of the Fisher matrices from the first to the $(n+1)$-th trial. Instead of minimizing $K_{n+1}(u^{n+1}, \boldsymbol{s})$, we consider the minimization of

$$\tilde{K}_{n+1}(u^{n+1}, \boldsymbol{s}|D^n) := \mathrm{tr}[H(\boldsymbol{s}) \tilde{F}_{n+1}(u^{n+1}, \boldsymbol{s}|D^n)^{-1}]. \tag{B10}$$

It is known that the convergence of $\tilde{K}_N(u^N, \boldsymbol{s}|D^N)$ to 0 is part of a sufficient condition for the convergence of a maximum likelihood estimator [26], and this justifies the use of this second approximation. After making these two approximations, we define the A-optimality criterion as

$$\boldsymbol{\Pi}_{n+1}^{\text{A-opt}} := u_{n+1}^{\text{A-opt}}(D^n) = \mathop{\mathrm{argmin}}_{\boldsymbol{\Pi}_{n+1} \in \mathcal{M}_{n+1}} \mathrm{tr}[H(\hat{\boldsymbol{s}}_n^{\rm est}) \tilde{F}_{n+1}(u^{n+1}, \hat{\boldsymbol{s}}_n^{\rm est}|D^n)^{-1}]. \tag{B11}$$

Finding $\boldsymbol{\Pi}_{n+1}^{\text{A-opt}}$ is a nonlinear minimization problem with high computational cost in general. In this paper, we derive the analytic solution of Eq. (B11) in the 1-qubit case, reducing the computational cost significantly.

d. *Estimation setting*

We consider a one-qubit mixed state estimation problem. We identify the Bloch parameter space $\{\boldsymbol{s} \in \mathbb{R}^3 | \|\boldsymbol{s}\| < 1\}$ with $\mathcal{O}$, where we restrict the true state space to be strictly the interior in order to avoid the possible divergence of the Fisher matrix. Suppose that we can choose any rank-1 projective measurement in each trial. Let $\boldsymbol{\Pi}(\boldsymbol{a}) = \{\Pi_x(\boldsymbol{a})\}_{x=\pm}$ denote the POVM corresponding to the projective measurement onto the $\boldsymbol{a}$-axis ($\boldsymbol{a} \in \mathbb{R}^3, \|\boldsymbol{a}\| = 1$), whose elements can be represented as

$$\Pi_\pm(\boldsymbol{a}) = \frac{1}{2}(\mathbb{1} \pm \boldsymbol{a} \cdot \boldsymbol{\sigma}). \tag{B12}$$

This is the Bloch parametrization of projective measurements. We identify the set of parameters $\mathcal{A} = \{\boldsymbol{a} \in \mathbb{R}^3 |\ \|\boldsymbol{a}\| = 1\}$ with the measurement class $\mathcal{M} =$ {All rank-1 projective measurements on a one-qubit system}.

The asymptotic behavior of the average expected fidelity $\bar{\Delta}_N^{\rm IFave}$ is known in the 1-qubit state estimation case [17, 18, 27]. The measure used for calculating this average is the Bures distribution, $d\mu(\boldsymbol{s}) = \frac{1}{\pi^2}(1 - \|\boldsymbol{s}\|^2)^{-1/2} d\boldsymbol{s}$. If we limit our available measurements to be sequential and independent (i.e., nonadaptive), $\bar{\Delta}_N^{\rm IFave}$ behaves at best as $O(N^{-3/4})$ [17, 27]. On the other hand, if we are allowed to use adaptive, separable, or collective measurements, $\bar{\Delta}_N^{\rm IFave}$ can behave as $O(N^{-1})$ [18]. In [17, 18, 27], the coefficient of the dominant term in the asymptotic limit is also derived.

2. Results and analysis

As explained in Sec. B 1 d, we consider the A-optimality criterion for one-qubit state estimation using

projective measurements. In Sec. B 2 a we give the analytic solution.

a. *Analytic solution for A-optimality in 1-qubit state estimation*

First, we give the explicit form of the Fisher matrix for projective measurements. The probability distribution for the rank-1 projective measurement $\boldsymbol{\Pi}(\boldsymbol{a})$ is given by

$$p(\pm;\boldsymbol{a}|\boldsymbol{s}) = \tfrac{1}{2}(1 \pm \boldsymbol{s}\cdot\boldsymbol{a}), \tag{B13}$$

and the Fisher matrix is

$$F(\boldsymbol{a},\boldsymbol{s}) = \frac{\boldsymbol{a}\boldsymbol{a}^T}{1-(\boldsymbol{a}\cdot\boldsymbol{s})^2}. \tag{B14}$$

In this case, Eq. (B11) is rewritten in the Bloch vector representation as

$$\boldsymbol{a}_{n+1}^{\text{A-opt}} := \operatorname*{argmin}_{\boldsymbol{a}\in\mathcal{A}} \operatorname{tr}\left[H(\hat{\boldsymbol{s}}_n^{\text{est}})\{\tilde{F}_n(\boldsymbol{a}^n,\hat{\boldsymbol{s}}_n^{\text{est}}|D^n) + F(\boldsymbol{a},\hat{\boldsymbol{s}}_n^{\text{est}})\}^{-1}\right]. \tag{B15}$$

We present the analytic solution of Eq.(B15) in the form of the following theorem.

Theorem 1 *Given a sequence of data $D^n = \{(\boldsymbol{a}_1,x_1),\ldots,(\boldsymbol{a}_n,x_n)\}$, the n-th estimate $\hat{\boldsymbol{s}}_n^{\text{est}}$, and a real positive matrix H, the A-optimal POVM Bloch vector is given by*

$$\boldsymbol{a}_{n+1}^{\text{A-opt}} = \frac{B_n \boldsymbol{e}_{\min}(C_n)}{\|B_n \boldsymbol{e}_{\min}(C_n)\|}, \tag{B16}$$

where

$$B_n = \sqrt{\tilde{F}_n(\boldsymbol{a}^n,\hat{\boldsymbol{s}}_n^{\text{est}}|D^n)H(\hat{\boldsymbol{s}}_n^{\text{est}})^{-1}\tilde{F}_n(\boldsymbol{a}^n,\hat{\boldsymbol{s}}_n^{\text{est}}|D^n)}, \tag{B17}$$

$$C_n = B_n(I - \hat{\boldsymbol{s}}_n^{\text{est}}\hat{\boldsymbol{s}}_n^{\text{est}\,T} + \tilde{F}_n(\boldsymbol{a}^n,\hat{\boldsymbol{s}}_n^{\text{est}}|D^n)^{-1})B_n, \tag{B18}$$

$\boldsymbol{e}_{\min}(C_n)$ is the eigenvector of the matrix C_n corresponding to the minimal eigenvalue, and I is the identity in the parameter space.

Here we omit the proof of Theorem 1. That is in the Appendix of [5].

[1] M. Paris and J. Řeháček, eds., *Quantum State Estimation*, Lecture Notes in Physics (Springer, Berlin, 2004).
[2] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[3] M. Hayashi, ed., *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (World Scientific, Singapore, 2005).
[4] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, Phys. Rev. A **78**, 052122 (2008).
[5] T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. A **85**, 052107 (2012).
[6] J. Nunn, B. J. Smith, G. Puentes, I. A. Walmsley, and J. S. Lundeen, Phys. Rev. A **81**, 042109 (2010).
[7] H. Nagaoka, in *Proc. Int. Symp. on Inform. Theory* (1988), p. 198.
[8] H. Nagaoka, in *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, edited by M. Hayashi (World Scientific, 2005), chap. 10.
[9] A. Fujiwara, J. Phys. A: Math. Gen. **39** (2006).
[10] D. G. Fischer, S. H. Kienle, and M. Freyberger, Phys. Rev. A **61**, 032306 (2000).
[11] C. J. Happ and M. Freyberger, Phys. Rev. A **78**, 064303 (2008).
[12] C. J. Happ and M. Freyberger, Eur. Phys. J. D **64**, 579 (2011).
[13] D. G. Fischer and M. Freyberger, Phys. Lett. A **273**, 293 (2000).
[14] F. Huszár and N. M. T. Houlsby (2011), quant-ph/1107.0895.
[15] T. Hannemann, D. Reiss, C. Balzer, W. Neuhauser, P. E. Toschek, and C. Wunderlich, Phys. Rev. A **65**, 050303 (2002).
[16] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, Phys. Rev. Lett. **109**, 130404 (2012).
[17] E. Bagan, M. Baig, R. Muñoz-Tapia, and A. Rodriguez, Phys. Rev. A **69**, 010304(R) (2004).
[18] E. Bagan, M. A. Ballester, R. D. Gill, R. Muñoz-Tapia, and O. Romero-Isart, Phys. Rev. Lett. **97**, 130501 (2006).
[19] T. Sugiyama, P. S. Turner, and M. Murao, New J. Phys. (2012).
[20] E. Prugovečki, Int. J. Theor. Phys. **16**, 321 (1977).
[21] Note1, there are also different approaches to evaluating the precision of estimators, including error probabilities [28], region estimators [29, 30].
[22] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley series in probability and statistics (Wiley, New York, 2002), 2nd ed., (originally published in 1973).
[23] S. G. Self and K. Y. Liang, J. Am. Stat. Assoc. **82**, 605 (1987).
[24] S. Watanabe, K. Hagiwara, S. Akaho, Y. Motomura, K. Fukumizu, M. Okada, and M. Aoyagi, *Theory and Implimentation of Learning Systems* (Morikita Publishing Co., 2005), (in Japanese).

[25] F. Pukelsheim, *Optimal Design of Experiments*, Classics in Applied Mathematics (SIAM, Philadelphia, 2006).
[26] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, Probability and mathematical statistics (Academic Press, New York, 1980).
[27] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz-Tapia, Phys. Rev. A **73**, 032301 (2006).
[28] T. Sugiyama, P. S. Turner, and M. Murao, Phys. Rev. A **83**, 012105 (2011).
[29] K. M. R. Audenaert and S. Scheel, New J. Phys. **11**, 023028 (2009).
[30] R. Blume-Kohout (2012), quant-ph/1202.5270.