Nuisance parameter problem in quantum estimation theory: tradeoff relation and qubit examples

Jun Suzuki

Graduate School of Informatics and Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585 Japan

E-mail: junsuzuki@uec.ac.jp

Received 10 October 2019, revised 20 March 2020
Accepted for publication 3 April 2020
Published 9 June 2020

Abstract

In this paper, we analyze quantum-state estimation problems when some of the parameters are of no interest to be estimated. In classical statistics, these irrelevant parameters are called nuisance parameters, and this problem is of great importance in many practical applications of statistics. However, little is known regarding the effects of nuisance parameters in quantum-state estimation problems. The main contribution of this paper is first to formulate the nuisance parameter problem in the quantum-state estimation theory for the finite sample size case. We then derive the precision limit for the parameters of interest based on the concept of locally unbiasedness for the parameters of interest. We also propose a method of how to eliminate the nuisance parameters to obtain an upper bound for estimation error of the parameters of interest. Based on the proposed method, we clarify an intrinsic tradeoff relation to estimate the nuisance parameters and parameters of interest. The general qubit models are examined in detail to emphasize that we cannot ignore the effects of the nuisance parameters in general. Several examples are worked out to illustrate our findings.

Keywords: quantum state estimation, nuisance parameter, tradeoff relation

(Some figures may appear in colour only in the online journal)

1. Introduction

This paper addresses an open problem in quantum-state estimation, the nuisance parameter problem, which is the common issue when dealing with the state-estimation problem in
practice. Consider a family of quantum states specified by the large number of unknown parameters. In many quantum information processing tasks, one is not interested in estimating all the parameters but only in a certain subset of parameters. In classical statistics, this subset of relevant parameters is called parameters of interest, whereas the remaining irrelevant parameters appearing in a parametric model are called nuisance parameters. This nuisance parameter problem is of great importance in many practical applications of statistics as was formulated by Fisher [1]. In the classical case, it is known how much estimation errors become worse in the presence of nuisance parameters. Further, many studies were devoted to find effective construction of good estimators. See textbooks [2–5] and classical results in [6–10].

The nuisance parameter problem in the quantum system has just gotten great attention in the quantum information community. This is partly because of advances in quantum metrology in a noisy environment. See, for example, [11–16], where many studies found that imprecise knowledge about environments may ruin advantages of quantum mechanically enhanced precision measurements. Another motivation is more general multi-parameter quantum metrology [17, 18]. In both cases, one faces problems of estimating a certain subset of parameters in the presence of noise parameters, which should be treated as nuisance parameters.

We note in passing that several authors analyzed the nuisance parameter problem for specific examples [19–25] and discussed achievability of bounds. However, most authors simply treated the problem as the multi-parameter problem. For example [19], discussed a tradeoff relation between two kinds of parameters for a specific noise model; the phase and diffusion constant without mentioning this is due to the effect of nuisance parameters. Later, their theoretical result was demonstrated in [20]. One of the key results in the nuisance parameter problem was due to Yang et al [26]. The authors made a solid foundation on the asymptotically achievable bound for the infinite sample size limit based on the Holevo bound. (See also the recent review [27].) However, this problem also needs to be formulated for the finite sample size case to apply any realistic experimental setting. This is one of the primarily motivations to this work.

Here we summarize the main contribution of this paper. We first establish a general formulation for the problem of nuisance parameters in the quantum estimation theory for any finite sample size case. To clarify the meaning of achievability of estimation error bounds, we introduce an important class of estimators, called locally unbiasedness for the parameters of interest. We then derive the precision limit for estimating the parameters of interest together with other equivalent expressions. (Theorem 2.2) We propose a method of how to eliminate the nuisance parameters from a given estimation error bound. The proposed method, the weight matrix elimination method, provides an upper bound for the precision limit for the parameters of interest. It can be used to derive a non-trivial error tradeoff relation between the parameters of interest and the nuisance parameters. We also discuss a possible application of our method. We compare several different classes of estimators and bounds to be used and find that the effect of nuisance parameters in general cannot be ignored for the finite sample size case. Based on our approach, we can clearly argue when we can ignore the effect of nuisance parameters and in what sense the bound is achievable.

As a specific example of the nuisance parameter problem, let us consider a familiar qubit state parametrized by the standard Stokes parameters. Suppose that one is only interested in knowing the expectation value of $\sigma_x$, but not the other two parameters. What is then the best measurement and estimator for this task? There is no physicist who objects to use the projection measurement along $x$ axis and returns the sample mean as an estimator. Why is this optimal or we ask what limits us from using other measurements to do better than this? Putting differently,
optimality is within which class of measurements and estimators? By formulating this simple problem as the nuisance parameter problem, we can give an affirmative answer to this question from the statistical point of view.

Next example is relevant to the problem in quantum metrology. A typical example is to estimate the value of phase accompanied by some unitary transformation when the system undergoes unavoidable quantum noise. This class of estimation problem is of great importance for realizing quantum metrological enhancement in laboratory where unknown noises from environment are present. In the absence of these noises, the ultimate precision limit for the one-parameter estimation metrology is set by the symmetric logarithmic derivative (SLD) quantum Fisher information, which overcomes the classical shot noise limit. Important is to investigate what happens to the precision limit when some of noise parameters are not completely known. It many previous analyses on quantum metrology with quantum noise, one calculates the SLD Fisher information about variation with respect only to the parameter of interest. Although this bound is commonly regarded as the ultimate one even in the presence of nuisance parameter(s), this is not proven when some of noise parameters are not known precisely. Over the last few years, noisy quantum metrology has been focusing just on the multi-parameter estimation aspect of the problem by estimating phase and noise parameters. In this paper, we provide a general formulation of the problem and discusses a tradeoff relation between the parameters of interest and the nuisance parameters. We explicitly show why the above naive bound cannot be attained in general. We also derive a sufficient condition that guarantees such a common approach. A realistic noise model analyzed in this paper also shows that the effects of nuisance parameters cannot be omitted.

The outline of this paper is as follows. In section 2.1, we give a summary on the multi-parameter estimation theory based on the locally unbiasedness condition. In section 2.2, we formulate the nuisance parameter problem in the quantum case, and define the most informative bound for the parameters of interest. Section 2.3 provides the main theorem proving three equivalent expressions for the most informative bound. In section 2.4, we propose a method of eliminating the nuisance parameters from a given precision bound. In section 2.5, discussions on our results are given together with a tradeoff relation in our problem setting. In section 3, we analyze the nuisance parameter problem for qubit models when the parameter of interest is scalar. Section 4, we study qubit models when two parameters are to be estimated as the parameters of interest. In section 5, two examples in open quantum systems are worked out to illustrate non-trivial effects of the nuisance parameters. The last section 6 summarizes our result and state some of open problems on this topics. Appendix A provides supplemental materials for the main text. Proofs are postponed in appendix B. Appendix C proves that the proposed method recovers the result in classical statistics.

2. Formulation of the problem

2.1. Quantum parameter estimation problem

In this subsection, we introduce notation and give a brief summary of the multi-parameter estimation problem. See textbooks [5,28–32] for more details and a review [27] in the context of the nuisance parameter problem.

A quantum system is represented by a $d$-dimensional complex vector space $\mathcal{H} = \mathbb{C}^d$. To simplify our discussion we only consider quantum systems with a fixed dimension $d < \infty$. A quantum state $\rho$ is a nonnegative matrix on $\mathcal{H}$ with unit trace. A measurement $\Pi$ is a set of nonnegative matrices $\Pi = \{\Pi_x\}_{x \in X}$ such that the condition $\sum_{x \in X} \Pi_x = I_d$ ($I_d$: $d \times d$ identity
matrix) is satisfied, where $X$ is a label set for the measurement outcomes. Hereafter, we shall only concern POVMs with finite outcomes, and the case of continuous measurement will not be discussed\(^1\). The set $\Pi$ is referred to as the positive operator-valued measure (POVM). When the POVM consists of mutually orthogonal projectors, we call it the projection valued measure (PVM) or simply the projection measurement. The probabilistic rule when a POVM $\Pi$ is performed on $\rho$ is given by

$$p_{\rho}(x|\Pi) := \text{tr}(\rho \Pi_x),$$

which defines the probability distribution of the measurement outcome $X = x$.

A quantum statistical model or simply a model is defined by a parametric family of quantum states $\rho_\theta$ on $\mathcal{H}$:

$$\mathcal{M}_n := \{\rho_\theta | \theta = (\theta_1, \theta_2, \ldots, \theta_n) \in \Theta \subset \mathbb{R}^n\}.\quad (2)$$

As in the classical statistics, we assume regularity conditions. For example, a mapping $\theta \mapsto \rho_\theta$ is one-to-one and smooth so that we can differentiate $\rho_\theta$ sufficiently many times. $\partial \rho_\theta / \partial \theta_i$ ($i = 1, 2, \ldots, n$) are linearly independent. In the quantum case, we need to draw attention to the rank of quantum states. For the sake of clarity, we only consider full-rank states in this paper.

A set of a measurement $\Pi$ and an estimator $\hat{\theta}$, $\hat{\Pi} = (\Pi, \hat{\theta})$, is called a quantum estimator or simply an estimator. We measure an estimation error by the mean square error (MSE) matrix defined by

$$V_{\theta}[\hat{\Pi}] = \sum_{x \in X} \text{tr}(\rho_\theta \Pi_x)(\hat{\theta}_i(X) - \theta_i)(\hat{\theta}_j(X) - \theta_j)$$

$$= \left[ E_{\theta} \left[ (\hat{\theta}_i(X) - \theta_i)(\hat{\theta}_j(X) - \theta_j) | \Pi \right] \right],$$

where $E_{\theta}[f(X)|\Pi]$ is the expectation value of a random variable $f(X)$ with respect to the probability distribution $p_{\rho_\theta}(x|\Pi) = \text{tr}(\rho_\theta \Pi_x)$. Note that it is in general not possible to minimize the MSE matrix over all possible measurements in the sense of a matrix inequality. This is already well-known fact in the theory of optimal design of experiments in statistics. See \cite{33–37} and \cite{38} in the context of the quantum estimation theory. One of possible approaches is to optimize the weighted trace of the MSE matrix:

$$\text{Tr} \left\{ W_n V_{\theta}[\hat{\Pi}] \right\},$$

for a given $n \times n$ positive matrix $W_n$. Here, the matrix $W_n$ is called a weight matrix. As we will discuss in this paper, a role of the weight matrix is important when discussing the nuisance parameter problem in the quantum case.

2.1.1. Most informative bound under the locally unbiasedness. The fundamental question is to find the precision limit and the lower bond for the MSE under an appropriate condition on estimators. One of the standard class of estimators within the framework of the point-estimation theory is the locally unbiasedness condition. This is because the primal objective is to find the

\(^1\) It is known that there exists a discrete POVM attaining the lower bound, which is optimal \cite{32}.\footnote{It is known that there exists a discrete POVM attaining the lower bound, which is optimal \cite{32}.}
best estimation strategy at the true yet unknown point $\theta$ [2–5]. An estimator $\hat{\Pi}$ is said \textit{locally unbiased at $\theta$}, if it satisfies

$$E_{\theta}[\hat{\theta}(X)|\Pi] = \theta_i \quad \text{and} \quad \frac{\partial}{\partial \theta_j} E_{\theta}[\hat{\theta}(X)|\Pi] = \delta_{ij},$$

at $\theta$ for all indices $i, j \in \{1, 2, \ldots, n\}$. In contrast to the unbiasedness condition, where one imposes locally unbiasedness at all points, a locally unbiased estimator always exists. Thus, we can avoid non-existence of an optimal estimator for the problem at hand.

The precision limit is the minimum of the weighted trace of the MSE matrix over all possible locally unbiased estimators:

$$C_{\theta}[W_n, M_n] := \min_{\Pi \text{ l.u. at } \theta} \text{Tr} \left\{ W_n V_{\theta}^{-1} \left[ \Pi \right] \right\},$$

where the locally unbiasedness condition is indicated by l.u. at $\theta$. In this paper, any bound for the weighted trace of the MSE matrix is referred to as the \textit{CR type bound}. When a CR type bound is tight as in (5), we call it as the \textit{most informative bound} in our discussion\(^2\). The most informative bound represents the fundamental precision limit for all quantum estimators under a specific class of estimation strategies. When considering all possible estimation strategies, we can regard it as the ultimate precision limit. Some of well-known examples of CR type and most informative bounds are: the SLD CR bound, which is most informative for a one-parameter model [28], the right logarithmic derivative (RLD) CR bound, which is most informative for a Gaussian shift model [29, 40], the Nagaoka bound, which is most informative for a two-parameter qubit model [41], and the Hayashi–Gill–Massar bound, which is most informative for three parameter qubit model [42, 43].

We show an alternative expression for the most informative bound. Nagaoka proved that the most informative bound can be expressed as

$$C_{\theta}[W_n, M_n] = \min_{\Pi \text{ POVM}} \text{Tr} \left\{ W_n J_{\theta}[\Pi]^{-1} \right\},$$

where the minimum is taken over all possible POVMs. Here, $J_{\theta}[\Pi]$ denotes the classical Fisher information matrix about the classical statistical model: $M(\Pi) := \{ p_\theta(\cdot|\Pi) \mid \theta \in \Theta \}$, which is determined by measurement outcomes $p_\theta(x|\Pi) = \text{tr}(\rho_\theta \Pi_x)$.

2.1.2. \textit{Estimation strategy}. So far we have considered the single copy setting to define the bound (5). We now list three possible estimation strategies upon dealing with the multiple-copy setting together with the meaning for achievability of the bound (5).

Suppose we are given identically and independently distributed (i.i.d.) quantum states of the sample size $N$ represented by the $N$-tensor state: $\rho_\theta^N = \rho_0 \otimes \rho_0 \otimes \cdots \otimes \rho_0$. We perform a suitable POVM and then make an estimate for the unknown parameter value $\theta$. In this multiple-copy setting, we can consider the following three cases.

(A1) Repetitive strategy (i.i.d. strategy) In this setting, we choose a POVM and repeat the same POVM $N$ times. Thus, each POVM is perform on each state without any (classical) correlation. The resulting measurement outcomes are i.i.d.

---

\(^2\) It is known that the minimum exists for a given positive-definite weight matrix [39]. When the weight matrix is positive semi-definite, on the other hand, only infimum exists in general. This problem is also known as the singular design problem in the theory of optimal design of experiments [33–38].
(A2) Adaptive strategy (sequential strategy): in the next setting, we still perform POVMs on each state individually. However, the choice of POVMs can depend on measurement outcomes of the previous POVMs. Thus, resulting measurement outcomes can be correlated classically.

(A3) Collective strategy: in the last setting, arbitrary POVM on the $N$ tensor state $\rho_\theta^{\otimes N}$ is considered. This includes a collective or joint POVM on the state utilizing quantum resource.

It is clear that these three classes exhaust all possible measurements. Further, A2 contains A1, and A3 contains A2 as special cases.

For the class of A2 estimation strategies, we have several classifications. Nagaoka proposed the adaptive state-estimation method in the context of quantum-state estimation problems [44]. Rigorous treatment of such a scheme is due to Fujiwara [39]. In this case, we perform each POVM sequentially based on the previous outcomes. Experimentally, this is demanding since we need to change measurement apparatus every time.

The other extreme case is known as the two-step estimation method [45, 46]. In this case, we divide $N$ sample states in two groups. We perform a suitable POVM on the first group individually and then we choose the next POVM based on the measurement outcomes on the first group. This method is often used as a method of proof.

Of course, there exist intermediate settings between the adaptive state-estimation method and the two-step estimation method: for example, a block-wise adaptation method, where $N$ sample states are divided into sub-blocks. We then choose a POVM for the next sample block based on the previous measurement outcomes. Other various types of adaptive schemes have been intensively studied recently both in theory and experiment. See, for example, [47–52] and a review paper [53] and references therein.

Lastly, A3 strategy (Collective one) includes all possible POVMs allowed in quantum theory. Therefore, the precision limit derived under this strategy represents an ultimate one. The fundamental precision limit in the asymptotic setting was investigated in [26]. We will not cover A3 estimation strategy in this paper. See [27] for more detail.

2.1.3. Operational meaning of the most informative bound. Let us now discuss the meaning of the most informative bound (5). One of the main objectives for the point estimation problem is to know the best estimation strategy upon inferring a given unknown parameter at each fixed point $\theta$. Hence, the optimal estimator is defined by

$$\hat{\Pi}_{\text{opt}} = \arg \min_{\Pi \; \text{at} \; \theta} \text{Tr} \left\{ W_n \rho_\theta \Pi \right\}.$$  \hspace{1cm} (7)

The operational meaning of the bound (5) is thus what one can do best for a single copy of an unknown state. When extending this argument to the multiple-copy setting, we can easily see that the strategy A1 (Repetitive one) will not change anything. In this case, the weighted trace of the MSE matrix behaves $C_0[W_n, M_n]/N$ by performing the best POVM and estimator $\hat{\Pi}_{\text{opt}}$. We now remark that this optimal estimator depends on the unknown parameter value $\theta$ in general. In the language of the optimal design of experiments, it is called a local optimal design problem [33–38]. Therefore, it is impossible to realize such a local optimal POVM without knowing the parameter to be estimated. If and only when the optimal estimator is independent of the parameter $\theta$, the bound (5) can be regarded as the achievable bound for the finite sample size. The necessary and sufficient condition for existing such $\theta$-independent POVM and estimator is known [5] and is reviewed in [27].

We next consider the strategy A2 (adaptive one). In [39, 45, 46], it was proven that the bound (5) can be achieved asymptotically. In this case, the operational meaning of the bound
is as follows. For the properly chosen sequence of estimators (≈ optimal one), the weighted trace of the MSE matrix behaves approximately for the sufficiently large $N$ case as

$$\text{Tr} \left\{ W_n V^\theta \left[ \hat{\Pi} \right] \right\} \simeq \frac{1}{N} C^\theta [W_n, \mathcal{M}_n]. \quad (8)$$

As noted above, utilization of adaptive strategies is more general than the repetitive strategy with fixed a POVM (strategy A1). However, it is known that adaptive strategies cannot go beyond the precision limit for the locally unbiased estimators within strategy A1. Therefore, $C^\theta [W_n, \mathcal{M}_n]$ represents the best decaying coefficient for the MSE matrix. (First order asymptotics) To conclude, the above precision limit (5) is the ultimate limit for any separable POVMs. See [27] for more detailed discussion on this point.

Last, let us consider the strategy A3 (collective one). In this case, optimization for the finite sample size case gets harder in general. In the asymptotic theory, we can no longer rely on the locally unbiased condition. But, one needs to impose other conditions depending on the problems at hand. Some of familiar ones are the asymptotically unbiasedness or asymptotic covariance. Under additional (mathematical) assumptions, it was shown that the optimal decay coefficient for the weighted trace of the MSE matrix is equal to the Holevo bound. In this paper, we will not discuss on this asymptotic setting for A3 strategy. See [27] for more detailed discussion on this point.

### 2.2. Quantum nuisance parameter problem

We now introduce a quantum statistical model with nuisance parameters. Consider an $n$-parameter model and divide the parameters into two groups, one consists of parameters of interest $\theta_i = (\theta_1, \theta_2, \ldots, \theta_m)$ and the other subset is the nuisance parameters $\theta_N = (\theta_{m+1}, \theta_{m+2}, \ldots, \theta_n)$. Thus, the family of quantum states are parametrized by two different kinds of parameters:

$$\mathcal{M}_n = \{ \rho_\theta \mid \theta = (\theta_i, \theta_N) \in \Theta \}. \quad (9)$$

Our goal is to perform a good measurement to extract as much information possible for the parameters of interest $\theta_i$ and then to infer these values. Let $\hat{\Pi}_I = (\Pi, \theta_i)$ be an estimator for the parameters of interest and define its MSE matrix for the parameters of interest by

$$V^\theta_I[\hat{\Pi}_I] = \sum_{x \in \mathcal{X}} \text{tr}(\rho_\theta \Pi(x) \hat{\theta}_i(x) \hat{\theta}_j(x) - \theta_i \theta_j)$$

$$= E^\theta \left[ \left( \hat{\theta}_i(X) - \theta_i \right) \left( \hat{\theta}_j(X) - \theta_j \right) \right], \quad (10)$$

where the matrix indices $(i,j)$ run from 1 to $m$, but not from 1 to $n$. Hence, the MSE matrix is an $m \times m$ matrix. Our concern is to find the precision limit for this MSE matrix under the locally unbiasedness condition.

To deal with the nuisance parameter problem, we need to impose a certain condition for estimators. Following the point-estimation problem setting in section 2.1, we should look for locally unbiased estimators for the parameters of interest. Therefore, a natural question is how to define the concept of locally unbiasedness for the parameters of interest. First, we consider the unbiasedness condition for the parameters of interest. If an estimator $\hat{\Pi}_I$ for the parameters of interest satisfies the condition, $E^\theta [\hat{\theta}_i(X) | \Pi] = \theta_i$, for all $i = 1, 2, \ldots, m$ and for all $\theta \in \Theta$, it is called unbiased for the parameters of interest.
The concept of locally unbiasedness for the parameters of interest is introduced as follows.

**Definition.** If an estimator \( \hat{\Pi} \) for the parameters of interest satisfies

\[
E_\theta[\hat{\theta}(X)|\Pi] = \theta_i \quad \text{and} \quad \frac{\partial}{\partial \theta_j} E_\theta[\hat{\theta}(X)|\Pi] = \delta_{ij}, \tag{11}
\]

for \( \forall i \in \{1, \ldots, m\} \) and \( \forall j \in \{1, \ldots, n\} \) at a given point \( \theta \), it is said locally unbiased for \( \theta \) at \( \theta \).

An important condition here is the requirement: \( \frac{\partial}{\partial \theta_j} E_\theta[\hat{\theta}(X)|\Pi] = 0 \) for \( i = 1, 2, \ldots, m \) and \( j = m + 1, m + 2, \ldots, n \). This additional condition comes from the partial derivatives of the unbiasedness condition with respect to the nuisance parameters. This is trivially satisfied if a probability distribution from a POVM is independent of the nuisance parameters. But this can only happen in special cases. In general, non-vanishing of \( \frac{\partial}{\partial \theta_j} E_\theta[\hat{\theta}(X)|\Pi] \) for \( j = m + 1, m + 2, \ldots, n \) contributes the estimation error bound for the parameters of interest.

Note that the following lemma shows the above definition does not depend on reparametrization of the nuisance parameters. Its proof is given in appendix B.1.

**Lemma 2.1.** If an estimator \( (\Pi, \hat{\Pi}) \) is locally unbiased for \( \theta \) at \( \theta \), then it is also locally unbiased for any parametrization defined by the transformation of the form\(^3\):

\[
\theta = (\theta_1, \theta_N) \mapsto \xi = (\xi_1, \xi_N) \text{ s.t. } \xi_l = \theta_l, \tag{12}
\]

*That is, if two conditions (11) are satisfied, then the following conditions also hold.

\[
E_\xi[\hat{\theta}(X)|\Pi] = \xi_i \quad \text{and} \quad \frac{\partial}{\partial \xi_j} E_\xi[\hat{\theta}(X)|\Pi] = \delta_{ij}, \tag{13}
\]

for \( \forall i \in \{1, \ldots, m\} \) and \( \forall j \in \{1, \ldots, n\} \).

Having introduced the locally unbiasedness condition for the parameters of interest, we now define the most informative bound for the parameters of interest by the following optimization:

**Definition.** For a given \( m \times m \) weight matrix \( W_i > 0 \), the most informative bound about the parameters of interest is defined by

\[
C_{\theta_l}[W_i, \mathcal{M}_l] := \min_{\hat{\Pi}_{l, 1,u} \text{ for } \theta_l} \text{Tr} \left\{ W_i V_{\theta_l}[\hat{\Pi}_{l, 1,u}] \right\}, \tag{14}
\]

where the condition for the minimization is such that estimators \( \hat{\Pi}_{l, 1,u} \) are locally unbiased for \( \theta_l \) at \( \theta^4 \).

The operational meaning of the bound (14) is as follows. It defines the best estimator \( \hat{\Pi}_{l, \text{opt}} = \arg \min_{\Pi_{l, 1,u} \text{ for } \theta_l} \text{Tr} \left\{ W_i V_{\theta_l}[\hat{\Pi}_{l, 1,u}] \right\} \) and the fundamental precision limit at each fixed point \( \theta \). The class of estimators in this framework is the locally unbiasedness for the parameters of interest. Following the same argument in section 2.1.3, we can identify this bound as the ultimate one under any separable estimation strategy: A2 (adaptive one). Further, its achievability can be

\(^3\)The condition \( \xi_l = (\xi_1, \xi_2, \ldots, \xi_n) = \theta_l \) ensures that the parameters of interest are unchanged while the nuisance parameters can be changed arbitrary.

\(^4\)Note that this bound depends on \( \theta \) in general. The subscript \( \theta_l \) has a symbolic meaning only rather than its dependence on the parameter. The same remark applies to the rest of the paper.
proven in the strategy A2 for the asymptotic sample size limit. One of the main purposes of this paper is to show other possible characterizations of this bound (14). An important question here is how to rewrite this bound in a more ready to use form. This will be the subject of the next section.

By extending the argument of deriving the relationship (6) and the classical CR inequality (A.3) in the presence of nuisance parameter, we can show that the following alternative expression holds (see appendix B.2 for its derivation).

\[
C_{\theta}[W, M] = \min_{\Pi\text{ POVM}} \text{Tr} \left\{ W_{\theta} J_{\theta}^{-1} \right\}.
\]

(15)

In this alternative expression, we use the block matrix representation of the Fisher information matrix about a statistical model \(M(\Pi)\) for given a POVM \(\Pi\). Let \(J_{\theta}\) be the classical Fisher information matrix and denote its inverse by \(J_{\theta}^{-1}\). The block matrix representations of the Fisher information matrix according to the partition \(\theta = (\theta_I, \theta_N)\) are defined by

\[
J_{\theta} = \begin{pmatrix}
J_{\theta I} & J_{\theta I \theta_N} \\
J_{\theta I \theta_N} & J_{\theta N}
\end{pmatrix}, \quad J_{\theta}^{-1} = \begin{pmatrix}
J_{\theta I}^{-1} & -J_{\theta I}^{-1} J_{\theta I \theta_N} J_{\theta N}^{-1} \\
-J_{\theta N}^{-1} J_{\theta I \theta_N} J_{\theta I}^{-1} & J_{\theta N}^{-1}
\end{pmatrix}.
\]

(16)

See appendix A.1 for more detail.

2.3. Class of estimators and equivalent representation of the precision limit

For our discussion, it is convenient to introduce the following notation and definitions. For an \(n\)-parameter model, we also consider an estimator for the nuisance parameter \(\theta_N\) and denote it by \(\hat{\theta}_N\). We then have a quantum estimator for all parameters as \(\hat{\Pi} = (\Pi, \hat{\theta}_I, \hat{\theta}_N)\). Let \(V_\theta[\hat{\Pi}]\) be the \(n \times n\) MSE matrix about the estimator \(\hat{\Pi}\), and introduce the following block matrix decomposition:

\[
V_\theta[\hat{\Pi}] = \begin{pmatrix}
V_{\theta I}[\hat{\Pi}] & V_{\theta I \theta_N}[\hat{\Pi}] \\
V_{\theta I \theta_N}[\hat{\Pi}] & V_{\theta N}[\hat{\Pi}]
\end{pmatrix},
\]

(17)

which is same as in the classical case (A.8). Likewise, for the weight matrix \(W_n\), consider the same decomposition:

\[
W_n = \begin{pmatrix}
W_I & W_{IN} \\
W_{NI} & W_N
\end{pmatrix}.
\]

(18)

Let us compare two important classes of quantum estimators in our discussions. Denote by \(\mathcal{E}_\theta\) a set of all estimators that are locally unbiased about \(\theta = (\theta_I, \theta_N)\) at \(\theta\) and by \(\mathcal{E}_{\theta I}\), locally unbiased estimators about \(\theta_I\) at \(\theta\):

\[
\mathcal{E}_\theta := \{ \hat{\Pi} | \hat{\Pi} \text{ is locally unbiased about } \theta \},
\]

\[
\mathcal{E}_{\theta I} := \{ \hat{\Pi} | \hat{\Pi} \text{ is locally unbiased about } \theta_I \}.
\]

(19)

Obviously, the set \(\mathcal{E}_\theta\) is a subset of \(\mathcal{E}_{\theta I}\), since we impose additional conditions for \(\hat{\Pi}\), that is \(\mathcal{E}_\theta \subset \mathcal{E}_{\theta I}\) holds. We now show two alternative estimation error bounds and discuss their relations to the most informative bound for the parameters of interest (14).

Method 1. (Locally unbiased estimation method)

Having in mind that optimal POVMs depend on the nuisance parameter \(\theta_N\) in general, we may look for a restricted class of estimators. For an estimator \(\hat{\Pi} = (\Pi, \hat{\theta}_I, \hat{\theta}_N)\) about \(\theta\), we restrict an estimator to consider \(\hat{\Pi}_I = (\Pi, \hat{\theta}_I)\) for the parameters of interest. We impose the
locally unbiasedness condition on both parameters $\theta = (\theta_I, \theta_N)$ and consider the optimization problem:

$$C_{\theta I}^L [W_I, \mathcal{M}_n] = \inf_{\Pi \in \mathcal{C}_\theta} \text{Tr} \left\{ W_I V_{\theta I} [\hat{\Pi}] \right\}.$$  

(20)

The difference from (14) is that estimators here need to satisfy the additional locally unbiasedness condition$^5$.

Method 2. (The weight-matrix limit method)

We consider an $n$-parameter model $\mathcal{M}_n$ including nuisance parameter. Let $C_\theta [W_n, \mathcal{M}_n]$ be a CR type bound, and we take the following limit for the weight matrix $W_n$: letting the $n \times n$ weight matrix to the $m \times m$ matrix by

$$W_n \rightarrow \lim_{\epsilon \rightarrow 0^+} \begin{pmatrix} W_I & 0 \\ 0 & \epsilon I_N \end{pmatrix} = \begin{pmatrix} W_I & 0 \\ 0 & 0 \end{pmatrix},$$  

(21)

where $I_N$ denotes the $(n - m) \times (n - m)$ identity matrix. There exist several other limiting procedures, but we only consider the above case in this paper. In this way, we obtain the relevant bound:

$$\lim_{W_n \rightarrow W_I} C_{\theta I} [W_n, \mathcal{M}_n].$$

Here the matrix limit indicates procedure (21). When we start with the most informative bound for $\theta = (\theta_I, \theta_N)$, we define

$$C_{\theta I}^{\text{lim}} [W_I, \mathcal{M}_n] = \lim_{W_n \rightarrow W_I} \min_{\Pi \in \mathcal{C}_\theta} \text{Tr} \left\{ W_n V_{\theta I} [\hat{\Pi}] \right\}.$$  

(22)

We now discuss relations among three precision bounds defined so far: $C_{\theta I} [W_I, \mathcal{M}_n]$, $C_{\theta I}^{E} [W_I, \mathcal{M}_n]$, $C_{\theta I}^{\text{lim}} [W_I, \mathcal{M}_n]$. Note that three bounds are defined based on different classes of estimators. The following theorem proves that they are all equivalent. The proof is given in appendix B.3.

Theorem 2.2. For our setting, $C_{\theta I} [W_I, \mathcal{M}_n] = C_{\theta I}^{E} [W_I, \mathcal{M}_n] = C_{\theta I}^{\text{lim}} [W_I, \mathcal{M}_n]$ holds for all $W_I > 0$.

Let us now discuss consequences of theorem 2.2 and operational meanings of these bounds. First, theorem 2.2 establishes that two classes of estimators are essentially same as long as separable measurements are concerned. Therefore, we have $C_{\theta I} [W_I, \mathcal{M}_n] = C_{\theta I}^{E} [W_I, \mathcal{M}_n]$. Furthermore, this bound can be easily evaluated from the precision limit for all parameters simply by taking the weight-matrix limit $C_{\theta I}^{\text{lim}} [W_I, \mathcal{M}_n]$.

Second, let us discuss achievability of $C_{\theta I} [W_I, \mathcal{M}_n]$. This bound represents what one could do best to estimate the parameters of interest $\theta_I$ at each fixed point $\theta$. Its achievability can be proven within the estimation strategy A2 (adaptive one). If the optimal POVM is independent of the nuisance parameters $\theta_N$ (case 2 and case 4 in table 1 below), then every step of proof goes same as the usual multi-parameter estimation problem. Properly designed two-step methods and the adaptive state-estimation method can be used to attain the bound. On the other hand, if the optimal POVM depends on the nuisance parameters (case 1 and case 3 in table 1),

$^5$ Upon imposing locally unbiasedness for all parameters, we may not be able to find an optimal estimator. However, we can always find an optimal locally unbiased estimator, which is $\epsilon$ close to any small $\epsilon$. This is the reason why the bound is given by the infimum.
Table 1. Summary of the nuisance parameter problem based on $\theta = (\theta_I, \theta_N)$ dependence ($\theta$ dep.) on the optimal POVM, and achievability of the most informative bound for the parameters of interest (14). Estimation strategy A1 (repetitive strategy) and A2 (adaptive strategy) are explained in section 2.1.2. Estimation strategy $A2^\prime$ represents a restricted class of adaptive strategies described in the second remark about theorem 2.2.

| Case | $\theta_I$ Dep. | $\theta_N$ Dep. | Tradeoff relation | Optimal strategy |
|------|----------------|----------------|-----------------|-----------------|
| 1    | Yes            | Yes            | Important       | $A2^\prime$     |
| 2    | Yes            | No             | Not important   | $A2$            |
| 3    | No             | Yes            | Important       | $A2^\prime$     |
| 4    | No             | No             | Not important   | $A1$            |

one has to design adaptive strategies more carefully. Not only estimating the values of $\theta_I$, but knowing the values of $\theta_N$ is also important in this case. We remark that some of the adaptive state-estimation methods fail to converge in this case. This happens when the local optimal POVM cannot extract information about the nuisance parameters. A typical method of proof for achievability of this bound is by the two-step method: first to estimate all parameters by a locally unbiased estimator for the first set of sample. Then, we perform an almost optimal estimator for the second group. In this way, we can achieve the bound $C_{\theta_I}[W_I, M_n]$ in the infinite sample size limit. Since this achievability is in the asymptotic limit, any small estimation error for the nuisance parameters does not matter in the end. In the following, estimation strategy $A2^\prime$ denotes this class of adaptive estimation strategy with additional constraints in the presence of the nuisance parameters.

Third, the operational meanings of two bounds $C_{\theta_I}[W_I, M_n]$, $C_{\theta_I}^E[W_I, M_n]$ are as follows. The class of estimators $E_{\theta_I}$, consisting of all possible locally unbiased ones for all parameters, is a suitable class, when estimating all parameters $\theta = (\theta_I, \theta_N)$. The bound $C_{\theta_I}^E[W_I, M_n]$ hence represents what one could do best by knowing all the parameters. The other class of estimators $E_{\theta_N}$ is wider than $E_{\theta_I}$, since we only impose the locally unbiasedness condition for the parameters of interest. As long as the point-estimation problem about the parameters of interest is concerned, this is the only requirement we should impose. This is because we do not need to estimate the nuisance parameters. Thus, the bound $C_{\theta_I}[W_I, M_n]$ gives the ideal precision limit for our problem. One may naively expect that the precision limit can be lower by estimating $\theta_I$ only. However, these two precision limits are exactly same from theorem 2.2. This is an important consequence of our paper.

Fourth, an advantage of the weight-matrix limit $C_{\theta_I}^{\text{lim}}[W_I, M_n]$ is obvious, since we can easily evaluate it. However, we need to be careful about the matrix limit. Beside some mathematical subtlety in matrix limits, we still have a question of achievability for a such derived bound in the limit (21). This is because it is not obvious whether the limit of optimal estimators exits or not. (In the worst case, an optimal POVM could diverge in the weight-matrix limit.) That is to say, the limit (21) may results in a meaningless measurement. However, theorem 2.2 states that the limit of the bound is well-defined quantity and has the operational meaning discussed above.

Last, tradeoff relations among estimation errors. One of our motivation in this work is to discuss tradeoff relations for estimation errors between the parameters of interest and the nuisance parameter. When dealing with the finite sample size case, it is extremely important how to design measurements to acquire as much information about the parameters of interest as possible. In case 2 and case 4 in table 1, it is clear that we do not need to estimate the nuisance parameters at all. (In fact, case 4 is a trivial one in the sense that the best is to repeat the optimal POVM.) Thus, such designing the two-step method and the adaptive state-estimation method.
can be handled with the existing methods studied in literatures. On the other hand, we need to properly design POVMs for the other two cases (case 1 and case 3 in table 1). In these cases, estimation errors for the nuisance parameters do matter for the finite sample size case, and a tradeoff relation between two sets of parameters should play an important role in our discussion. Note that case 3 is an extreme one, since an optimal POVM only depends on the nuisance parameters. In other words, we also need to perform a good measurement for extracting as much information about the nuisance parameters as possible even though we are not interested in them.

From the above discussion so far, we will not be able to discuss such tradeoff relations within the framework of the point-estimation problem under the locally unbiasedness condition. This is mainly due to the following two reasons. One is that we are looking at an optimal estimation method, which is defined locally. The other is that its achievability is proven in the asymptotic limit. Thus, only estimation errors for the parameters of interest matter in this limit. In next section, we shall propose a method that can capture an intrinsic tradeoff relation systematically.

2.4. Elimination of nuisance parameters in quantum-state estimation problems

2.4.1. Formulation of the problem. In this section, we propose a method of deriving an upper bound for the precision limit for the parameters of interest. Thereby, we discuss a tradeoff relation between the parameters of interest and the nuisance parameters. In literature, there exist various forms of tradeoff relations between MSE components were reported. See, for example, [19, 21, 29, 54–56] and references therein. Normally, these results were discussed within the context of the uncertainty relation in quantum estimation theory. The main difference from existing results is that we wish to derive a form of precision (upper) bound for the parameters of interest given a certain MSE for the nuisance parameters.

As stated in the previous section, the locally unbiasedness condition for the parameters of interest cannot include an intrinsic tradeoff relation between the parameters of interest and the nuisance parameters. To circumvent this problem, we derive an upper bound for the precision limit $C_{\theta}[W, M_n]$ (definition in section 2.2), which is equal to other two forms $C^{opt}_{\theta}[W, M_n]$ and $C^{lim}_{\theta}[W, M_n]$ (theorem 2.2). This method, called the weight-matrix elimination method, is different from the classical case, since our figure of merit is the weighted trace of the MSE matrix. Yet, as we shall demonstrate later (proposition 2.3), this method is a generalization of the classical elimination method in the sense that we can also recover the same result as the classical case.

Given an $n$-parameter model $M_n$, we consider an estimation problem for all parameters under the locally unbiasedness condition for all parameters $\theta = (\theta_I, \theta_N)$. Let $\tilde{\Pi} = (\Pi, \theta)$ be an estimator for $\theta = (\theta_I, \theta_N)$ and denote its MSE matrix $V_{\theta}[\tilde{\Pi}]$. Suppose its most informative bound $C^{0}_{\theta}[W_n]$ for the weighted trace of the MSE matrix is given and consider the following inequality:

$$\text{Tr} \left\{ W_{\theta} V_{\theta}[\tilde{\Pi}] \right\} \geq C^{0}_{\theta}[W_n], \quad (23)$$

where we drop the model dependence $M_n$ in the bounds to simplify notation. We split the MSE and the weight matrix according to parameter partition $\theta = (\theta_I, \theta_N)$ as before:

$$V_{\theta}[\tilde{\Pi}] = \begin{pmatrix} V_{\theta_I} & V_{\theta_I \theta_N} \\ V_{\theta_I \theta_N} & V_{\theta_N} \end{pmatrix}, \quad W_{\theta} = \begin{pmatrix} W_I & W_{IN} \\ W_{NI} & W_N \end{pmatrix}. \quad (24)$$

Here the off-diagonal block matrices satisfy $W_{NI} = W_{IN}$ and $V_{\theta_I \theta_N} = V_{\theta_I \theta_N}$. Rewrite inequality (23) as
2.4.2. Alternative evaluation of the upper bound

J. Phys. A: Math. Theor. 53 (2020) 264001 J Suzuki

Since this holds for all weight matrices $W_n$, which is positive-definite, we can evaluate the maximum of the right-hand side to define

$$C_{\theta_N}^{\text{Nui}}[W_I, V(\theta_N)] := \max_{W_n, V_N>0} \left\{ C_{\theta_N}^{(0)}[W_n] - \text{Tr} \left\{ W_n V_{\theta_N} \right\} - 2\text{Tr} \left\{ W_N V_{\theta_I}/\theta_I \right\} \right\}.$$

(25)

where $V(\theta_N)$ represents $V_{\theta_I}/\theta_I = (V_{\theta_I}/\theta_I)^T$ and $V_{\theta_N}$. Thus, we obtain the following bound for the weighted trace of the MSE about the parameters of interest as

$$\text{Tr} \left\{ W_I V_{\theta_I} \right\} \geq C_{\theta_N}^{\text{Nui}}[W_I, V(\theta_N)].$$

(26)

Importantly, this inequality is derived under the condition of locally unbiasedness for all parameters. A natural interpretation of this result is that the MSE matrix must satisfy this inequality irrespective of the choice of weight matrix for the nuisance parameters.

2.4.2. Alternative evaluation of the upper bound (25). We have an important remark regarding the positivity condition for the weight matrix. As we emphasize in section 2.1, we impose the condition $W_n > 0$ in the maximization problem (25). It happens in some case that a maximum does not exists and we need to evaluate the supremum. If this is the case, we say that the bound is not explicitly achievable. To avoid nonexistence of a maximum, we can then relax the positivity condition as follows.

$$\tilde{C}_{\theta_N}^{\text{Nui}}[W_I, V(\theta_N)] := \max_{W_n, V_N>0: W_n>0, W_I>0} \left\{ C_{\theta_N}^{(0)}[W_n] - \text{Tr} \left\{ W_n V_{\theta_N} \right\} - 2\text{Tr} \left\{ W_N V_{\theta_I}/\theta_I \right\} \right\}.$$

(27)

Here the positivity condition applies only for the weight matrix for the parameters of interest. By definition, $C_{\theta_N}^{\text{Nui}}[W_I, V(\theta_N)] \geq \tilde{C}_{\theta_N}^{\text{Nui}}[W_I, V(\theta_N)]$ holds, since the latter optimization is carried out under less constraints.

We note that the maximization to get the bound $C_{\theta_N}^{\text{Nui}}[W_I, V(\theta_N)]$ seems hard in general and we propose the following bound. Instead of maximizing $W_N$ and $W_I$ under the matrix condition $W_n > 0$, we set the weight matrix as the block diagonal matrix by imposing $W_N = 0$. This is to look for the maximum within a restricted class of weight matrices. Therefore, we can bound the maximization from below as

$$C_{\theta_N}^{\text{Nui}}[W_I, V(\theta_N)] \geq \tilde{C}_{\theta_N}^{\text{Nui}}[W_I, V_{\theta_N}] := \max_{W_N>0} \left\{ C_{\theta_N}^{(0)}[W_n = W_I \oplus W_N] - \text{Tr} \left\{ W_N V_{\theta_N} \right\} \right\}.$$

(28)

Summarizing the above result, bound (25) (or the approximated one (28)) represents what we can do best upon estimating the parameters of interest $\theta_I = (\theta_1, \ldots, \theta_m)$ in the presence of nuisance parameter $\theta_N = (\theta_{m+1}, \ldots, \theta_n)$. The proposed method of eliminating the nuisance parameters is thus based on a similar philosophy to derive the bound (A.3) in the classical case. It is straightforward to observe that the above optimization (25) is equivalently expressed as the alternative one:

$$\min_{W_N, W_I} \left\{ \text{Tr} \left\{ W_n V_{\theta_N} \right\} - C_{\theta_N}^{(0)}[W_n] \right\}.$$

(29)
2.4.3. Discussions on the upper bound (25). Below we list several remarks concerning the above bound. First of all, as $C^\text{opt}_{\theta_1} [W_n]$ is the most informative bound upon estimating all parameters, we can easily derive the relation:

$$C^\text{Nu}_{\theta_1} [W_1, V(\theta_N)] \geq C_{\theta_1} [W_1, \mathcal{M}_a].$$

(30)

Therefore, the propose quantity is an upper bound for the precision limit for the parameters of interest. This inequality follows from i) $C^\text{Nu}_{\theta_1} [W_1, V(\theta_N)] \geq C^\text{opt}_{\theta_1} [W_1, V(\theta_N)]$, ii) $C^\text{Nu}_{\theta_1} [W_1, V(\theta_N)] \geq C^\text{opt}_{\theta_1} [W_1, \mathcal{M}_a]$ and iii) $C^\text{opt}_{\theta_1} [W_1, \mathcal{M}_a] = C_{\theta_1} [W_1, \mathcal{M}_a]$ (theorem 2.2).

Second, the proposed bound (25) depends on the estimation errors concerning the nuisance parameter $V_{\theta_1 \theta_N}$ and $V_{\theta_N}$. This is in contrast to the classical case where the bound in the presence of nuisance parameter is solely expressed in terms of the given model, i.e., the partial Fisher information. One reason for this dependence is that nuisance parameters enter in POVMs as described before. There is a tradeoff relation for the estimation errors between $\theta_1$ and $\theta_N$ in general. Therefore, bound (25) includes this kind of tradeoff. Alternative interpretation is to regard the dependence of $V_{\theta_1 \theta_N}$ and $V_{\theta_N}$ as an unavoidable ‘biased effect’. It is well known in classical statistics that the generalized CR bound for biased estimators does depend on estimation errors in general. For example, consider a one-parameter problem, the generalized CR inequality with

$$V_a[\hat{\theta}] \geq \left( \frac{\partial}{\partial \theta} E_a[\hat{\theta}] \right)^2 J^{-1}_a + (E_a[\hat{\theta}] - \theta)^2.$$

Clearly, the right-hand side contains terms depending on the performance of an estimator $\hat{\theta}$.

Third, it is important to compare the approximated bound $C^\text{Nu}_{\theta_1} [W_1, V_{\theta_1}]$ and to the achievable bound for the parameters of interest $C_{\theta_1} [W_1, \mathcal{M}_a]$. If $C^\text{Nu}_{\theta_1} [W_1, V_{\theta_1}] \geq C_{\theta_1} [W_1, \mathcal{M}_a]$ does not hold, then used approximation is not accurate enough to capture the tradeoff relation.

Fourth, we can also extend the proposed method for any CR type bound. If we start with a bound that is not achievable, the obtained bound according to the above procedure is also not achievable in general. For example, let us consider an $n$-parameter model with the parameter partition $\theta = (\theta_1, \theta_N)$ as before. The SLD CR bound $C^\text{opt}_{\theta} [W_n] := \text{Tr} \left\{ W_n (J^S)^{-1} \right\}$ gives a CR type bound:

$$\text{Tr} \left\{ W_n V_{\theta}[\Pi] \right\} \geq C^\text{opt}_{\theta} [W_n],$$

(31)

where $J^S$ is the SLD Fisher information matrix, which is defined by (A.14) in appendix A.2. It is known that this bound is not achievable unless all SLD operators commute with each other. Let us evaluate the approximated bound (28) to see that our proposal reduces to a simple result. By splitting the inverse of the SLD Fisher information into the block matrix analogously to the classical case (16),

$$J^S = \begin{pmatrix} J^S_{\theta_1 \theta_1} & J^S_{\theta_1 \theta_N} \\ J^S_{\theta_N \theta_1} & J^S_{\theta_N \theta_N} \end{pmatrix}, \quad (J^S)^{-1} = \begin{pmatrix} J^S_{\theta_1} & J^S_{\theta_1 \theta_N} \\ J^S_{\theta_N \theta_1} & J^S_{\theta_N \theta_N} \end{pmatrix},$$

(32)

we get

$$\tilde{C}^\text{Nu}_{\theta_1} [W_1, V_{\theta_1}] = \sup_{W_N > 0} \left\{ \text{Tr} \left\{ W_1 J^S_{\theta_1 \theta_1} \right\} + \text{Tr} \left\{ W_N (J^S_{\theta_1 \theta_N}) - V_{\theta_N} \right\} \right\}$$

$$= \text{Tr} \left\{ W_1 J^S_{\theta_1 \theta_1} \right\}.$$
Here we use the fact that \( \inf_{W>0} \text{Tr}(W^A) = 0 \) for a given positive semi-definite matrix \( A \) and \( V\theta_N - J_S^{\theta\theta} \geq 0 \). Thus, we obtain the same expression as the classical case by replacing the Fisher information matrix by the SLD Fisher information matrix. In fact, we can work out the maximization without involving the block diagonalization assumption of the weight matrix, and then we can show that the maximum exists.

**Proposition 2.3.** For a CR type bound of the form \( C_\theta[0][W_n] = \text{Tr}\{W_nJ^{-1}_\theta\} \), the elimination of the weight matrix for the nuisance parameters (27) yields:

\[
\overline{C}_{\theta I}^{\text{nu}}[W_n, V(\theta_N)] = \text{Tr}\{W_nJ^{\theta\theta}\},
\]

where \((J^{\theta\theta})^{-1} = J_{\theta_1\theta_1} - J_{\theta_1\theta_2}J_{\theta_2\theta_2}^{-1}J_{\theta_2\theta_1} =: J_{\theta_1|\theta_2N}\) can be regarded as the partial SLD quantum Fisher information matrix.

The proof of this proposition is given in appendix C. We refer [27] for more detail discussion on the properties of the partial quantum Fisher information. From this proposition, we see that our elimination method also recovers the classical result.

**2.4.4. Utility of the upper bound (25).** The fundamental precision limit for estimating the parameters of interest is given by (14). The propose method of deriving the upper bound (25) at first sight does not provide any practical use of it. This is true in the infinite sample size limit where one can suppress the effects of the nuisance parameters. This upper bound is also of no use when there is no tradeoff relation to be discussed (case 2 and case 4 in table 1). For the finite sample size case, on the other hand, one faces a problem of designing optimal POVMs such that we can extract as much as information about the parameters of interest for a given limited knowledge about the nuisance parameters. This is the subject of the optimal design of experiments [33–38]. In this case, we also have to concern the estimation error for the nuisance parameters.

To provide essential part of the problem, we consider the following optimization. For a given estimation error for the nuisance parameters, what is the lowest estimation error for the parameters of interest. Mathematically, this is expressed as

\[
C_\theta[W_n, M_n] := \min_{\beta : \text{I.a.t. } \theta} \text{Tr}\left\{W_nV\theta_n[\hat{\Pi}_n]\right\},
\]

where the minimization is done for the locally unbiased estimator for all parameters in the presence of an additional constraint \( \text{Tr}\{W_nV\theta_n\} \leq \epsilon \) for a given \( W_n > 0, \epsilon > 0 \). This kind of optimization is a typical problem in theory of multi-objective decision making, and are called \( \epsilon \)-constraint problem\(^6\) [57, 58]. It is clear that this optimization can be solved numerically in general. The proposed method of eliminating the nuisance parameters (25) now can give a lower bound for the above \( \epsilon \)-constraint problem. In some special case, our method provides the exact answer to this problem. All these practical applications deserve further studies and shall be presented in the future publication.

**2.5. Information loss and tradeoff relation**

We now discuss information loss and tradeoff relation based on our results. First, let us define the information loss due to the presence of nuisance parameters for the quantum case. Consider

\(^6\) More generally, we impose different values of \( \epsilon \) constraint for the MSE about the nuisance parameters.
an $m$-parameter model $\mathcal{M}_m$ where all nuisance parameters are known. Assume that we have a bound $C_{\theta}[W_I, \mathcal{M}_m]$ for this model, then the difference
\[ \Delta C_{\theta}[W_I] = C_{\theta}[W_I, \mathcal{M}_m] - C_{\theta}[W_I, \mathcal{M}_m], \] (36)
measures how much information we lose by not knowing the nuisance parameter. Owing to theorem 2.2, the information loss can be easily evaluated from two bounds by taking the weight-matrix limit of $C_{\theta}[W_n, \mathcal{M}_n]$ ($n$-parameter model) and $C_{\theta}[W_I, \mathcal{M}_I]$ ($m$-parameter model).

Unlike the classical case, it is not obvious to derive the condition of $\Delta C_{\theta}[W_I, \theta_N] = 0$ in terms of a given model and weight matrix $W_I$. Another difference is that the orthogonal condition does not provide a direct consequence for the zero loss of information. One obvious reason is that the concept of orthogonality depends on the choice of the Fisher information matrix and hence it is not unique in the quantum case [31]. The other is that a precision limit is not in general expressed as a simple closed-form in terms of the quantum Fisher information matrices. These will be discussed in the subsequent sections.

We next discuss a tradeoff relation between the parameters of interest and the nuisance parameters based on the proposed upper bound (25). In our setting, we do not need to know the value of $\theta_N$, and we are only interested in minimizing the MSE about the parameters of interest $\theta_I$. However, an optimal measurement to extract information about $\theta_I$ requires exact knowledge of $\theta_N$ in general. Therefore, we face a problem of designing experiments in such a way that we need to extract both $\theta_I$ and $\theta_N$.

We remind that this kind of tradeoff relation will not be present in the precision limit. In section 2.3, it was proven that the fundamental precision limit in the presence of nuisance parameters can be evaluated simply by taking the weight-matrix limit from the precision limit for estimating all parameters. Furthermore, this bound can be achieved at each point by separable POVMs, which are locally unbiased for the parameters of interest. A possible form of the tradeoff relation can then be defined by the difference between the upper bound $C_{\theta_I}^{\text{Trafo}}[W_I, V(\theta_N)]$ and the achievable bound $C_{\theta_I}[W_I, \mathcal{M}_N]$ as
\[ \Delta C_{\theta_I}^{\text{Trafo}}[W_I] := C_{\theta_I}^{\text{Trafo}}[W_I, V(\theta_N)] - C_{\theta_I}[W_I, \mathcal{M}_N]. \] (37)
This quantity is non-negative by definition, and zero if and only if the upper bound coincides with the most informative bound.

When the optimal POVM is independent of the nuisance parameters, this quantity does not make sense. This is because we do not need to estimate $\theta_N$, and hence we can look for an optimal estimator within the locally unbiasedness only for the parameters of interest. In this case, estimators for the nuisance parameters $\theta_N$ can be arbitrary. If the optimal POVM depends on the nuisance parameters, we need to impose the locally unbiasedness condition for the nuisance parameters to make an estimate for them. We naturally expect that the tradeoff quantity $\Delta C_{\theta_I}^{\text{Trafo}}[W_I]$ becomes smaller as the MSE for the nuisance parameters increases. This will be examined by several examples in the subsequent sections.

From proposition 2.3, we observe $\Delta C_{\theta_I}^{\text{Trafo}}[W_I] = 0$ for the nuisance parameter problem in classical statistics. Thus, we conclude that there is no tradeoff relation exists in the classical case.

3. Qubit models with nuisance parameter: scalar parameter of interest

In this section and the next section, we consider all possible mixed-state models with nuisance parameters when the dimension of the Hilbert space is two, that is, qubit models. First, note
that the possible maximum number of parameters for the qubit models is three. Otherwise, the quantum Fisher information matrix becomes singular and is not invertible. We thus have the three generic classes of possible qubit models:

1 + 1 model: \( \theta_I = \theta_1, \theta_N = \theta_2 \),
1 + 2 model: \( \theta_I = \theta_1, \theta_N = (\theta_2, \theta_3) \),
2 + 1 model: \( \theta_I = (\theta_1, \theta_2), \theta_N = \theta_3 \).

(38)

We shall call the above models as 1 + 1, 1 + 2, and 2 + 1 qubit models, respectively. Relevant most informative bounds for the qubit case are known, and they are given in terms of the SLD Fisher information matrix \( J^S_\theta \). In the following, we denote the \((i, j)\) component of \( J^S_\theta \) by \( J^S_{i,j} \). The \((i, j)\) component of the inverse of the SLD Fisher information matrix is denoted by the upper index as \( J^S_{i,j} \). In this section, we focus on the case of single parameter of interest. Namely, 1 + 1 and 1 + 2 models. 2 + 1 qubit model is discussed in the next section.

In the following, we first discuss the general formula for arbitrary qubit models, and then specific examples. In these qubit examples, we use the standard Pauli matrices \( \sigma_i (i = 1, 2, 3) \) to represent qubit states. For example, the most general qubit state in the Stokes parametrization is given by

\[
\rho_\theta = \frac{1}{2} (I + \theta_1 \sigma_1 + \theta_2 \sigma_2 + \theta_3 \sigma_3).
\]

The method of parameter orthogonalization due to Cox and Reid [7] is extended to the quantum case [27] to demonstrate its usefulness.

3.1. 1 + 1 qubit model

3.1.1. General formula. The achievable CR type bound for two-parameter qubit models is the Nagaoka bound [41]:

\[
C^N_\theta [W_2] := \text{Tr} \left\{ W_2 (J^S_\theta)^{-1} \right\} + 2 \sqrt{\text{det} W_2 (J^S_\theta)^{-1}}.
\]

(39)

This bound holds for all locally unbiased estimators \( \hat{\Pi} \) at \( \theta = (\theta_1, \theta_2) \). We stress that the Nagaoka bound is strictly larger than the SLD CR bound since the second term in (39) is positive for \( W_2 > 0 \). Nagaoka also gave an explicit construction of an optimal estimator [59].

We first examine the most informative bound for the parameter of interest. Owing to theorem 2.2, this is evaluated from the weight-matrix limit method (22). Without loss of generality, we can set the weight matrix, which is a scalar, for the parameter of interest as 1. Since \( \text{det} W_2 \) vanishes, this method yields the following bound for the parameter of interest:

\[
C_\theta [W_I = 1] = \lim_{W_2 \to (1, 0)^T(1, 0)} C^N_\theta [W_2] = (J^S_\theta)^{-1})_{11} = J^{S,11}_\theta,
\]

(40)

where we use the limit \( W_2 \to (1, 0)^T(1, 0) \). This expression is exactly same as the classical case except for having the SLD Fisher information instead of the Fisher information.

Information loss due to the nuisance parameter (36) can be evaluated as follows. Since the precision limit without the nuisance parameter is \((J^S_{\delta,11})^{-1}\), the relevant information loss is

\[
\Delta_\theta [W_I = 1] = J^{S,11}_\theta - (J^S_{\delta,11})^{-1}.
\]

(41)

17
This result immediately implies that information loss is zero if and only if the parameter of interest and the nuisance parameter are orthogonal with respect to the SLD quantum Fisher information matrix.

Next, we analyze the proposed method by eliminating the nuisance parameter from the Nagaoka bound. To perform the maximization, we parametrize the weight matrix as follows. We introduce a positive variable $\epsilon$ as follows.

$$W_2 = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} = \begin{pmatrix} w_1 \sqrt{w_1 w_2} & w_1 w_2 \epsilon \\ w_1 w_2 \epsilon & w_2 \end{pmatrix},$$

(42)

where the positivity conditions are $w_1, w_2 > 0$ and $|\epsilon| < 1$. With this form of the weight matrix, we rewrite the maximization as follows. We introduce a positive variable $\delta = (w_2/w_1)^{1/2}$, and the problem is maximization over $\delta$ and $\epsilon$. Or equivalently, we analyze $C_{\theta_1}^{\text{Nui}}[w_{11}] / w_{11}$ as follows.

$$\frac{1}{w_{11}} C_{\theta_1}^{\text{Nui}}[W_1, V(\theta_N)] = \max_{w_{12}, w_{22} > 0} \left\{ \frac{1}{w_{11}} \left( C_{\theta_1}^{\text{Nui}}[W_2] - 2 w_{12} V_{\theta_1 \theta_1} - w_{22} V_{\theta_2} \right) \right\}$$

$$= J_{\theta}^{\text{S11}} + \max_{w_2 > 0} \left\{ -(V_{\theta_1 \theta_2} - J_{\theta}^{\text{S22}}) w_2 \right\}$$

$$- 2 \sqrt{w_2} \epsilon (V_{\theta_1 \theta_2} - J_{\theta}^{\text{S12}}) + 2 \sqrt{\det(J_{\theta}^{\text{S}})^{-1}} \sqrt{w_2} \epsilon w_1 \sqrt{w_2} \epsilon w_1 \delta (\epsilon)$$

$$= J_{\theta}^{\text{S11}} + \max_{\delta > 0, |\epsilon| < 1} \left\{ -(V_{\theta_1 \theta_2} - J_{\theta}^{\text{S22}}) \delta^2 \right\}$$

$$- 2 \delta \epsilon (V_{\theta_1 \theta_2} - J_{\theta}^{\text{S12}}) + 2 \sqrt{\det(J_{\theta}^{\text{S}})^{-1}} \sqrt{1 - \epsilon^2} \sqrt{1 - \epsilon^2}$$

$$= J_{\theta}^{\text{S11}} + \max_{|\epsilon| < 1} \left\{ \sqrt{\det(J_{\theta}^{\text{S}})^{-1}} \sqrt{1 - \epsilon^2} \right\}$$

$$= J_{\theta}^{\text{S11}} + \frac{\det(J_{\theta}^{\text{S}})^{-1} + (V_{\theta_1 \theta_2} - J_{\theta}^{\text{S12}})^2}{V_{\theta_1 \theta_2} - J_{\theta}^{\text{S22}}}.$$  

(43)

The first maximization over $\delta > 0$ is simply due to a quadratic function of $\delta$ and the second one about $\epsilon$ can be done by the Cauchy–Schwarz inequality. That is $(a \sqrt{1 - \epsilon^2} + b \epsilon)^2 \leq (a^2 + b^2)(1 - \epsilon^2 + \epsilon^2) = a^2 + b^2$.

We now analyze the obtained upper bound $C_{\theta_1}^{\text{Nui}}$ and discuss a tradeoff relationship in detail. First, this is greater than the bound due to the weight-matrix limit method, since the second term in (43) is strictly positive. Second, this upper bound represents a tradeoff relation between the estimation errors about $\theta_1$ and $\theta_N$. Bounding $(V_{\theta_1 \theta_2} - J_{\theta}^{\text{S12}})^2 \geq 0$ gives

$$C_{\theta_1}^{\text{Nui}}[W_1 = 1, V(\theta_N)] \geq J_{\theta}^{\text{S11}} + \frac{\det(J_{\theta}^{\text{S}})^{-1}}{V_{\theta_1 \theta_2} - J_{\theta}^{\text{S22}}},$$

(44)

which is symmetric between $\theta_1$ and $\theta_2$. Third, this result shows that the estimation error regarding the nuisance parameter $\theta_2$ has to diverge in order to suppress the second term of (43). However, if this is the case, we completely lose information about the nuisance parameter $\theta_2$. Since the optimal POVM depends on the nuisance parameter in general, we cannot perform this optimal measurement with the finite sample size in this case. Fourth, the parameter orthogonality condition (with respect to the SLD Fisher information) alone does not play an important
role in this bound. This condition in this particular example only states that the first term in (43) is equal to $(J_{\theta}^{(1)})^{-1}$, but the second term is still finite value in general.

Last, the MSE inequality by bound (43) can be expressed after a little algebra as

$$\det (V_\theta - (J_{\theta}^{0})^{-1}) \geq \det(J_{\theta}^{0})^{-1}. \tag{45}$$

We note that this relation itself was known before as follows, see for example reference [54].

Combining the Gill–Massar inequality [43]:

$$\text{Tr} \left\{ J_{\theta}^{0} \left[ \Pi \right] (J_{\theta}^{0})^{-1} \right\} \leq 1 \text{ for all POVMs } \Pi, \tag{46}$$

and the CR inequality $V_\theta \left[ \hat{\Pi} \right] \geq (J_{\theta}(\Pi))^{-1}$ for all (locally) unbiased estimators, we get

$$\text{Tr} \left\{ (V_\theta)^{-1}(J_{\theta}^{0})^{-1} \right\} \leq 1. \tag{47}$$

When the number of parameters is two, this is equivalent to the tradeoff relationship (45). However, the Gill–Massar inequality for higher dimensional case are not tight in general, and the above derivation does not lead to achievable bounds.

We finally examine the tradeoff relation (37) due to the nuisance parameter.

$$\Delta C_{\theta}^{\text{Tradeoff}}[W_j = 1] = C_{\theta}^{\text{Stir}}[W_j = 1, V(\theta_N)] - C_{\theta}[W_j = 1, \mathcal{M}_A]$$

$$= \frac{\det(J_{\theta}^{0})^{-1} + (V_{\theta_N} - J_{\theta}^{S,12})^2}{V_{\theta_N} - J_{\theta}^{S,22}}. \tag{48}$$

We can bound this term by $(\det(J_{\theta}^{0})^{-1})/(V_{\theta_N} - J_{\theta}^{S,22})$ to conclude that the loss of information due to the nuisance parameter can be suppressed by making large errors in the nuisance parameter. However, this in turn gives less information about it, and this makes sense only when the optimal POVM is independent of $\theta_N$.

### 3.1.2. Two-parameter qubit model.

$$\mathcal{M}_A = \{ \rho_\theta = \frac{1}{2}(I + \theta_1 \sigma_1 + \theta_2 \sigma_2) \mid \theta \in \Theta \}, \tag{49}$$

where $\theta_1$ is the parameter of interest and $\theta_2$ is the nuisance parameter. The parameter region $\Theta$ is any subset of $\mathbb{R}^2 = \{ (\theta_1, \theta_2) \in \mathbb{R}^2 \mid (\theta_1)^2 + (\theta_2)^2 < 1 \}$. The inverse of the SLD Fisher information matrix for this model is

$$(J_{\theta}^{S})^{-1} = \begin{pmatrix} 1 - \theta_1^2 & -\theta_1 \theta_2 \\ -\theta_1 \theta_2 & 1 - \theta_2^2 \end{pmatrix}, \tag{50}$$

and hence $\theta_1$ and $\theta_2$ are not orthogonal in this model.

We apply the parameter orthogonalization procedure and obtain the following new parametrization:

$$\theta_1 = \xi_1, \theta_2 = c(\xi_2)\sqrt{1 - (\xi_1)^2}, \tag{51}$$

where $c(\xi_2)$ is any differentiable function, which is not constant and $\forall \xi_2, c(\xi_2) := dc(\xi_2)/d\xi_2 \neq 0$. With this new parametrization, the corresponding SLD Fisher information matrix is diagonalized as.
\[
(J^S_{\xi})^{-1} = \begin{pmatrix}
1 - (\xi_1)^2 & 0 \\
0 & 1 - \xi_1 c(\xi_2)^{-1} (\xi_1)^{-1} \xi_2 c(\xi_2)^{-1}
\end{pmatrix}. \tag{52}
\]

The SLD operator for the coordinate \(\xi_1\) now becomes
\[
L^{S}_{\xi,1} = \frac{1}{1 - (\xi_1)^2} (\xi_1 I + \sigma_1), \tag{53}
\]
which gives the optimal measurement about \(\theta_1\). This is the projection measurement about \(\sigma_1\), which is independent of the nuisance parameter. In fact, this optimal measurement is also independent of the parameter of interest \(\theta_1\). This corresponds case 4 in table 1 of section 2.3. Therefore, the projection measurement along \(x\) axis is the optimal measurement. The bound is
\[
V_{\xi,11}[\Pi] \geq 1 - (\xi_1)^2 = 1 - (\theta_1)^2, \tag{54}
\]
that is to say \(J^{S,11}_0 = 1 - (\theta_1)^2\) can be achieved explicitly by a locally unbiased estimator about the parameter of interest \(\theta_1\). It also happens that this measurement does not allow us to construct a locally unbiased estimator for both parameters \(\theta = (\theta_1, \theta_2)\). This is because the resulting probability distribution does not depend on the nuisance parameter \(\theta_2\) at all. Hence, the MSE for \(\theta_2\) formally diverges with this measurement. However, as long as we are only interested in estimating \(\theta_1\), this does not matter at all. Since the bound \(C_{\theta_1}[W]\) is achievable explicitly in this example, we conclude that the projection measurement along \(x\) axis is the optimal among all locally unbiased estimators about the parameter of interest.

3.1.3. Two-parameter submodel. We next consider a two-parameter model:
\[
\mathcal{M}_B = \{ \rho_\theta = \frac{1}{2} (I + \theta_1 \sigma_1 + \theta_2 \sigma_2 + \theta_0 \sigma_3) | \theta \in \Theta \}. \tag{55}
\]
Here, unknown parameters are \(\theta = (\theta_1, \theta_2)\) and \(\theta_0 \neq 0\) is a fixed parameter. That is to say, we have complete knowledge about \(\theta_0\) in the standard Stokes parameterization of the qubit state, see section 3.2.2 in the subsequent section. We are only interested in estimating the value of \(\theta_1\), whereas the nuisance parameter \(\theta_2\) is of no interest. What is the best estimation strategy in this case? At first thought, one may still performs a projection measurement about \(\sigma_1\). However, precise knowledge about \(\theta_3 = \theta_0\) can lower the estimation error for the parameter of interest as we see below.

The inverse of the SLD Fisher information matrix is calculated as
\[
(J^S_0)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{1 - \theta_0^2} \begin{pmatrix} \theta_1^2 & \theta_1 \theta_2 \\ \theta_1 \theta_2 & \theta_2^2 \end{pmatrix}. \tag{56}
\]
\(\theta_1\) and \(\theta_2\) are not orthogonal with respect to the SLD Fisher information matrix in this model. With the new parametrization,
\[
\theta_1 = \xi_1, \quad \theta_2 = c_2(\xi_1) \sqrt{1 - (\theta_0)^2 - (\xi_1)^2}, \tag{57}
\]
we can diagonalize the SLD Fisher information as
\[ G^{-1}_\xi = \begin{pmatrix} 1 - (\xi_1)^2 & 0 \\ 0 & 1 - c^2 \end{pmatrix} \begin{pmatrix} 1 - (\theta_0)^2 & 1 - c^2 \\ c^2(1 - (\theta_0)^2 - (\xi_1)^2) \end{pmatrix}. \] (58)

The SLD operator in this model is

\[ L_{\xi,1} = \frac{1}{1 - (\theta_0)^2 - (\xi_1)^2}(-\xi_1 I + (1 - (\theta_0)^2)\sigma_1 + \xi_1 \sigma_3). \] (59)

and this gives an optimal measurement to estimate \( \xi_1 = \theta_1 \). It depends only on the parameter \( \xi_1 \) but not on the nuisance parameter \( \xi_2 \). (Case 2 in table 1 of section 2.3.)

Thus, the SLD CR bound for the parameter of interest is

\[ J_{\theta}^S_{11} = 1 - \frac{(\theta_1)^2}{1 - (\theta_0)^2}, \] (60)

which is lower than the previous example \( 1 - (\theta_1)^2 \). The optimal measurement is similarly worked out to find

\[ \Pi_{opt} = \left\{ \frac{1}{2} I \pm \frac{1}{\sqrt{1 + \phi_0(\theta_1)^2}}(\sigma_1 + \phi_0(\theta_1)\sigma_3) \right\}, \] (61)

where \( \phi_0(\theta_1) = \theta_0\theta_1/(1 + (\theta_0)^2) \) is a function of unknown parameter \( \theta_1 \). Note that this optimal measurement is independent of the nuisance parameter, we can achieve the above bound in the infinite sample size limit by using an appropriate adaptive strategy \( A2 \). We conclude that the locally unbiased estimator for the parameter of interest is useful here to achieve the bound \( C_{\theta}[W_I] \).

3.1.4. Nuisance parameter as a qubit phase. In the previous examples, optimal measurements are independent of nuisance parameters. And hence, we can safely ignore tradeoff relations between the MSEs for the parameter of interest and the nuisance parameter. In this example, we provide a simple example in which we cannot ignore the existence of a nuisance parameter.

Suppose we are interested in estimating the value of \( \theta_1 \) of the following model:

\[ \mathcal{M}_C = \left\{ \rho_\theta = \frac{1}{2} (I + \theta_1 \cos \theta_2 \sigma_1 + \theta_1 \sin \theta_2 \sigma_2) \mid \theta_1 \in (0,1), \theta_2 \in [0,2\pi) \right\}, \]

where the phase \( \theta_2 \) is the nuisance parameter. The inverse of the SLD Fisher information matrix of this model is

\[ (J_{\theta}^S)^{-1} = \begin{pmatrix} 1 - (\theta_1)^2 & 0 \\ 0 & (\theta_1)^{-2} \end{pmatrix}. \] (62)

that is \( \theta_1 \) and \( \theta_2 \) is globally orthogonal with respect to the SLD Fisher information.

When the precise value of the phase \( \theta_2 \) is known, the optimal measurement is given by the projection measurement about the SLD operator:

\[ L_{\theta,1}^S = \frac{1}{1 - (\theta_1)^2}(-\theta_1 I + \theta_1 \cos \theta_2 \sigma_1 + \theta_1 \sin \theta_2 \sigma_2). \] (63)
That is

$$\Pi_{\text{opt}} = \left\{ \frac{1}{2} \left[ I \pm (\cos \theta_2 \sigma_1 + \sin \theta_2 \sigma_2) \right] \right\}, \quad (64)$$

and the Fisher information about this optimal estimator is

$$J_\theta[\Pi] = J_{\theta,11}^S = \left[ 1 - (\theta_1)^2 \right]^{-1}. \quad (65)$$

Next, let us consider the case when $\theta_2$ is not known, that is $\theta_2$ is the nuisance parameter. Since this measurement $\Pi_{\text{opt}}(\theta_2)$ depends only on the nuisance parameter $\theta_2$, there is no way to perform it in the presence of nuisance parameter $\theta_2$. We also note that the measurement $\Pi_{\text{opt}}(\theta_2)$ gives a probability distribution with only two outcomes. It is then impossible to infer two parameters $\theta_1$ and $\theta_2$. The score functions become linearly dependent, and hence, the Fisher information matrix is singular.

$$J_\theta[\Pi_{\text{opt}}] = \frac{1}{1 - \theta_1^2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (66)$$

We then consider more restricted class of estimators, which are locally unbiased for all parameters. In this case we follow the proposed method to get the upper bound (44) for any locally unbiased estimator at $\theta$:

$$C_{\theta}^{\text{Nui}}[W_I = 1, W_I, V(\theta_N)] = (J_{11}^S)^{-1} + \left( \frac{J_{12}^S}{V_{0,22}} - (J_{22}^S)^{-1} \right)^2, \quad (66)$$

which is strictly greater than $J_{\theta,11}^S = (J_{11}^S)^{-1}$.

This simple model exhibits a unique feature of the nuisance parameter problem in the quantum case. Estimating a two-parameter qubit state in the presence of unknown phase as a nuisance parameter was discussed in [21], which provides a discussion on asymptotic achievability of the bound.

3.2. 1 + 2 qubit model

3.2.1. General formula. When the qubit model contains three unknown parameters, the achievable bound is given by the Hayashi–Gill–Massar (HGM) bound [42, 43]. This is defined by

$$C_{\theta}^{\text{HGM}}[W_3] := \left[ \text{Tr} \left\{ \sqrt{W_3(J_\theta^S)^{-1}\sqrt{W_3}} \right\} \right]^2. \quad (67)$$

First, the most informative bound for the parameter of interest $\theta_1$ can be evaluated by the weight-matrix limit. With a straightforward manner, we get

$$C_{\theta}[W_I = 1] = \lim_{W_2 \to W_I = 1} C_{\theta}^{\text{Nui}}[W_2] = J_{\theta}^{S,11}, \quad (66)$$

where the weight matrix for the parameter of interest is set to as 1 without loss of generality. From this expression, we can evaluate information loss due to the nuisance parameter (36).
Since the precision limit without the nuisance parameter is $(J_{θ,11}^{S})^{-1}$, the relevant information loss is

$$\Delta_θ[W_Y = 1] = J_θ^{S,11} - (J_{θ,11}^{S})^{-1}. \quad (68)$$

This result is identical to the previous $1 + 1$ qubit model.

For this class of model, it is hard to get a closed expression of the HGM bound since it involves calculations of the fidelity between $3 \times 3$ matrices. Below we derive an approximated bound by setting the weight matrix as a block matrix, i.e., the approximated bound $(28)$ and then making additional approximations. To simplify our result, we analyze the case where the SLD Fisher information is block diagonalized according to the parameter partition.

This can be done by performing the parameter orthogonalization. With these assumptions, the approximated bound $(28)$ is obtained by maximizing the $2 \times 2$ weight matrix $W_N$ as

$$\tilde{C}_N^{\text{HGM}}[W_Y, V(θ_N)] = \max_{W_N > 0} \left\{ C_θ^{\text{HGM}}[W_Y \oplus W_N] - \text{Tr} \{ W_N V_θ \} \right\}$$

$$= \max_{W_N > 0} \left\{ -\text{Tr} \{ W_N V_θ \} + \left( \sqrt{w_{11}J_θ^{S,11}} + \sqrt{\text{Tr} \{ W_N J_θ^{Θ,Θ} \}} + 2\sqrt{\det W_N J_θ^{Θ,Θ}} \right)^2 \right\}$$

$$\geq w_{11}J_θ^{S,11} + \max_{W_N > 0} \left\{ -\text{Tr} \{ W_N (V_θ - J_θ^{Θ,Θ}) \} + 2\sqrt{\det W_N J_θ^{Θ,Θ}} + 2\sqrt{w_{11}J_θ^{S,11}}(4\det W_N J_θ^{Θ,Θ})^{1/4} \right\}, \quad (69)$$

where the above inequality is due to the fact that $\text{Tr} \{ WA \} \geq 2\sqrt{\det W} \det A$ holds for all positive matrices $W$ and $A$. Since $V_θ ≠ J_θ^{Θ,Θ} > 0$ holds, we can reparametrize the weight matrix $W_N$ as $(V_θ - J_θ^{Θ,Θ})^{-1/2}W(V_θ - J_θ^{Θ,Θ})^{-1/2}$. Define the quantity

$$\delta_N = \sqrt{\frac{\det J_θ^{Θ,Θ}}{\det(V_θ - J_θ^{Θ,Θ})}} \quad (70)$$

and note $\delta_N < 1$, we get the following chain of inequalities:

$$\tilde{C}_N^{\text{HGM}}[W_Y, V(θ_N)] \geq w_{11}J_θ^{S,11} + \max_{W > 0} \left\{ -\text{Tr} \{ W \} + 2\delta_N\sqrt{\det W} + 4\sqrt{w_{11}J_θ^{S,11}}\delta_N(\det W)^{1/4} \right\}$$

$$\geq w_{11}J_θ^{S,11} + \max_{T = \text{Tr}[W] > 0} \left\{ -(1 - δ_N)T + \sqrt{8w_{11}J_θ^{S,11}}\delta_N\sqrt{T} \right\}$$

$$= w_{11}J_θ^{S,11} + \frac{2w_{11}J_θ^{S,11}\delta_N}{1 - \delta_N}$$

$$= w_{11}J_θ^{S,11} \left( 1 + \frac{2\delta_N}{1 - \delta_N} \right), \quad (71)$$

where the second inequality follows from $\text{Tr}[W] \geq 2\sqrt{\det W}$ and the monotonicity of a function of the form $f(x) = ax + b\sqrt{x}$ with $a, b > 0$. Third inequality is due to that of a quadratic
function and it is attained by \( \text{Tr}\{W\} = (2w_1d_{\theta}^{S,11}\delta_N)^{1/4}(1 - \delta_N)^{-1/2} \). Combining all yields the approximated upper bound for the parameter of interest as

\[
\tilde{C}_{\theta_1}^{\text{Nuis}}[W_1 = 1, V(\theta_N)] = J_{\theta}^{S,11} \left( 1 + \frac{2\delta_N}{1 - \delta_N} \right). \tag{72}
\]

The obtained upper bound (72) clearly shows that the estimation error is greater than the value \( J_{\theta}^{S,11} \) in the presence of nuisance parameters. Since \( 1 > \delta_N > 0 \), the second term in the right-hand side of (72) is positive. We note in passing that the inequality \( \delta_N < 1 \) has the same structure as the tradeoff relation (45) for the \( 1 + 1 \) qubit model.

Let us evaluate the tradeoff relation (37) due to the nuisance parameters. Since the above expression is approximated one, the tradeoff relation is lower bounded as

\[
\Delta C_{\text{Tradeoff}}[W_1 = 1] = C_{\theta_1}^{\text{Nuis}}[W_1 = 1, V(\theta_N)] - C_{\theta_1}[W_1 = 1, M_n] \\
\geq \tilde{C}_{\theta_1}^{\text{Nuis}}[W_1 = 1, V(\theta_N)] - C_{\theta_1}[W_1 = 1, M_n] \\
= J_{\theta}^{S,11} \frac{2\delta_N}{1 - \delta_N} \\
= J_{\theta}^{S,11} \frac{2\sqrt{\text{det} J_{\theta_N}^{\theta_N,\theta_N}}}{\sqrt{\text{det} (V_{\theta_N} - J_{\theta_N}^{\theta_N,\theta_N})} - \sqrt{\text{det} J_{\theta_N}^{\theta_N,\theta_N})}}, \tag{73}
\]

Since this quantity is positive, the approximated upper bound (72) is valid in our discussion. As in the previous class of \( 1 + 1 \) qubit model, when the value of MSE for the nuisance parameters gets larger, \( \text{det}(V_{\theta_N} - J_{\theta_N}^{\theta_N,\theta_N}) \) becomes large. This then makes the right-hand side of (73) smaller.

### 3.2.2. Estimating one parameter of the standard Stokes parameters

Consider the following family of qubit states, which is parametrized the standard Stokes parameters.

\[
\mathcal{M}_D = \{ \rho_b = \frac{1}{2}(I + \theta_1\sigma_1 + \theta_2\sigma_2 + \theta_3\sigma_3) \mid \theta \in \Theta \}. \tag{74}
\]

Let us assume that \( \theta_1 \) is the parameter of interest and \( \theta_N = (\theta_2, \theta_3) \) are the nuisance parameters. The parameter region \( \Theta \) is any subset of a three dimensional ball with the radius one: \( \mathcal{B}_3 = \{ (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3 \mid (\theta_1)^2 + (\theta_2)^2 + (\theta_3)^2 < 1 \} \). The inverse of the SLD Fisher information matrix for this model is

\[
(J_{\theta}^{S})^{-1} = \begin{pmatrix} 1 - \theta_1^2 & -\theta_1\theta_2 & -\theta_1\theta_3 \\ -\theta_1\theta_2 & 1 - \theta_2^2 & -\theta_2\theta_3 \\ -\theta_1\theta_3 & -\theta_2\theta_3 & 1 - \theta_3^2 \end{pmatrix}. \tag{75}
\]

Thus, \( \theta_1 \) and \( \theta_N \) are not orthogonal in this model. Following the parameter orthogonalization method, we introduce the new parameterization:

\[
\theta_1 = \xi_1, \quad \theta_2 = c_2(\xi_2)\sqrt{1 - (\xi_1)^2}, \quad \theta_3 = c_3(\xi_3)\sqrt{1 - (\xi_1)^2}, \tag{76}
\]
where $c_2(\xi_2)$ and $c_3(\xi_3)$ are arbitrary functions satisfying the same condition as in example of section 3.1.2. The inverse of the SLD Fisher information matrix in the $\xi$ representation is

$$
(J^S_\xi)^{-1} = \begin{pmatrix}
1 - (\xi_1)^2 & 0 & 0 \\
0 & \frac{1 - c_2^2}{c_2^2(1 - (\xi_1)^2)} & -\frac{c_2c_3}{c_2^2c_3(1 - (\xi_1)^2)} \\
0 & -\frac{c_2c_3}{c_2^2c_3(1 - (\xi_1)^2)} & \frac{1 - c_3^2}{c_3^2(1 - (\xi_1)^2)}
\end{pmatrix}.
$$

(77)

The relevant SLD operator is exactly same as (53). We calculate an optimal measurement from

$$
L_{ij}^{S,1} = \sum_{j=1}^{3} J_{ij}^S \epsilon_{ij}^S = -\theta I + \sigma_1,
$$

to find that the optimal measurement is the $\sigma_1$ measurement. We see that this optimal measurement is independent of both the parameter of interest and nuisance parameter. Therefore, the same conclusion as in example of section 3.1.2 holds. The SLD CR bound $J_{\theta}^{S,11} = 1 - (\theta_1)^2$ can be achieved explicitly by estimation strategy A1 under the condition of locally unbiasedness about $\theta_1$.

4. Qubit models with nuisance parameter: 2 + 1 qubit model

4.1. General formula

The achievable bound for this class of models is the HGM bound (67). The most informative bound for the parameters of interest can be evaluated by taking the weight-matrix limit as before.

$$
C_{\theta}[W_I, M_3] = \text{Tr} \left\{ W_2 J^S_{\theta} \right\} + 2 \sqrt{\text{det} W_2 J^S_{\theta}} ,
$$

(78)

where we use the block matrix representation of the SLD quantum Fisher information matrix (32). The precision limit without nuisance parameter is given by the Nagaoka bound $C_{\theta}^N[W_I, M_2] = \text{Tr} \left\{ W_I (J^S_{\theta})^{-1} \right\} + 2 \sqrt{\text{det} W_I (J^S_{\theta})^{-1}}$. Information loss in the presence of the nuisance parameter $\theta_N = \theta_3$ is

$$
\Delta_{\theta}[W_I] = C_{\theta}[W_I, M_3] - C_{\theta}^N[W_I, M_2] = \text{Tr} \left\{ W_2 \left[ J^S_{\theta} - (J^S_{\theta})^{-1} \right] \right\} + 2 \sqrt{\text{det} W_2} \left[ \sqrt{J^S_{\theta}} - \sqrt{\text{det}(J^S_{\theta})^{-1}} \right].
$$

(79)

It is easy to see that this can be zero if and only if $J^S_{\theta} = (J^S_{\theta})^{-1}$ holds. In other words, the parameters of interest $\theta_I = (\theta_1, \theta_2)$ and the nuisance parameter $\theta_N = \theta_3$ are orthogonal with respect to the SLD quantum Fisher information matrix. Combining this result with those of the previous section, we conclude that information loss due to the nuisance parameter(s) in the qubit case can be zero if and only if two sets of parameters are orthogonal with respect to the SLD Fisher information matrix.

We next examine the proposed method of eliminating the weigh matrix for the nuisance parameter. The same remark applies here as the case of 1 + 2 model in section 3.2. For this model, we assume that orthogonality condition holds between $\theta_I = (\theta_1, \theta_2)$ and $\theta_N = \theta_3$. We
set the weight matrix as a block diagonal one. A similar procedure can be carried out to derive the approximated upper bound as follows.

\[
\tilde{C}_{\theta}^{\text{Nai}}[W_I, V(\theta_N)] = \max_{w_{33} > 0} \left\{ C_{\theta}^{\text{HGM}}[W_I \oplus w_{33}] - \text{Tr} \{ w_{33} V_{0,33} \} \right\} = \max_{w_{33} > 0} \left\{ (\sqrt{C_{\theta}[W_I]} + \sqrt{w_{33} J_{\theta}^{S,33}})^2 - \text{Tr} \{ w_{33} V_{0,33} \} \right\} = C_{\theta}[W_I] + \max_{w_{33} > 0} \left\{ - \text{Tr} \{ w_{33}(V_{0,33} - J_{\theta}^{S,33}) \} + 2 \sqrt{C_{\theta}[W_I] w_{33} J_{\theta}^{S,33}} \right\} = C_{\theta}[W_I] \left( 1 + \frac{J_{\theta}^{S,33}}{V_{0,33} - J_{\theta}^{S,33}} \right),
\]

(80)

where the maximization in the third line follows from that about a quadratic function. In the above expression, \( C_{\theta}[W_I] = \text{Tr} \{ W_I J_{\theta}^{S,\theta_N} \} + 2 \sqrt{\det W_I J_{\theta}^{S,\theta_N}} \) denotes the most informative bound for the parameters of interest.

The above result indicates a similar structure as the other qubit models by noting \( \det(V_{0,33} - J_{\theta}^{S,33}) = V_{0,33} - J_{\theta}^{S,33} \) and \( \det J_{\theta}^{S,\theta_N} = J_{\theta}^{S,33} \). To suppress the effect of the nuisance parameter, we need to make \( V_{0,33} - J_{\theta}^{S,33} \) as large as possible, but this in turn implies large estimation errors for \( \theta_N \). Further, we observe that the quantity \( J_{\theta}^{S,33} / (V_{0,33} - J_{\theta}^{S,33}) \) (1 takes again the form \( \det J_{\theta}^{S,\theta_N} / \det(V_{0,33} - J_{\theta}^{S,33}) \), (Cf, the tradeoff relation (45) for the 1 + 2 qubit model).

Lastly, we discuss a tradeoff relation based on the above result. Following the same argument for the 1 + 2 qubit model case, we obtain

\[
\Delta C_{\theta}^{\text{Tradeoff}}[W_I] \geq C_{\theta}^{\text{Nai}}[W_I, V(\theta_N)] - C_{\theta}[W_I, M_a] = C_{\theta}[W_I] \frac{J_{\theta}^{S,33}}{V_{0,33} - J_{\theta}^{S,33}}.
\]

(81)

This quantity is positive and vanishes as the MSE for the nuisance parameter \( V_{0,33} \) diverges.

### 4.2. Estimating two parameters for a qubit model

We consider the same model \( M_2 \) as the example in section 3.2.2, but with different parameters of interest, \( \theta_1 = (\theta_1, \theta_2) \) are the parameters of interest and \( \theta_3 \) is the nuisance parameter. The inverse of the SLD Fisher information matrix is given by (75). Therefore, the relevant block submatrix of the inverse SLD Fisher information is

\[
J_{\theta,\theta_3}^{S,3} = \begin{pmatrix}
1 - \theta_1^2 & -\theta_1 \theta_2 \\
-\theta_1 \theta_2 & 1 - \theta_2^2
\end{pmatrix}.
\]

(82)

An immediate question here is whether we can achieve the Nagaoka bound for the parameter of interest in the presence of nuisance parameter \( \theta_3 \). Below, we show the following result about the CR bound for the parameters of interest.

\[
C_{\theta}[W_I] = \text{Tr} \{ W_I J_{\theta}^{S,\theta_N} \} + 2 \sqrt{\det W_I J_{\theta}^{S,\theta_N}}.
\]

(83)

To make our discussion clear, it is convenient to utilize the method of parameter orthogonalization. Introduce the new parametrization \( \xi = (\xi_1, \xi_2, \xi_3) \) by...
\[ \theta_1 = \xi_1, \quad \theta_2 = \xi_2, \quad \theta_3 = c(\xi_3)\sqrt{1 - (\xi_1)^2 - (\xi_2)^2}, \quad (84) \]

where \( c(\xi_3) \) is arbitrary function satisfying the same conditions as previous examples. In this new parametrization, the inverse of SLD Fisher information takes

\[ (J_\xi^S)^{-1} = \begin{pmatrix} J_{\xi_1}^{G\xi} & 0 \\ 0 & \frac{1 - c(\xi_3)^2}{\xi(\xi_1)^2(1 - (\xi_1)^2 - (\xi_2)^2)} \end{pmatrix}, \quad (85) \]

where the \( 2 \times 2 \) matrix \( J_{\xi_1}^{G\xi} \) is

\[ J_{\xi_1}^{G\xi} = \begin{pmatrix} 1 - \xi_1^2 & -\xi_1\xi_2 \\ -\xi_1\xi_2 & 1 - \xi_2^2 \end{pmatrix}. \quad (86) \]

With this parameter orthogonalization method, the weight-matrix limit immediately gives

\[ C_{\lim}^{\text{sl}}[W] = \text{Tr} \left\{ W J_{\xi}^{G\xi} \right\} + 2 \sqrt{\text{Det} \left\{ W J_{\xi}^{G\xi} \right\}}. \]

Two relevant SLD operators for the parameter of interest \( \xi = (\xi_1, \xi_2) \) are

\[ L_{\xi_1}^S = \frac{1}{1 - |\xi_1|^2}[-\xi_1 I + \xi_1 \xi_2 \sigma_1 + (1 - (\xi_1)^2) \sigma_2], \]
\[ L_{\xi_2}^S = \frac{1}{1 - |\xi_2|^2}[-\xi_1 I + (1 - (\xi_2)^2) \sigma_1 + \xi_1 \xi_2 \sigma_2]. \quad (87) \]

with \( |\xi_1|^2 = (\xi_1)^2 + (\xi_2)^2 \). They depend only on the parameters of interest. From the SLD operators about \( \xi_1 \), we see that an optimal measurement for \( \xi_1 \) is solely determined by certain combinations of \( L_{\xi_1}^S \) and \( L_{\xi_2}^S \), see for example [60]. We then conclude that it is independent of the nuisance parameter \( \xi_3 \). Furthermore, the resulting probability distribution from the optimal measurement only depends on the parameters of interest, we can construct a locally unbiased estimator for the parameter of interest at any point. We thus have \( C_\theta[W] = \text{Tr} \left\{ W J_{\xi}^{G\xi} \right\} + 2 \sqrt{\text{Det} \left\{ W J_{\xi}^{G\xi} \right\}} \). This is, of course, true from the general theorem 2.2. Noting \( \theta = (\theta_1, \theta_2) = (\xi_1, \xi_2) \), we prove our claim: \( C_\theta[W] = C_\xi[W] \) is the adaptively achievable bound for the parameter of interest.

5. Examples in open quantum systems

5.1. The isotropic noise

In this example, we shall analyze a qubit model given by the solution to the quantum master equation of the form:

\[ \frac{\partial}{\partial t} \rho(t) = i[H_0, \rho(t)] - \frac{1}{4} \sum_{i=1,2,3} \gamma_i [\sigma_i, [\sigma_i, \rho(t)]] , \quad (88) \]

in the unit of \( \hbar = 1 \). Here, the free Hamiltonian is \( H_0 = \omega \sigma_3 / 2 \) and we wish to estimate the frequency \( \omega \) from the evolution of the system. If we specify an initial state as \( s_0 = (s_1, s_2, s_3) \)
in terms of the Bloch vector and for a given later time \( t \), which is fixed, we can consider a model:

\[
\mathcal{M}_F = \{ \rho(\theta) | \theta \in \Theta \},
\]

(89)

where \( \theta = (\theta_1, \theta_N) \) with \( \theta_1 = \omega t \) and \( \theta_N \) represents the damping parameters \( \gamma_i \).

When the precise values of all damping parameters are known, the model is a single parameter. The SLD CR bound provides the achievable bound:

\[
V_0[\hat{\Pi}] \geq \left( J_{S,11}^\theta \right)^{-1}.
\]

(90)

If the precise values are not known, the model is specified by the four parameters \((\omega, \gamma_1, \gamma_2, \gamma_3)\). Then, it is in general impossible to attain the above precision limit unless the model satisfies special conditions.

Further, let us analyze the simplest case where all damping parameters are equal, i.e., the isotropic noise model. Set \( \theta_2 = \gamma_1 t = \gamma_2 t = \gamma_3 t \), and thus have a two-parameter model. The SLD Fisher information matrix of this model is

\[
J_\theta^S = \begin{pmatrix}
  e^{-2\theta_2}(s_1^2 + s_2^2) & 0 \\
  0 & e^{-2\theta_2}|s_0|^2 \\
  1 - e^{-2\theta_2}|s_0|^2 & 1 - e^{-2\theta_2}|s_0|^2
\end{pmatrix},
\]

(91)

which is diagonal and is independent of the parameter of interest \( \theta_1 \). Here and below, we assume that \( s_1^2 + s_2^2 \neq 0 \).

It happens that the projectors about the SLD operator \( L_{S,11}^\theta \) is independent of \( \theta_2 \). Therefore, we conclude that the bound

\[
\left( J_{S,11}^\theta \right)^{-1} = J_{11} = e^{2\theta_2}(s_1^2 + s_2^2)^{-1},
\]

(92)

can be achieved by performing the following projection measurement:

\[
\left\{ \frac{1}{2} \left[ I + \frac{1}{|s_0|}(s_1 \sin \theta_1 + s_2 \cos \theta_1)\sigma_1 \pm \frac{1}{|s_0|}(s_1 \cos \theta_1 - s_2 \sin \theta_1)\sigma_2 \right] \right\}.
\]

It should be emphasized that the above simple result holds only for a special class of noise models. One of key ingredients is that two processes of unitary and damping are completely factorized in the following sense. Denoting the state in terms of the Bloch vector \( s(t) \) at later time \( t \), the dynamics is given

\[
s(t) = \Gamma(t)u_0(t)s_0,
\]

(93)

where \( u_0(t) \) denotes the rotation according to the free Hamiltonian and \( \Gamma(t) \) accounts for the damping process, which is equivalent to the dephasing channel. If a solution to a given master equation is not factorized or the damping matrix \( \Gamma(t) \) also depends on the parameter of interest \( \theta \), we can no longer use the bound \( \left( J_{S,11}^\theta \right)^{-1} \) as the ultimate bound. This point will be discussed in the next example.

5.2. A realistic noise model

So far, we have provided relatively simple examples in which we cannot ignore the nuisance parameter. For the last example, we shall consider a physically motivated model, a single-spin model affected by random magnetic fields. To simplify our notation we set \( \hbar = 1 \) and the magnetic moment \( \mu_B = 1 \) unless noted otherwise.
5.2.1. Model setting. Consider a frequency estimation problem described by the Hamiltonian 
\[ H = H_0 + H_{\text{noise}}(t) \], where \( H_0 = \omega_3/2 \) generates a unitary evolution as in the previous example and 
\( H_{\text{noise}}(t) \) describes the effect of random magnetic fields defined by

\[ H_{\text{noise}}(t) = \frac{1}{2} \sum_{i=1}^{3} b_i(t) \sigma_i. \] (94)

Here the time-dependent magnetic field \( b(t) = (b_1(t), b_2(t), b_3(t)) \) obeys the Maxwell equation
and is fluctuating with the following time average characteristics:

\[ \overline{b_i(t)} = 0, \quad \overline{b_i(t) b_j(t')} = \delta_{ij} b^2 e^{-\Gamma|t-t'|}, \] (95)

where the bars denote the time average over a classical probability density function (See [61]
for the derivation and more details).

We follow the standard quantum master equation approach to derive the differential equation
for the qubit-state density matrix:

\[ \frac{\partial}{\partial t} \rho(t) = -i[H_0, \rho(t)] + \mathcal{L}[\rho(t)], \]

\[ \mathcal{L}[\rho(t)] = -\frac{i}{2} [\delta \omega(t) \sigma_3, \rho(t)] - \frac{1}{4} \sum_{i=1}^{3} \gamma_i(t) [\sigma_i, [\sigma_i, \rho(t)]] , \] (96)

where the time-dependent model parameters are

\[ \delta \omega(t) = b^2 \int_0^t d\tau' e^{-\Gamma \tau'} \sin \omega \tau', \]

model 1: \[ \gamma_1(t) = \gamma_2(t) = b^2 \int_0^t d\tau' e^{-\Gamma \tau'} \cos \omega \tau', \]

\[ \gamma_3(t) = b^2 \int_0^t d\tau' e^{-\Gamma \tau'}. \] (97)

To get the above equation, we only used the following approximation: \[ \rho(t') \approx e^{-iH_0(t'-t')/\omega} \rho(t) e^{iH_0(t'-t')/\omega} \], which is valid when the coupling is small \((b/\omega)^2 \ll 1\).

In the Born-Markov approximation, we further set \( t \to \infty \) in the upper limits for the integrations to get time-independent model parameters:

model 2: \[ \delta \omega = \frac{b^2 \omega}{\Gamma^2 + \omega^2}, \quad \gamma_1 = \gamma_2 = \frac{b^2 \Gamma}{\Gamma^2 + \omega^2}, \quad \gamma_3 = \frac{b^2}{\Gamma}. \] (98)

This approximation is valid when time is much later than the decoherence time \( \Gamma^{-1} \).

When the noise parameters \((b, \Gamma)\) are much smaller than \(\omega\), we can further simplify the model to arrive at a so-called parallel noise model:

model 3: \[ \delta \omega = 0, \quad \gamma_1 = \gamma_2 = 0, \quad \gamma_3 = \frac{b^2}{\Gamma} = : \gamma. \] (99)

This approximation is valid, for example, the strength of the fluctuating magnetic field is of order \( b \sim 10^{-9} \) T and the noise correlation is \( \Gamma \sim 100 \) Hz in the usual units such that the relation \( b, \Gamma \ll \omega \) holds [61].
In the following we compare the above three models described by different approximations ((97)–(99)). For convenience, we call them model 1, model 2, and model 3, respectively. Under the assumption that time $t$ is precisely known, three models are characterized by three parameters $(\theta_1, \theta_2, \theta_3) = (\omega, b^2, \Gamma)$. (Model 3 depends essentially on two parameters ($\omega, \gamma$) only.) Here $\theta_1 = \omega$ is the parameter of interest whereas other noise parameters are the nuisance parameters. We can explicitly solve the equation (96) about the time evolution of the Bloch vector for a given initial state $s_0 = (s_1, s_2, s_3)$. Let $s^{\theta_1}_q(t)$ be the Bloch vector at time $t$ for model $q$ ($q = 1, 2, 3$). These parametric models $\mathcal{M}^{\theta_1}$ are denoted by

$$\mathcal{M}^{\theta_1} = \{s^{\theta_1}_q(t) \mid \theta = (\omega, b^2, \Gamma) \in \Theta\}. \quad (100)$$

More explicitly, each model is given by the following form:

$$\mathcal{M} = \{s_\theta(t) = A_\theta(t)s_0 \mid \theta = (\omega, b^2, \Gamma) \in \Theta\},$$

$$A_\theta(t) := \begin{pmatrix} e^{-\Gamma_1(t)} & \cos \Omega(t) \sin \Omega(t) & 0 \\ -\sin \Omega(t) \cos \Omega(t) & 0 & e^{-\Gamma_3(t)} \\ 0 & 0 & 0 \end{pmatrix},$$

$$\Gamma_1(t) = \int_0^t dt' [\gamma_1(t') + \gamma_2(t')], \quad \Gamma_3(t) = 2 \int_0^t dt' \gamma_3(t'),$$

$$\Omega(t) = \omega t + \int_0^t dt' \delta \omega(t'). \quad (101)$$

Here, we omit the model index $q$ for simplicity. We then compute the SLD Fisher information matrix $J^{S^{\theta_1}_q}$ for model $q$ ($q = 1, 2, 3$). The explicit expressions of $J^{S^{\theta_1}_q}$ for model 1, 2 are rather complicated and are omitted. Model 3 takes a similar form as the previous example in section 5.1.

First note that, in the absence of external noise, the SLD Fisher information about the parameter $\theta_1 = \omega$ is $J^{S^{\theta_1}_q} = (s_1^2 + s_2^2)\omega^2$. The maximum sensitivity is achieved by arbitrary pure initial state on $xy$ plane, and the optimal MSE is $r^{-2}$. This observation also holds in model 3 as previously known, and the maximum SLD Fisher information matrix with this optimal initial state is $(J^{S^{\theta_1}_3})_{ij} = (J^{S^{\theta_1}_3}_{ij})^{-1} = \exp(-2\gamma t) r^{-2}$. Note that the effect of nuisance parameter $\gamma$ is irrelevant in model 3 and this bound can be achieved asymptotically by the adaptive scheme. However, model 2 and 3 do not satisfy the condition of parameter orthogonality with respect to the SLD Fisher information matrix, and further analysis is needed.

5.2.2. Comparison and discussion. We now compare three models to see the validity of frequently used approximations. Model 3 is the simplest one and has been widely adopted. (Cf, the isotropic noise model in section 5.1 is another familiar example of this kind.) Model 2 takes into account the effect of random fields up to the second order in the coupling constant $b$. This is based on the assumption of small $b/\omega$, and the solution is valid for large $\Gamma t$ regime. Model 1 is more general than others, in which the kernel for the master equation $\gamma_i(t)$ is time-dependent. Some authors regard this model as non-Markovian in the sense of finite correlation time for the decaying factor. In this paper, however, we prefer to call model 3 as the weak-coupling model with a finite memory, and this model is valid as long as $b/\omega$ is small.

It is widely known that a non-Markovian environment may bring back coherence in some cases when compared with the Markovian case. In quantum metrology, several authors reported
Figure 1. Dynamics of the SLD Fisher information (in units of $r^{-2}$) about the parameter of interest $\theta_1 = \omega$ for three different approximations ((97)–(99)). Parameters $(\omega, b^2, \Gamma)$ are figure 1(a); $(10^5, 10^2, 10)$, figure 1(b); $(10^2, 10^2, 10)$, and figure 1(c); $(10^2, 10^2, 1)$. Five curves are model 1; $(J_{11}^{(1)})^{-1}$ (blue solid line), $J_{11}^{(1)}$ (red solid line). Model 2; $(J_{12}^{(1)})^{-1}$ (blue dotted line), $J_{12}^{(1)}$ (red dotted line). Model 3; $(J_{11}^{(3)})^{-1} = J_{11}^{(3)}$ (gray solid line). Insets show the logarithmic plot.
this ‘memory effect’ showing that the SLD Fisher information decays much slower in time than the Markovian case, see for example [62–64]. We now analyze if this memory effect is true even when noise parameters are not completely known, that is, they are treated as the nuisance parameters.

In the following, we compare five different quantities related to the SLD Fisher information about the parameter of interest \( \theta_1 = \omega \). Fix a pure initial state \( s_0 = (\sqrt{1 - \omega^2}, 0, \omega) \) and let it evolve according to three noise models. The Bloch vector at later time \( t \) for model \( q \) \((q = 1, 2, 3)\) is denoted by \( s_q^0(t) \) corresponding to (101). We write \( (1, 1) \) components of the SLD Fisher information matrix and the inverse of SLD Fisher information matrix about model \( q \) by

\[
J_{11}^{(q)}(\omega, b^2, \Gamma) := \left( J^{(q)}_\theta \right)_{11},
\]

\[
J_{11}^{11}(\omega, b^2, \Gamma) := \left( J^{(q)}_\theta \right)^{-1}_{11},
\]

respectively. (In the following, we drop the superscript for the SLD Fisher information to simplify notation.) The quantity \((J_{11}^{(q)})^{-1}\) represents the precision limit when all noise parameters \((b^2, \Gamma)\) are known, i.e., without nuisance parameters. The quantity \(J^{(q)}_{\theta \theta}\), on the other hand, is the precision limit when we treat the noise parameters as the nuisance parameters. (Cf, its inverse is called the partial SLD Fisher information.) The general relationship \(J_{11}^{(q)} \geq (J^{(q)}_{\theta \theta})^{-1}\) holds for all parameters, and \(J_{11}^{(3)} = (J^{(3)}_{\theta \theta})^{-1}\) holds as a special case.

In the following, we plot \(J_{11}^{(q)}/t^2\) and \((J_{11}^{(q)})^{-1}/t^2\) \((q = 1, 2, 3)\) as functions of time for fixed parameters \((\omega, b^2, \Gamma)\) and the initial state \(z \in [0, 1]\). In figure 1, we plot five functions: model 1; \((J_{11}^{(1)})^{-1}\) (blue solid line), \(J_{11}^{(1)}\) (red solid line), model 2; \((J_{11}^{(2)})^{-1}\) (blue dotted line), \(J_{11}^{(2)}\) (red dotted line). Model 3; \((J_{11}^{(3)})^{-1} = J_{11}^{(1)}\) (gray solid line). Parameter values \((\omega, b^2, \Gamma)\) are set to \((10^3, 10^2, 10)\) for figure 1(a), \((10^2, 10^2, 10)\) for figure 1(b), and \((10^2, 10^2, 1)\) for figure 1(c), respectively. The initial state parameter is chosen as \(s_0 = (\sqrt{0.91}, 0, 0.3)\). The ratio \((b/\omega)^2\) represents the weakness of the coupling, which is to be weak in order for the master equation approach to be valid. The values for above three choices figure 1(a), 1(b), 1(c) are \((b/\omega)^2 = 10^{-4}, 10^{-2}, 10^{-4}\), that satisfy the weak-coupling condition \((b/\omega)^2 \ll 1\). Insets of the figures show the logarithmic plot for longer time behaviors.

Let us now discuss the result shown in figure 1(a). In the absence of nuisance parameters, model 2 and 3 are almost identical confirming that model 3 is a good approximation to model 2 within our parameter setting. Model 1, which is more accurate than others, has larger SLD Fisher information about the parameter of interest \(\theta_1 = \omega\) for the short time scale. This effect is known as the memory effect in literature and is believed to bring an advantage upon estimating \(\omega\). This observation is seen in our example.

When the noise parameters \((b^2, \Gamma)\) are treated as the nuisance parameters, on the other hand, we have to use \((J_{\theta \theta}^{(q)})^{-1}\) as the correct SLD Fisher information about the parameter of interest \(\theta_1 = \omega\). In this case, we observe completely different behaviors for model 1 and model 2. First, model 2 (blue dotted line) immediately drops down when compared with the case of no nuisance parameters (red dotted line). Model 1 (blue solid line), which is valid for the short time scale, exhibits amplification of the SLD Fisher information, however, this effect is much weaker than the case of no nuisance parameters (red solid line). Figures 1(b) and (c) also show qualitatively same behaviors as figure 1(a). The major difference is that the enhancement due to the memory effect is more visible than figure 1(a). In particular, good amplification is observed in figure 1(c). The above results show that the effect of nuisance parameters are not negligible even in the weak-coupling limit. Furthermore, the memory effect, which is due to a
time-dependent kernel for the master equation, strongly depends on approximations that need to be justified.

6. Summary and outlook

In this paper, we have formulated the nuisance parameter problem in the quantum estimation theory for any finite sample size case. Unlike the classical case, the estimation error bound for multi-parameter case cannot be expressed as a simple formula in terms of the partial quantum Fisher information. This is partly due to the intrinsic nature of the problem involving measurement degrees of freedoms. We have obtained the fundamental precision limit for estimating the parameters of interest under the locally unbiasedness for the parameters of interest. This bound can be also evaluated by taking the weight-matrix limit. We have clarified the operational meanings of obtained bounds.

We have proposed a method, the weight-matrix elimination method, to derive an upper bound for estimating the parameters of interest by eliminating the effects of the nuisance parameters from a given bound when estimating all parameters. This is based on the similar philosophy as in the classical case, i.e., the best estimation strategy is to estimate all parameters including the nuisance parameters. In the general case, we should not ignore a tradeoff relation between the mean square error of the parameters of interest and the nuisance parameters. Another key observation in our results is that a class of estimators needs to be clarified when dealing with the nuisance parameter problem in the quantum case. Otherwise, a claim of achievability for a bound might not be conclusive. Based on our general formulation of the problem, the general qubit models and several examples are examined to illustrate these findings.

There are several issues that have not been explored in this paper. First is adaptive or sequential measurement schemes to implement an optimal measurement in the presence of nuisance parameters. Optimal design of experiments [33–37] is a proper statistical language to deal with this problem.

Second, the nuisance parameter problem is also important when estimating quantum channel parameters. For the channel estimation problem, we need to optimize over input quantum states to extract as much information as possible. In the presence of nuisance parameters, an optimal input state also depends on them, and hence a similar tradeoff relation involves in general. Based on our formalism, we investigated the problem of detecting asymmetry in the qubit Pauli channel and showed that the effect of the nuisance parameter cannot be ignored [38].

Last, practically efficient estimation strategies. In classical statistics, we do not intend to apply the best estimator when dealing with the nuisance parameter problem. But rather, we look for a sub-optimal estimator to suppress the effect of nuisance parameters. A typical example is to construct a pretty good estimator for the parameters of interest without any or partial knowledge about the nuisance parameters. A challenge in the quantum case is to find a good measurement that does not depend on the nuisance parameters and to discuss how efficient we can extract information about the parameters of interest.

The nuisance parameter problem in quantum systems is a common and practical problem when analyzing any statistical decision problem such as estimation, discrimination, control, and so on. Hence, it is relevant to any quantum information processing protocol in a noisy environment. As in the classical case, we need to develop proper statistical tools to handle them, since we cannot break the laws of statistics as well as quantum theory. This paper is a beginning of research along this line, and we shall develop useful statistical tools in the subsequent publication.
Acknowledgments

The work is partly supported by JSPS KAKENHI Grant Number JP17K05571. The author is indebted to Prof A Fujiwara and Prof H Nagaoka for invaluable discussions and suggestions. He would like to thank Dr K Yamagata for useful comments on the work. He would also like to thank Prof H K Ng for discussions and her kind hospitality at Centre for Quantum Technologies in Singapore where part of this work was done. Constructive discussions with Prof M Hayashi and Dr Y Yang are also acknowledged upon completing the review paper on the same subject [27]. We lastly wish to thank anonymous referee and the Guest Editor for constructive comments to improve the manuscript. After we completed this work [66], we have noticed a preprint [67] formulating the nuisance parameter problem in the framework of quantum semiparametric estimation.

Appendix A. Supplemental materials

A.1. Classical statistics

This appendix briefly summarizes the nuisance parameter problem in classical statistics. More details can be found in books [2–5] and relevant papers for this work [6–10]. Other various aspects of the nuisance parameter problem are explained in the recent review paper [27].

Consider an \( n \)-parameter family of probability distributions \( p_\theta(x) \) on a real-valued set \( X \), where the \( n \)-dimensional real parameter \( \theta = (\theta_1, \theta_2, \ldots, \theta_n) \) takes values in an open subset of \( n \)-dimensional Euclidean space \( \Theta \subset \mathbb{R}^n \). When we are interested only in estimating the values of a certain subset of parameters \( \theta_I = (\theta_1, \theta_2, \ldots, \theta_m) \) \( (m < n) \), we call them parameters of interest. The remaining set of \( n - m \) parameters \( \theta_N = (\theta_{m+1}, \theta_{m+2}, \ldots, \theta_n) \) are the nuisance parameters. We denote this partition as \( \theta = (\theta_I, \theta_N) \). Let \( \hat{\theta}_I \) be an estimator for the parameters of interest, and we define the mean square error (MSE) matrix for the parameters of interest by

\[
V_{\theta_I}[\hat{\theta}_I] = \sum_{x \in X} p_\theta(x)(\hat{\theta}_i(x) - \theta_i)(\hat{\theta}_j(x) - \theta_j)
= E_{\theta}[\hat{\theta}_I(X) - \theta_I)(\hat{\theta}_J(X) - \theta_J)].
\]

Here, the matrix index \((i, j)\) takes values in the index set of parameters of interest, i.e., \( i, j \in \{1, 2, \ldots, m\} \), and \( E_{\theta}[f(X)] \) denotes the expectation value of a random variable \( f(X) \) with respect to \( p_\theta(x) \).

The well-established result in classical statistics proves the following Cramér–Rao (CR) inequality:

\[
V_{\theta_I}[\hat{\theta}_I] \geq \begin{cases} 
(J_{\theta_I\theta_I})^{-1} & (\theta_N \text{ is known}) \\
J_{\theta_I\theta_I} & (\theta_N \text{ is not known}),
\end{cases}
\]

(A.2)

(A.3)

for any (locally) unbiased estimators \( \hat{\theta}_I \) about the parameters of interest (see for example [27] about the detailed derivation). In this formula, two matrices \( J_{\theta_I\theta_I} \) and \( J_{\theta_I\theta_N} \) are defined by the block matrices of the Fisher information matrix \( J_\theta \) according to the partition \( \theta = (\theta_I, \theta_N) \);

\[
J_\theta = \begin{pmatrix}
J_{\theta_I\theta_I} & J_{\theta_I\theta_N} \\
J_{\theta_N\theta_I} & J_{\theta_N\theta_N}
\end{pmatrix}, \quad J_\theta^{-1} = \begin{pmatrix}
J_{\theta_I\theta_I}^{-1} & J_{\theta_I\theta_N} J_{\theta_I\theta_I}^{-1}
\\
J_{\theta_N\theta_I} J_{\theta_I\theta_I}^{-1} & J_{\theta_N\theta_N}^{-1}
\end{pmatrix}.
\]

(A.4)
where the \((i, j)\) component of the Fisher information matrix is defined by

\[
J_{\theta_i \theta_j} = \sum_{x \in X} p_\theta(x) \left[ \frac{\partial}{\partial \theta_i} \log p_\theta(x) \right] \left[ \frac{\partial}{\partial \theta_j} \log p_\theta(x) \right]
\]

\[
= E_\theta \left[ \frac{\partial}{\partial \theta_i} \log p_\theta(X) \frac{\partial}{\partial \theta_j} \log p_\theta(X) \right] .
\] (A.5)

The inverse of the sub-block matrix in inequality (A.3),

\[
(J_{\theta_i \Theta_N})^{-1} = (J_{\theta_i \Theta} - J_{\theta_i \Theta_N} J_{\Theta_N \Theta_N}^{-1} J_{\theta_i \Theta_N} )^{-1}
\]

is known as the partial Fisher information matrix about \(\theta_i\).

A simple proof of inequalities (A.2) and (A.3) is as follows. When \(\theta_N\) is known, the model \(\mathcal{M}_n\) is reduced to an \(m\)-dimensional model. Hence, we can apply the standard CR inequality to get inequality (A.2). When \(\theta_N\) is not completely known, on the other hand, consider an estimator \(\hat{\theta} = (\hat{\theta}_i, \hat{\theta}_N)\) for all parameters \(\theta = (\theta_i, \theta_N)\) and denote its MSE matrix by \(V_\theta[\hat{\theta}]\). Then, the CR inequality for the model \(\mathcal{M}_n\) is

\[
V_\theta[\hat{\theta}] \geq J_{\theta_i}^{-1}.
\] (A.7)

for any locally unbiased estimator \(\hat{\theta}\) at \(\theta\). Let us decompose the MSE matrix as

\[
V_\theta[\hat{\theta}] = \left( \begin{array}{cc} V_{\theta_i} & V_{\theta_i \theta_N} \\ V_{\theta_N \theta_i} & V_{\theta_N} \end{array} \right),
\] (A.8)

then, applying the projection onto the subspace \(\theta_i\) gives the desired result (A.3).

It is well known that the following matrix inequality holds.

\[
J_{\theta_i} = (J_{\theta_i \Theta} - J_{\theta_i \Theta_N} J_{\Theta_N \Theta_N}^{-1} J_{\theta_i \Theta_N})^{-1} - (J_{\theta_i \Theta})^{-1}.
\] (A.9)

Here the equality holds if and only if the off-diagonal block matrix vanishes, i.e., \(J_{\theta_i \theta_N} = 0\). When \(J_{\theta_i \theta_N} = 0\) holds at \(\theta\), we say that two sets of parameters \(\theta_i\) and \(\theta_N\) are orthogonal with respect to the Fisher information matrix at \(\theta\) or simply \(\theta_i\) and \(\theta_N\) are orthogonal at \(\theta\). When \(J_{\theta_i \theta_N} = 0\) holds for all \(\theta \in \Theta\), \(\theta_i\) and \(\theta_N\) are said globally orthogonal.

Summarizing the result known in classical statistics, the MSE becomes worse in the presence of nuisance parameters when compared with the case of no nuisance parameters. We can regard the difference of two bounds as the loss of information due to the presence of nuisance parameters. This quantity is defined by

\[
\Delta J_{\theta_i}^{-1} := J_{\theta_i}^{-1} - (J_{\theta_i \Theta})^{-1}.
\] (A.10)

When the values of \(\Delta J_{\theta_i}^{-1}\) is large (in the sense of matrix inequality), the effect of nuisance parameters is more important. From the above mathematical fact, we have that no loss of information is possible if and only if two sets of parameters are globally orthogonal, i.e.,

\[
\Delta J_{\theta_i}^{-1} = 0 \iff J_{\theta_i \theta_N} = 0,
\] (A.11)

for all values of \(\theta = (\theta_i, \theta_N) \in \Theta\).
From the above inequality (A.9), the orthogonality condition is a key ingredient when discussing parameter estimation problems in the presence of nuisance parameters. This was pointed out in the seminal paper by Cox and Reid [7]. The concept of orthogonality condition was also investigated in the framework of information geometry [2].

It is well known that any two sets of parameters can be made orthogonal at each point locally by an appropriate invertible map from a given parameterization to the new parameters as

$$\theta = (\theta_I, \theta_N) \mapsto \xi = (\xi_I, \xi_N) \text{ s.t. } \xi_I = \theta_I.$$  \hfill (A.12)

The condition $\xi_I = (\xi_1, \xi_2, \ldots, \xi_m) = \theta_I$ ensures that the parameters of interest are unchanged while the nuisance parameters can be changed arbitrary. The detailed exposition of this method in the quantum case can be found in [27].

**A.2. Logarithmic derivative operators and quantum Fisher information matrix**

In this appendix, we briefly summarize about the quantum score function and quantum Fisher information [28, 29, 31]. We only consider the symmetric logarithmic (SLD) operators and the SLD Fisher information matrix. For a given smooth family of quantum states $\rho_\theta$ and any (bounded) linear operators $X, Y$, define the symmetric inner product by

$$\langle X, Y \rangle_{\rho_\theta} := \frac{1}{2} \text{tr} (\rho_\theta (YX^\dagger + X^\dagger Y)),$$

where $X^\dagger$ denotes the hermite conjugate of $X$. The $i$th SLD operator, $L_{ij}^S$, is formally defined by the solution to the operator equation:

$$\frac{\partial}{\partial \theta_i} \rho_\theta = \frac{1}{2} (\rho_\theta L_{ij}^S + L_{ij}^S \rho_\theta),$$  \hfill (A.13)

for $i = 1, 2, \ldots, n$. The SLD Fisher information matrix is defined by

$$J_{ij}^S := \langle L_{ij}^S, L_{ij}^S \rangle_{\rho_\theta}. \hfill (A.14)$$

It is convenient to introduce the following linear combinations of the logarithmic derivative operators.

$$L_{ij}^{S_i} := \sum_{j=1}^n (J_{ij}^S)^{-1}_j L_{ij}^S.$$

By definitions, $\{L_{ij}^{S_i}\}$ forms a dual basis for the inner product space $\langle \cdot, \cdot \rangle_{\rho_\theta}; \langle L_{ij}^{Si}, L_{ij}^{Sj} \rangle_{\rho_\theta} = \delta_{ij}$. The inverse of the SLD Fisher information matrix is expressed as

$$(J_{ii}^S)^{-1} = [J_{ij}^{Si}] \quad \text{with} \quad J_{ij}^{Si} = \langle L_{ij}^{Si}, L_{ij}^{Sj} \rangle_{\rho_\theta}.$$  

**Appendix B. Proofs**

**B.1. Proof of lemma 2.1**

First, it is straightforward to see that the transformation:

$$\theta = (\theta_I, \theta_N) \mapsto \xi = (\xi_I, \xi_N) \text{ s.t. } \xi_I = \theta_I.$$  \hfill (B.1)
preserves the first condition of (11). By elementary calculus, we can show that the partial derivatives are transformed as
\[
\frac{\partial}{\partial \xi_i} = \sum_{j=1}^{n} \frac{\partial \theta_j}{\partial \xi_i} \frac{\partial}{\partial \theta_j} = \frac{\partial}{\partial \theta_i} - \sum_{j=m+1}^{n} \frac{\partial \theta_j}{\partial \xi_i} \frac{\partial}{\partial \theta_j} \quad (i = 1, 2, \ldots, m),
\]
\[
\frac{\partial}{\partial \xi_j} = \sum_{i=1}^{n} \frac{\partial \theta_i}{\partial \xi_j} \frac{\partial}{\partial \theta_i} = \frac{\partial}{\partial \theta_j} \quad (j = m + 1, \ldots, n).
\]

The second condition of (11) is verified as follows. Let \( \hat{\theta} = (\Pi, \hat{\theta}) \) be an arbitrary locally unbiased estimator for the parameters of interest at \( \theta \). For \( i, j = 1, 2, \ldots, m \), using (B.1) reads
\[
\frac{\partial}{\partial \xi_i} E_\xi \left[ \hat{\theta}(X) | \Pi \right] = \frac{\partial}{\partial \theta_i} E_{\theta} \left[ \hat{\theta}(X) | \Pi \right] - \sum_{k=m+1}^{n} \frac{\partial \theta_k}{\partial \xi_i} \frac{\partial}{\partial \theta_k} E_{\theta} \left[ \hat{\theta}(X) | \Pi \right]
\]
\[
= \frac{\partial}{\partial \theta_i} E_{\theta} \left[ \hat{\theta}(X) | \Pi \right] = \delta_{i,j}.
\]

The second term in the first line vanishes because of the assumption of locally unbiasedness. For \( i = 1, 2, \ldots, m \) and \( j = m + 1, m + 2, \ldots, n \), we can directly check
\[
\frac{\partial}{\partial \xi_i} E_\xi \left[ \hat{\theta}(X) | \Pi \right] = \frac{\partial}{\partial \theta_i} E_{\theta} \left[ \hat{\theta}(X) | \Pi \right] = 0 = \delta_{i,j}.
\]

Therefore, we prove the relation \( \frac{\partial}{\partial \xi_j} E_{\xi} \left[ \hat{\theta}(X) | \Pi \right] = 0 \) for \( i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, n \).

**B.2. Derivation of expression (15)**

Let us define another bound by
\[
C_{\hat{\theta}}^{\Pi} [W, \mathcal{M}_n] := \min_{\Pi_{\text{POVM}}} \text{Tr} \left\{ W_{\Pi} f_{\hat{\theta}} | \Pi \right\}.
\]
then, we will prove \( C_{\hat{\theta}}^{\Pi} [W, \mathcal{M}_n] = C_{\theta} [W, \mathcal{M}_n] \). The proof here is almost same line of argument as [41]. Using the CR inequality for any locally unbiased estimator for \( \theta_i \), we have
\[
C_{\theta} [W, \mathcal{M}_n] = \min_{\Pi_{\text{1-a for } \theta_i}} \text{Tr} \left\{ W_{\Pi} V_{\theta} \left| \Pi \right| \right\} 
\]
\[
\geq \text{Tr} \left\{ W_{\Pi} f_{\theta} | \Pi \right\} \quad (B.5)
\]
This is true for all POVMs that belong to the set. Therefore, we the relation \( C_{\theta} [W, \mathcal{M}_n] \geq C_{\hat{\theta}}^{\Pi} [W, \mathcal{M}_n] \). To prove the other direction, note that we can always construct a locally unbiased estimator at \( \theta_0 = (\theta_1(0), \ldots, \theta_n(0)) \) for a given POVM \( \Pi \). For example,
\[
\hat{\theta}(X) = \theta_0(0) + \sum_{j=1}^{n} (J_{\theta_0} | \Pi \right|)^{-1} \frac{\partial}{\partial \theta_j} \log \left| p_0(X) | \Pi \right| \bigg|_{\theta_0} \quad (B.6)
\]
Since this estimator is also locally unbiased about the parameter of interest \( \theta_i \), and the MSE matrix about \( \Pi = (\Pi, \hat{\theta}) \) satisfies \( V_{\theta} | \Pi \right| = J_{\theta}^{-1} | \Pi \right| \). In turn, we have a relationship \( V_{\theta} | \Pi \right| = f_{\theta} | \Pi \right| \) for the parameter of interest. Thus, we obtain \( C_{\theta} [W, \mathcal{M}_n] \leq C_{\hat{\theta}}^{\Pi} [W, \mathcal{M}_n] \). This proves \( C_{\theta} [W, \mathcal{M}_n] = C_{\hat{\theta}}^{\Pi} [W, \mathcal{M}_n] \).
B.3. Proof of theorem 2.2

We shall prove the equivalence:

\[ C_\theta [W_I] = C^\theta_\phi [W_I] = C^\lim_\theta [W_I], \]

holds for the three bounds,

\[ C_\theta [W_I] = \min_{\Pi_I: 1.u. \text{ for } \theta} \text{Tr} \left\{ W_I V_\theta [\Pi_I] \right\}, \]
\[ C^\theta_\phi [W_I] = \inf_{\Pi_I: 1.u. \text{ for } \phi} \text{Tr} \left\{ W_I V_\phi [\Pi_I] \right\}, \]
\[ C^\lim_\theta [W_I] = \lim_{W_n \to W_I \Pi_I: 1.u. \text{ for } \theta} \text{Tr} \left\{ W_n V_\theta [\Pi_I] \right\}. \]  

(B.7)

In this section, we omit the dependence of models in the above bounds for simplicity.

We define two classes of POVMs as follows. Denote \( M_\theta \) by the set of POVMs \( \Pi \) whose classical model \( M_\Pi(\Pi) = \{ p_{\theta}(\cdot|\Pi) \} \) is regular, in particular, all score functions

\[ \frac{\partial}{\partial \theta_i} \log p_{\theta}(x|\Pi) \]

\( i = 1, 2, \ldots, m \), (B.8)

are linearly independent, and they span an \( n \)-dimensional tangent space at \( \theta \). Let \( M_\phi \) by the set of POVMs \( \Pi \) in which the linearly independent condition only holds for the parameter of interest, i.e.,

\[ \frac{\partial}{\partial \phi_i} \log p_{\phi}(x|\Pi) \]

\( i = 1, 2, \ldots, m \), (B.9)

are linearly independent. Clearly, the inclusion \( M_\theta \subset M_\phi \subset M \) holds.

First, let us prove the relation \( C^\phi_\phi [W_I] = C_\phi [W_I] \). As mentioned in the main text, if an estimator is locally unbiased for all parameters, it is also locally unbiased for the parameter of interest. Therefore, we have

\[ C^\phi_\phi [W_I] \geq C_\phi [W_I]. \]

Next, let \( \Pi^* = \arg \min_{\Pi \in M_\theta} \text{Tr} \left\{ W_I J_{\theta}(\Pi) \right\} \) be an optimal POVM attaining the minimum. Since the resulting probability distribution \( p_{\theta}(x) = \text{tr}(\rho \Pi^*) \) may not be regular, one cannot construct a locally unbiased estimator about \( \theta = (\theta_I, \theta_N) \) at \( \theta \). In this case, we consider a randomized POVM \( \Pi(\epsilon) \) as follows. Given \( \epsilon \in (0, 1) \), we perform the POVM \( \Pi(\epsilon) \) with a probability \( 1 - \epsilon \) and perform another POVM \( \Pi^0 \in M_\theta \) with a probability \( \epsilon \). The Fisher information matrix about this POVM is

\[ J_{\theta}(\Pi(\epsilon)) = (1 - \epsilon) J_{\theta}(\Pi^*) + \epsilon J_{\theta}(\Pi^0), \]

and this is regular. We can then find an locally unbiased estimator \( \hat{\theta}_I \) for \( \theta \) such that

\[ \text{Tr} \left\{ W_I V_{\theta} [\Pi(\epsilon), \hat{\theta}_I] \right\} = \text{Tr} \left\{ W_I J_{\theta}(\Pi(\epsilon))^{-1} \right\}. \]

Convexity of matrix inverse states

\[ J_{\theta}(\Pi(\epsilon))^{-1} \leq (1 - \epsilon) J_{\theta}(\Pi^*)^{-1} + \epsilon J_{\theta}(\Pi^0)^{-1}. \]
This then shows that
\[ \text{Tr } \left\{ W_f J_0 [\Pi(\epsilon)]^{-1} \right\} \leq (1 - \epsilon) C_{\theta}[W_f] + \epsilon \text{ Tr } \left\{ W_f J_0 [\Pi^0]^{-1} \right\}. \]
Since \( \text{Tr } \left\{ W_f J_0 [\Pi^0]^{-1} \right\} - C_{\theta}[W_f] \geq 0 \) for \( W_f > 0 \), we find that
\[ \text{Tr } \left\{ W_f V_{\theta_f}[\Pi(\epsilon), \hat{\theta}_f] \right\} \leq C_{\theta}[W_f] + \delta, \]
holds for arbitrary \( \delta > 0 \) by an appropriate choice of \( \epsilon \) and \( \Pi^0 \). Therefore, \( C_{\theta}[W_f] + \delta \) can be achieved by a locally unbiased estimator about \( \theta \) to conclude \( C_{\theta_f}^{\epsilon}[W_f] = C_{\theta}[W_f] \).

The other relation \( C_{\theta_f}^{\epsilon}[W_f] = C_{\theta_f}^{\text{lim}}[W_f] \) is proven as follows. Rewrite the bound as \( C_0[W_n] = \min_{\Pi \in \mathcal{M}_\theta} \text{Tr } \{ W_n J_0 [\Pi]^{-1} \} \). Then, we have
\[ C_{\theta_f}^{\text{lim}}[W_f] = \lim_{\epsilon \to 0} C_0[W_n = W_f \oplus \epsilon J_N] \]
\[ = \lim_{\epsilon \to 0} \min_{\Pi \in \mathcal{M}_\theta} \{ \text{Tr } \{ W_f J_0^{\epsilon}[\Pi] \} + \epsilon \text{ Tr } \{ J_0^{\epsilon}[\Pi] \} \}. \quad (B.10) \]
The term \( \text{Tr } \{ J_0^{\epsilon}[\Pi] \} \) is always positive for \( \Pi \in \mathcal{M}_\theta \). Therefore, the following inequality holds.
\[ C_{\theta_f}^{\text{lim}}[W_f] \geq \lim_{\epsilon \to 0} \inf_{\Pi \in \mathcal{M}_\theta} \text{Tr } \{ W_f J_0^{\epsilon}[\Pi] \} \]
\[ = C_{\theta_f}^{\epsilon}[W_f]. \quad (B.11) \]
Finally, we show that \( C_{\theta_f}^{\text{lim}}[W_f] + \delta \) can be achieved by a locally unbiased estimator about \( \theta \) for arbitrary \( \delta > 0 \). But this is clear from that for arbitrary \( \epsilon \), we can always find a POVM \( \Pi(\epsilon, \delta) \) such that \( \text{Tr } \{ W_f J_0^{\epsilon}[\Pi(\epsilon, \delta)] \} = C_{\theta_f}^{\text{lim}}[W_f] + \delta \) holds. Since \( \Pi(\epsilon, \delta) \in \mathcal{M}_\theta \), we can construct a locally unbiased estimator to conclude \( C_{\theta_f}^{\epsilon}[W_f] = C_{\theta_f}^{\text{lim}}[W_f] \).

Appendix C. Classical CR inequality from a weight matrix optimization

In this section, we give the proof for proposition 2.3. This shows that the classical CR inequality (A.3) in the presence of nuisance parameters can be derived from the proposed weight-matrix elimination method. Here, we consider the optimization problem (27). Using (29), the goal is to show
\[ \min_{W_0, W_n = W_m, W_0 \geq 0, W_f > 0} \left[ \text{Tr } \{ W_n V_{\theta_f} \} - \text{Tr } \{ W_n J_0 [\Pi]^{-1} \} \right] \geq 0 \quad (C.1) \]
\[ \Rightarrow V_{\theta_f} \geq J_0^{\epsilon}[\Pi]. \]

The following lemma is well-known fact in matrix analysis (see, for example, theorem 1.3.3 of [65]).

**Lemma appendix C.1.** For a \( 2 \times 2 \) block matrix on the complex number,
\[ M = \begin{pmatrix} A & B \\ B^\dagger & D \end{pmatrix}, \quad (C.2) \]
where \( A \in \mathbb{C}^{m \times m}, B \in \mathbb{C}^{m \times n}, \) and \( D \in \mathbb{C}^{n \times n}. \) Suppose \( A > 0 \) and \( D > 0, \) then \( M \geq 0 \) holds if and only if \( A \geq BD^{-1}B^\dagger. \)
It is easy to verify that we can exchange the role of $A$ and $D$. We can also extend the above lemma for the case: $A > 0$ and $D > 0$. With this observation, we can rewrite the constraint for the minimization problem as $W_N, W_IN = W_N^{1/2}: W_N \geq W_N^{-1/2} W_IN$. Denote the difference of two matrices and its block-matrix representation according to the partition $\theta = (\theta_I, \theta_N)$ as

$$M_\theta := V_\theta[\hat{\theta}] - J_\theta^{-1},$$

$$M_\theta = \begin{pmatrix} M_H & M_{IN} \\ M_{NI} & M_{NN} \end{pmatrix}, \quad M_\theta^{-1} = \begin{pmatrix} M^H & M^{IN} \\ M^{NI} & M^{NN} \end{pmatrix}.$$  \hfill (C.3)

Noting a matrix inequality $W_I \geq W_J$ implies $\text{Tr}\{V W_I\} \geq \text{Tr}\{V W_J\}$ for $V \geq 0$, we have

$$\min_{W_N \geq 0, W_I > 0} \text{Tr}\{W_N M_\theta\} \geq \min_{W_N \geq W_N^{-1/2}} \text{Tr}\{W_I M_H\} + \text{Tr}\{W_IN M_{NN}\}$$

$$= \min_{W_N \geq W_N^{-1/2}} \text{Tr}\{W_I M_H\} + \text{Tr}\{W_N W_I^{-1/2} M_{IN} W_I^{1/2}\} \times (M_{NN}^{-1/2} W_N W_I^{-1/2} M_{NN} W_I^{1/2})^T$$

$$- \text{Tr}\{(M_{NN}^{-1/2} W_N W_I^{-1/2} M_{NI} W_I^{1/2})^T\}$$

$$\geq \text{Tr}\{W_I M_H\} - \text{Tr}\{W_I M_{IN} M_{NN}^{-1/2} M_{MI}\}$$

$$= \text{Tr}\{W_I (M_H - M_{IN} M_{NN}^{-1} M_{MI})\}. \hfill (C.4)$$

Two inequalities can be saturated by the following choice of the weight matrix:

$$W_{IN}^* = -M_{NN}^{-1/2} M_{NI} W_I = (W_{NI}^*)^T,$$

$$W_N^* = W_N^{-1/2} W_{IN}^*.$$  \hfill (C.5)

Summarizing above argument, we prove

$$\min_{W_N, W_{IN} = W_{IN}^*, W_N \geq 0, W_I > 0} \text{Tr}\{W_N M_\theta\} = \text{Tr}\{W_I (M_H - M_{IN} M_{NN}^{-1} M_{MI})\}. \hfill (C.6)$$

The quantity in the right-hand side is nonnegative, and thus we obtain

$$\text{Tr}\{W_I M_H\} \geq \text{Tr}\{M_{IN} M_{NN}^{-1} M_{MI}\} \geq 0 \iff \text{Tr}\{W_I V_\theta\} \geq \text{Tr}\{W_I J_{\theta_I}\}. \hfill (C.7)$$

Since this inequality holds for all positive matrices $W_I > 0$, this is equivalent to the matrix inequality $V_\theta \geq J_{\theta_I}$. This proves our claim. \hfill \Box

We note that the above result immediately derives a tradeoff relationship among estimation errors in the classical parameter estimation problem. Expression (C.6) shows that the matrix inequality

$$V_\theta[\hat{\theta}] \geq J_{\theta_I} + (V_\theta[\hat{\theta}_N] - J_{\theta_I}) (V_{\theta_N} - J_{\theta_N})^{-1} (V_\theta[\hat{\theta}_N] - J_{\theta_N}), \hfill (C.8)$$

holds for any locally unbiased estimator $\hat{\theta}$ at $\theta = (\theta_I, \theta_N)$. In this inequality, the second term of the right-hand side, which is nonnegative, represents the tradeoff relationship. We can see that it
vanishes if $V_{\theta_\theta}^{\hat{\theta}} - J_{\theta_\theta}^{-1} = 0$ or the MSE matrix for the nuisance parameters $V_{\theta_\theta}$ diverges. Lastly, we comment that this result itself is not surprising at all. Since we are imposing the locally unbiasedness condition for all parameters, We have the usual CR inequality $M_\theta[M_{\theta}\theta_\theta]^{-1} \geq 0$. Using lemma appendix C.1, this is equivalent to $M_\theta - MN M_{\theta_\theta}^{-1} M_{\theta_\theta} \geq 0$. This is exactly the same statement as (C.8).

We point out that our weight-matrix elimination method can also provide the CR bound (A.2) without any nuisance parameter. Consider an $n$-parameter model and we wish to minimize the weighted trace of the inverse of the Fisher information matrix. This is expressed as $\min W_{\theta_\theta}^{-1}$ where the minimization is about $W_{\theta_\theta} = W_{\theta_\theta}^T$ and $W_N$ under the constraint $W_{\theta_\theta} \geq I$ and $W_N > 0$. Working the exactly same procedure, we get

$$\min \text{Tr} \{ W_{\theta_\theta}^{-1} \} = \text{Tr} \{ W_{\theta_\theta}^{-1} J_{\theta_\theta}^{-1} \} = \text{Tr} \{ W_{\theta_\theta}^{-1} J_{\theta_\theta}^{-1} \} \geq 0. \quad \text{(C.9)}$$

Here the second line is due to the Schur’s complement. Therefore, the CR inequality $V_{\theta_\theta}^{\hat{\theta}} \geq J_{\theta_\theta}^{-1}$ holds by considering the CR type bound among all possible weight matrices for the nuisance parameters.

ORCID iDs

Jun Suzuki © https://orcid.org/0000-0003-1975-6003

References

[1] Fisher R A 1935 J. Roy. Stat. Soc. 98 39
[2] Amari S 1985 Differential-Geometrical Methods in Statistics (Berlin: Springer)
[3] Lehmann E L and Casella G 1998 Theory of Point Estimation 2nd edn (Berlin: Springer)
[4] Barndorff-Nielsen O E and Cox D R 1994 Inference and Asymptotics (Boca Raton, FL: CRC)
[5] Amari S and Nagaoka H 2000 Methods of Information Geometry (Translations of Mathematical Monograph, vol 191) (Providence, RI: American Mathematical Society)
[6] Basu D 1977 J. Am. Stat. Assoc. 72 355
[7] Cox D R and Reid N 1987 J. Roy. Stat. Soc. B 49 1
[8] Amari S and Kumon M 1988 Ann. Stat. 16 1044
[9] Bhapkar V P and Srinivasan C 1994 Ann. Inst. Statist. Math. 46 593
[10] Zhu Y and Reid N 1994 Can. J. Stat. 22 111
[11] Giovannetti V, Lloyd S and Maccone L 2011 Nat. Photon. 5 222
[12] Escher B M, de Matos Filho R L and Davidovich L 2011 Nat. Phys. 7 406
[13] Demkowicz-Dobrzański R, Kołodyński J and Guţă M 2012 Nat. Commun. 3 1063
[14] Tóth G and Apellaniz I 2014 J. Phys. A: Math. Theor. 47 424006
[15] Demkowicz-Dobrzański R, Jarzyna M and Kołodyński J 2015 Prog. Opt. 60 345
[16] Pezzé L and Smerzi A 2014 Atom Interferometry ed GMTino and MA Kasevich (Amsterdam: IOS Press) p 691
[17] Szczylukska M, Baumgratz T and Datta A 2016 Adv. Phys. X 1 621
[18] Degen C L, Reinhard F and Cappellaro P 2017 Rev. Mod. Phys. 89 035002
[19] Crowley P J D, Datta A, Barbieri M and Walmsley I A 2014 Phys. Rev. A 89 023845
[20] Vidrighin M D, Donati G, Genoni M G, Jin X-M, Kolthammer W S, Kim M S, Datta A, Barbieri M and Walmsley I A 2014 Nat. Commun. 5 3532
[21] Suzuki J 2015 Int. J. Quant. Inf. 13 1450044
[22] Kok P, Dunningham J and Ralph J F 2017 Phys. Rev. A 95 012326
[23] Alenburg S, Oszmaniec M, Wolk S and Gühne O 2017 Phys. Rev. A 96 042319
[24] Apellaniz I, Urizar-Lanz I, Zumbrunen Z, Hyllus P and Tóth G 2018 Phys. Rev. A 97 053603

41
