Attaining the Ultimate Precision Limit in Quantum State Estimation

Yuxiang Yang[1,2,5]**, Giulio Chiribella**[1,3,4,5]**, Masahito Hayashi**[6,7,8]

[1] Department of Computer Science, The University of Hong Kong, Pokfulam Road, Pok Fu Lam, Hong Kong
[2] Institute for Theoretical Physics, ETH Zürich, Zurich, Switzerland. E-mail: yangyu@phys.ethz.ch
[3] Department of Computer Science, Parks Road, Oxford OX1 3QD, UK. E-mail: giulio.chiribella@cs.ox.ac.uk
[4] Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada
[5] The University of Hong Kong Shenzhen Institute of Research and Innovation, 5/F, Key Laboratory Platform Building, No. 6, Yuexing 2nd Rd., Nanshan, Shenzhen 518057, China
[6] Graduate School of Mathematics, Nagoya University, Furocho, Chikusaku, Nagoya 464-860, Japan. E-mail: masahito@math.nagoya-u.ac.jp; cqtmh@nus.edu.sg
[7] Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117542, Singapore
[8] Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Abstract: We derive a bound on the precision of state estimation for finite dimensional quantum systems and prove its attainability in the generic case where the spectrum is non-degenerate. Our results hold under an assumption called local asymptotic covariance, which is weaker than unbiasedness or local unbiasedness. The derivation is based on an analysis of the limiting distribution of the estimator's deviation from the true value of the parameter, and takes advantage of quantum local asymptotic normality, a useful asymptotic characterization of identically prepared states in terms of Gaussian states. We first prove our results for the mean square error of a special class of models, called *D*-invariant, and then extend the results to arbitrary models, generic cost functions, and global state estimation, where the unknown parameter is not restricted to a local neighbourhood of the true value. The extension includes a treatment of nuisance parameters, i.e. parameters that are not of interest to the experimenter but nevertheless affect the precision of the estimation. As an illustration of the general approach, we provide the optimal estimation strategies for the joint measurement of two qubit observables, for the estimation of qubit states in the presence of amplitude damping noise, and for noisy multiphase estimation.

1. Introduction

Quantum estimation theory is one of the pillars of quantum information science, with a wide range of applications from evaluating the performance of quantum devices [1,2] to exploring the foundation of physics [3,4]. In the typical scenario, the problem is specified by a parametric family of quantum states, called the *model*, and the objective is to design measurement strategies that estimate the parameters of interest with the highest possible precision. The precision measure is often chosen to be the mean square error (MSE), and is lower bounded through generalizations of the Cramér–Rao bound of classical statistics [5,6]. Given n copies of a quantum state, such generalizations imply that the product MSE $\cdot n$ converges to a positive constant in the large n limit.

Despite many efforts made over the years (see, e.g., [5–12] and [13] for a review), the attainability of the precision bounds of quantum state estimation has only been proven in a few special cases. Consider, as an example, the most widely used bound, namely the symmetric logarithmic derivative Fisher information bound (SLD bound, for short). The SLD bound is tight in the one-parameter case [5,6], but is generally non-tight in multiparameter estimation. Intuitively, measuring one parameter may affect the precision in the measurement of another parameter, and thus it is extremely tricky to construct the optimal measurement. Another bound for multiparameter estimation is the right logarithmic derivative Fisher information bound (RLD bound, in short) [5]. Its achievability was shown in the Gaussian states case [5], the qubits case [14,15], and the qudits case [16,17]. In this sense, the RLD bound is superior to the SLD bound. However, the RLD bound holds only when the family of states to be estimated satisfies an *ad hoc* mathematical condition. The most general quantum extension of the classical Cramér–Rao bound till now is the Holevo bound [5], which gives the maximum among all existing lower bounds for the error of unbiased measurements for the estimation of any family of states. The attainability of the Holevo bound was studied in the pure states case [10] and the qubit case [14,15], and was conjectured to be generic by one of us [18]. Yamagata et al. [19] addressed the attainability question in a local scenario, showing that the Holevo bound can be attained under certain regularity conditions. However, the attaining estimator constructed therein depends on the true parameter, and therefore has limited practical interest. Meanwhile, the need of a general, attainable bound on multiparameter quantum estimation is increasing, as more and more applications are being investigated [20–24].

In this work we explore a new route to the study of precision limits in quantum estimation. This new route allows us to prove the asymptotic attainability of the Holevo bound in generic scenarios, to extend its validity to a broader class of estimators, and to derive a new set of attainable precision bounds. We adopt the condition of *local asymptotic covariance* [18] which is less restrictive than the unbiasedness condition [5] assumed in the derivation of the Holevo bound. Under local asymptotic covariance, we characterize the MSE of the *limiting distribution*, namely the distribution of the estimator's rescaled deviation from the true value of the parameter in the asymptotic limit of $n \to \infty$.

Our contribution can be divided into two parts, the attainability of the Holevo bound and the proof that the Holevo bound still holds under the weaker condition of local asymptotic covariance. To show the achievability part, we employ *quantum local asymptotic normality* (Q-LAN), a useful characterization of n-copy d-dimensional (qudit) states in terms of multimode Gaussian states. The qubit case was derived in [14,15] and the case of full parametric models was derived by Kahn and Guta when the state has non-degenerate spectrum [16,17]. Here we extend this characterization to a larger class of models, called *D-invariant models*, using a technique of symplectic diagonalization. For models that are not *D*-invariant, we derive an achievable bound, expressed in terms of a quantum Fisher information-like quantity that can be straightforwardly evaluated. Whenever the model consists of qudit states with non-degenerate spectrum, this quantity turns out to be equal to the quantity in the Holevo bound [5]. Our evaluation has compact uniformity and order estimation of the convergence, which will allow us to prove the achievability of the bound even in the global setting.

We stress that, until now, the most general proof of the Holevo bound required the condition of local unbiasedness. In particular, no previous study showed the validity of the Holevo bound under the weaker condition of local asymptotic covariance in the

multiparameter scenario. To avoid employing the (local) unbiasedness condition, we focus on the discretized version of the RLD Fisher information matrix, introduced by Tsuda and Matsumoto [25]. Using this version of the RLD Fisher information matrix, we manage to handle the local asymptotic covariance condition and to show the validity of the Holevo bound in this broader scenario. Remarkably, the validity of the bound does not require finite-dimensionality of the system or non-degeneracy of the states in the model. Our result also provides a simpler way of evaluating the Holevo bound, whose original expression involved a difficult optimization over a set of operators.

The advantage of local asymptotic covariance over local unbiasedness is the following. For practical applications, the estimator needs to attain the lower bound globally, i.e., at all points in the parameter set. However, it is quite difficult to meet this desideratum under the condition of local unbiasedness, even if we employ a two-step method based on a first rough estimate of the state, followed by the measurement that is optimal in the neighbourhood of the estimate. In this paper, we construct a locally asymptotic covariant estimator that achieves the Holevo bound at every point, for any qudit submodel except those with degenerate states. Our construction proceeds in two steps. In the first step, we perform a full tomography of the state, using the protocol proposed in [26]. In the second step, we implement a locally optimal estimator based on Q-LAN [16,17]. The two-step estimator works even when the estimated parameter is not assumed to be in a local neighbourhood of the true value. The key tool to prove this property is our precise evaluation of the optimal local estimator with compact uniformity and order estimation of the convergence. Our method can be extended from the MSE to arbitrary cost functions. A comparison between the approach adopted in this work (in green) and conventional approaches to quantum state estimation (in blue) can be found in Fig. 1.

Besides the attainability of the Holevo bound, the method can be used to derive a broad class of bounds for quantum state estimation. Under suitable assumptions, we characterize the tail of the limiting distribution, providing a bound on the probability that the estimate falls out of a confidence region. The limiting distribution is a good approximation of the (actual) probability distribution of the estimator, up to a term vanishing in n. Then, we derive a bound for quantum estimation with *nuisance parameters*, *i.e.* parameters that are not of interest to the experimenter but may affect the estimation of the other parameters. For instance, the strength of noise in a phase estimation scenario can be regarded as a nuisance parameter. Our bound applies also to arbitrary estimation models, thus extending nuisance parameter bounds derived for specific cases (see, e.g., [27–29]). In the final part of the paper, the above bounds are illustrated in concrete examples, including the joint measurement of two qubit observables, the estimation of qubit states in the presence of amplitude damping noise, and noisy multiphase estimation.

The remainder of the paper is structured as follows. In Sect. 2 we introduce the main ideas in the one-parameter case. Our discussion of the one-parameter case requires no regularity condition for the parametric model. Then we devote several sections to introducing and deriving tools for the multiparameter estimation. In Sect. 3, we briefly review the Holevo bound and Gaussian states, and derive some relations that will be useful in the rest of the paper. In Sect. 4, we introduce Q-LAN. In Sect. 5 we introduce the ϵ-difference RLD Fisher information matrix, which will be a key tool for deriving our bounds in the multiparameter case. In Sect. 6, we derive the general bound on the precision of multiparameter estimation. In Sect. 7, we address state estimation in the presence of nuisance parameters and derive a precision bound for this scenario. Section 8 provides bounds on the tail probability. In Sect. 9, we extend our results
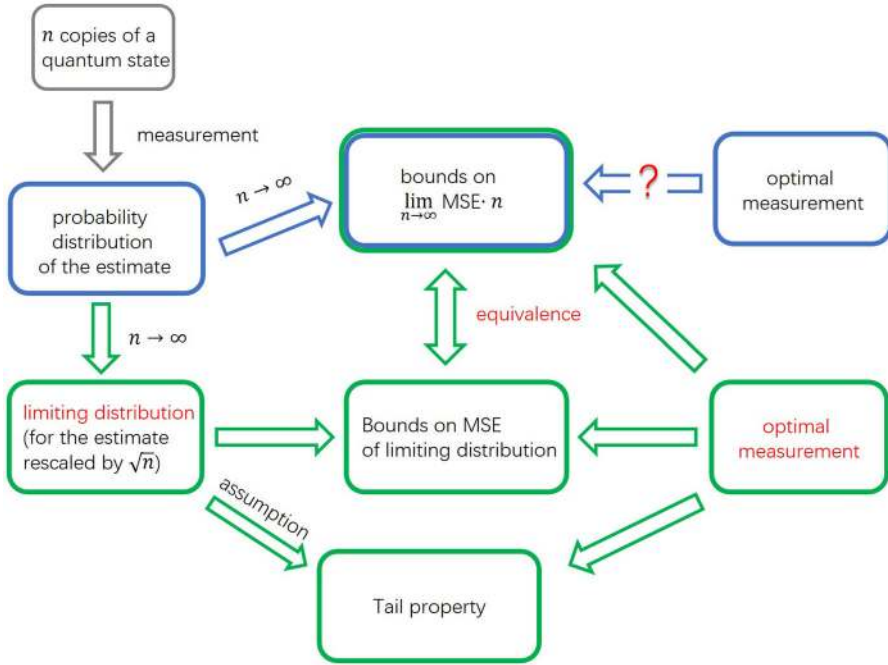
Fig. 1. Comparison between the approach of this work (in green) and the traditional approach of quantum state estimation (in blue). In the traditional approach, one derives precision bounds based on the probability distribution function (PDF) for measurements on the original set of quantum states. The bounds are evaluated in the large n limit and the task is to find a sequence of measurements that achieves the limit bound. In this work, we first characterize the limiting distribution and then work out a bound in terms of the limiting distribution. This construction also provides the optimal measurement in the limiting scenario, which can be used to prove the asymptotic attainability of the bound. The analysis of the limiting distribution also provides tail bounds, which approximate the tail bounds for finite n up to a small correction, under the assumption that the cost function and the model satisfy a certain relation (see Theorem 9)

Table 1. Table of terms and notations

Term	Notation	Definition
Limiting distribution	$\wp_{t_0, t \mid m}$	Eqs. (21) and (101)
SLD quantum Fisher information at t_0	J_{t_0}	Eq. (35)
RLD quantum Fisher information at t_0	$\tilde{J}_{t_0}$	Eq. (42) (for D-invariant models)
D-matrix at t_0	D_{t_0}	Eq. (41)
Gaussian state	$G[\boldsymbol{\alpha}, \boldsymbol{\gamma}]$	Eq. (55)
(Multi-mode) displaced thermal state	$\Phi[(\boldsymbol{\alpha}^R, \boldsymbol{\alpha}^I), \boldsymbol{\beta}]$	Eq. (54)
Gaussian shift operator	$T_{\boldsymbol{\alpha}}$	Eq. (59)
2m-dimensional symplectic form	Ω_m	Eq. (57)
A 2m-dimensional diagonal matrix	$E_m(\boldsymbol{x})$	Eq. (56)
Covariance matrix of a probability distribution $\wp$	$V[\wp]$	Eq. (108)

to global estimation and to generic cost functions. In Sect. 10, the general method is illustrated through examples. The conclusions are drawn in Sect. 11.

Remark on the notation In this paper, we use z^* for the complex conjugate of $z \in \mathbb{C}$ and $A^\dagger$ for the Hermitian conjugate of an operator A. For convenience of the reader, we list other frequently appearing notations and their definitions in Table 1.

2. Precision Bound Under Local Asymptotic Covariance: One-Parameter Case

In this section, we discuss estimation of a single parameter under the local asymptotic covariance condition, without any assumption on the parametric model.

2.1. Cramér–Rao inequality without regularity assumptions. Consider a one-parameter model $\mathcal{M}$, of the form

$$\mathcal{M} = \{\rho_t\}_{t \in \Theta} \tag{1}$$

where Θ is a subset of $\mathbb{R}$. In the literature it is typically assumed that the parametrization is differentiable. When this is the case, one can define the symmetric logarithmic derivative operator (SLD in short) at t_0 via the equation

$$\frac{d\rho_{t_0}}{dt_0} = \frac{1}{2}\left(\rho_{t_0} L_{t_0} + L_{t_0}\rho_{t_0}\right). \tag{2}$$

Then, the SLD Fisher information is defined as

$$J_{t_0} := \mathrm{Tr}\left[\rho_{t_0} L_{t_0}^2\right]. \tag{3}$$

The SLD L_{t_0} is not unique in general, but the SLD Fisher information J_{t_0} is uniquely defined because it does not depend on the choice of the SLD L_{t_0} among the operators satisfying (2). When the parametrization is C^1-continuous and $\epsilon > 0$ is a small number, one has

$$F\left(\rho_{t_0} || \rho_{t_0+\epsilon}\right) = 1 - \frac{J_{t_0}}{8}\epsilon^2 + o(\epsilon^2), \tag{4}$$

where

$$F(\rho \| \rho') := \mathrm{Tr}\, |\sqrt{\rho}\sqrt{\rho'}| \tag{5}$$

is the fidelity between two density matrices ρ and ρ'. It is called Bhattacharya or Hellinger coefficient in the classical case [30,31].

Here we do not assume that the parametrization (1) is differentiable. Hence, the SLD Fisher information cannot be defined by (3). Instead, following the intuition of (4), we define the SLD Fisher information J_{t_0} as the limit

$$J_{t_0} := \liminf_{\epsilon \to 0} \frac{8\left[1 - F\left(\rho_{t_0} || \rho_{t_0+\epsilon}\right)\right]}{\epsilon^2}. \tag{6}$$

In the n-copy case, we have the following lemma:

Lemma 1.

$$\liminf_{n \to \infty} \frac{8\left(1 - F\left(\rho_{t_0}^{\otimes n} || \rho_{t_0+\frac{\epsilon}{\sqrt{n}}}^{\otimes n}\right)\right)}{\epsilon^2} = \frac{8\left(1 - e^{-\frac{J_{t_0}\epsilon^2}{8}}\right)}{\epsilon^2}. \tag{7}$$

Proof. Using the definition (6), we have

$$\liminf_{n\to\infty} \frac{8\left(1 - F\left(\rho_{t_0}^{\otimes n} || \rho_{t_0 + \frac{\epsilon}{\sqrt{n}}}^{\otimes n}\right)\right)}{\epsilon^2}$$

$$= \liminf_{n\to\infty} \frac{8\left(1 - \left(1 - \frac{J_{t_0}\epsilon^2}{8n} + o(\frac{\epsilon^2}{n})\right)^n\right)}{\epsilon^2} = \frac{8\left(1 - e^{-\frac{J_{t_0}\epsilon^2}{8}}\right)}{\epsilon^2}. \tag{8}$$

In other words, the SLD Fisher information is constant over n if we replace ϵ by $\epsilon/\sqrt{n}$.

To estimate the parameter $t \in \Theta$, we perform on the input state a quantum measurement, which is mathematically described by a *positive operator valued measure (POVM)* with outcomes in $\mathcal{X} \subset \mathbb{R}$. An outcome x is then mapped to an estimate of t by an estimator $\hat{t}(x)$. It is often assumed that the measurement is unbiased, in the following sense: a POVM M on a single input copy is called *unbiased* when

$$\int_{\mathcal{X}} \hat{t}(x) \, \text{Tr}\left[\rho_t \, M(dx)\right] = t, \qquad \forall t \in \Theta. \tag{9}$$

For a POVM M, we define the mean square error (MSE) $V_t(M)$ as

$$V_t(M) := \int_{\mathcal{X}} \left(\hat{t}(x) - t\right)^2 \text{Tr}\left[\rho_t M(dx)\right]. \tag{10}$$

Then, we have the fidelity version of the Cramér–Rao inequality:

Theorem 1. *For an unbiased measurement M satisfying*

$$\int_{\mathcal{X}} \hat{t}(x) P_t(dx) = t \tag{11}$$

for any t, we have

$$\frac{1}{2}\left[V_{t_0}(M) + V_{t_0+\epsilon}(M) + \epsilon^2\right] \geq \frac{\epsilon^2}{8\left[1 - F(\rho_{t_0} || \rho_{t_0+\epsilon})\right]}. \tag{12}$$

When $\lim_{\epsilon\to 0} V_{t_0+\epsilon}(M) = V_{t_0}(M)$, taking the limit $\epsilon \to 0$, we have

$$V_{t_0}(M) \geq J_{t_0}^{-1}. \tag{13}$$

The proof uses the notion of fidelity between two classical probability distributions: for two given distributions P and Q on a probability space $\mathcal{X}$, we define the fidelity $F(P\|Q)$ as follows. Let f_P and f_Q be the Radon–Nikodým derivatives of P and Q with respect to $P + Q$, respectively. Then, the fidelity $F(P\|Q)$ can be defined as

$$F(P\|Q) := \int_{\mathcal{X}} \sqrt{f_P(x)} \sqrt{f_Q(x)} (P + Q)(dx). \tag{14}$$

With the above definition, the fidelity satisfies an information processing inequality: for every classical channel $\mathcal{G}$, one has $F(\mathcal{G}(P)\|\mathcal{G}(Q)) \geq F(P\|Q)$. For a family of probability distributions $\{P_\theta\}_{\theta\in\Theta}$, we define the Fisher information as

$$J_{t_0} := \liminf_{\epsilon\to 0} \frac{8\left[1 - F\left(P_{t_0} || P_{t_0+\epsilon}\right)\right]}{\epsilon^2}. \tag{15}$$

When the probability distributions are over a discrete set, their Fisher information coincides with the quantum SLD of the corresponding diagonal matrices.

Proof of Theorem 1. Without loss of generality, we assume $t_0 = 0$. We define the probability distribution P_t by $P_t(B) := \text{Tr}\,[\,\rho_t M(B)\,]$. Then, the information processing inequality of the fidelity [32] yields the bound $F(\rho_{t_0}\|\rho_{t_0+\epsilon}) \leq F(P_0\|P_\epsilon)$. Hence, it is sufficient to show (12) for the probability distribution family $\{P_t\}$.

Let f_0 and f_ϵ be the Radon–Nikodým derivatives of P_0 and P_ϵ with respect to P_0+P_ϵ. Denoting the estimate by $\hat{t}$, we have

$$V_\epsilon(\hat{t}) = \int_{\mathcal{X}} (\hat{t}(x) - \epsilon)^2 P_\epsilon(dx) = \int_{\mathcal{X}} \hat{t}(x)^2 P_\epsilon(dx) - 2\epsilon \int_{\mathcal{X}} \hat{t}(x) P_\epsilon(dx) + \epsilon^2$$

$$= \int_{\mathcal{X}} \hat{t}(x)^2 P_\epsilon(dx) - 2\epsilon^2 + \epsilon^2 = \int_{\mathcal{X}} \hat{t}(x)^2 P_\epsilon(dx) - \epsilon^2,$$

and therefore

$$2V_0(\hat{t}) + 2V_\epsilon(\hat{t}) + 2\epsilon^2 = \int_{\mathcal{X}} 2\hat{t}(x)^2 (P_0(dx) + P_\epsilon(dx))$$

$$= \int_{\mathcal{X}} 2\hat{t}(x)^2 (f_0(x) + f_\epsilon(x))(P_0 + P_\epsilon)(dx)$$

$$\geq \int_{\mathcal{X}} \hat{t}(x)^2 (\sqrt{f_0(x)} + \sqrt{f_\epsilon(x)})^2 (P_0 + P_\epsilon)(dx). \qquad (16)$$

Also, (14) implies the relation

$$\int_{\mathcal{X}} (\sqrt{f_0(x)} - \sqrt{f_\epsilon(x)})^2 (P_0 + P_\epsilon)(dx) = 2 - 2F(P_0\|P_\epsilon). \qquad (17)$$

Hence, Schwartz inequality implies

$$\int_{\mathcal{X}} \hat{t}(x)^2 (\sqrt{f_0(x)} + \sqrt{f_\epsilon(x)})^2 (P_0 + P_\epsilon)(dx) \cdot \int_{\mathcal{X}} (\sqrt{f_0(x)} - \sqrt{f_\epsilon(x)})^2 (P_0 + P_\epsilon)(dx)$$

$$\geq \left(\int_{\mathcal{X}} \hat{t}(x)(\sqrt{f_0(x)} + \sqrt{f_\epsilon(x)})(\sqrt{f_0(x)} - \sqrt{f_\epsilon(x)})(P_0 + P_\epsilon)(dx) \right)^2$$

$$= \left(\int_{\mathcal{X}} \hat{t}(x)(f_0(x) - f_\epsilon(x))(P_0 + P_\epsilon)(dx) \right)^2$$

$$= \left(\int_{\mathcal{X}} \hat{t}(x)(P_0 - P_\epsilon)(dx) \right)^2 = \epsilon^2. \qquad (18)$$

Combining (16), (17), and (18) we have (12). $\quad\square$

2.2. *Local asymptotic covariance.* When many copies of the state ρ_t are available, the estimation of t can be reduced to a local neighbourhood of a fixed point $t_0 \in \Theta$. Motivated by Lemma 1, we adopt the following parametrization of the n-copy state

$$\rho_{t_0,t}^n := \rho_{t_0+\frac{t}{\sqrt{n}}}^{\otimes n}, \qquad t \in \Delta_n := \sqrt{n}\,(\Theta - t_0), \qquad (19)$$

having used the notation $a\,\Theta+b := \{ax+b\,|x \in \Theta\}$, for two arbitrary constants $a, b \in \mathbb{R}$. With this parametrization, the local n-copy model is $\left\{\rho_{t_0,t}^n\right\}_{t \in \Delta_n}$. Note that, for every $t \in \mathbb{R}$, there exists a sufficiently large n such that Δ_n contains t. As a consequence, one has $\bigcup_{n \in \mathbb{N}} \Delta_n = \mathbb{R}$.

Assuming t_0 to be known, the task is to estimate the local parameter $t \in \mathbb{R}$, by performing a measurement on the n-copy state $\rho_{t_0,t}^n$ and then mapping the obtained data to an estimate $\hat{t}_n$. The whole estimation strategy can be described by a sequence of POVMs $\mathfrak{m} := \{M_n\}$. For every Borel set $\mathsf{B} \subset \mathbb{R}$, we adopt the standard notation

$$M_n(\mathsf{B}) := \int_{\hat{t}_n \in \mathsf{B}} M_n(d\hat{t}_n).$$

In the existing works on quantum state estimation, the error criterion is defined in terms of the difference between the global estimate $t_0 + \frac{\hat{t}_n}{\sqrt{n}}$ and the global true value $t_0 + \frac{t}{\sqrt{n}}$. Instead, here we focus on the difference between the local estimate $\hat{t}_n$ and the true value of the local parameter t. With this aim in mind, we consider the probability distribution

$$\wp_{t_0,t|M_n}^n(\mathsf{B}) := \mathrm{Tr}\left[\rho_{t_0,t}^n M_n\left(\frac{\mathsf{B}}{\sqrt{n}} + t_0\right)\right]. \tag{20}$$

We focus on the behavior of $\wp_{t_0,t|M_n}^n$ in the large n limit, assuming the following condition:

Condition 1 (Local asymptotic covariance for a single-parameter). *A sequence of measurements $\mathfrak{m} = \{M_n\}$ satisfies local asymptotic covariance[1] when*

1. *The distribution $\wp_{t_0,t|M_n}^n$ (20) converges to a distribution $\wp_{t_0,t|\mathfrak{m}}$, called the limiting distribution, namely*

$$\wp_{t_0,t|\mathfrak{m}}(\mathsf{B}) := \lim_{n\to\infty} \wp_{t_0,t|M_n}^n(\mathsf{B}) \tag{21}$$

   for any Borel set B.
2. *the limiting distribution satisfies the relation*

$$\wp_{t_0,t|\mathfrak{m}}(\mathsf{B} + t) = \wp_{t_0,0|\mathfrak{m}}(\mathsf{B}) \tag{22}$$

   for any $t \in \mathbb{R}$, which is equivalent to the condition

$$\lim_{n\to\infty} \mathrm{Tr}\left[\rho_{t_0,t}^n M_n\left(\frac{\mathsf{B}}{\sqrt{n}} + t_0 + \frac{t}{\sqrt{n}}\right)\right] = \lim_{n\to\infty} \mathrm{Tr}\left[\rho_{t_0,0}^n M_n\left(\frac{\mathsf{B}}{\sqrt{n}} + t_0\right)\right]. \tag{23}$$

Using the limiting distribution, we can faithfully approximate the tail probability as

$$\mathbf{Prob}\left\{\left|\hat{t}_n - t\right| > \epsilon\right\} = \wp_{t_0,t|\mathfrak{m}}((-\infty, -\epsilon) \cup (\epsilon, \infty)) + \epsilon_n \tag{24}$$

where the ϵ_n term vanishes with n for every fixed ϵ.

For convenience, one may be tempted to require the existence of a probability density function (PDF) of the limiting distribution $\wp_{t_0,t|\mathfrak{m}}$. However, the existence of a PDF is already guaranteed by the following lemma.

Lemma 2. *When a sequence $\mathfrak{m} := \{M_n\}$ of POVMs satisfies local asymptotic covariance, the limiting distribution $\wp_{t_0,t|\mathfrak{m}}$ admits a PDF, denoted by $\wp_{t_0,0|\mathfrak{m},d}$.*

The proof is provided in "Appendix A".

---

[1] The counterpart of this condition in classical statistics is known as asymptotic equivalent-in-law or regular. See, for instance, page 115 of [33].

2.3. MSE bound for the limiting distribution. As a figure of merit, we focus on the mean square error (MSE) $V[\wp_{t_0,t|\mathrm{m}}]$ of the limiting distribution $\wp_{t_0,t|\mathrm{m}}$, namely

$$V[\wp_{t_0,t|\mathrm{m}}] := \int_{-\infty}^{\infty} (\hat{t} - t)^2 \, \wp_{t_0,t|\mathrm{m}} \left(d\hat{t} \right) .$$

Note that local asymptotic covariance implies that the MSE is independent of t.

The main result of the section is the following theorem:

Theorem 2 (MSE bound for single-parameter estimation). *When a sequence $\mathrm{m} := \{M_n\}$ of POVMs satisfies local asymptotic covariance, the MSE of its limiting distribution is lower bounded as*

$$V[\wp_{t_0,t|\mathrm{m}}] \geq J_{t_0}^{-1} , \tag{25}$$

where J_{t_0} is the SLD Fisher information of the model $\{\rho_t\}_{t \in \Theta}$. The PDF of $\wp_{t_0,t|\mathrm{m}}$ is upper bounded by $\sqrt{J_{t_0}}$. When the PDF of $\wp_{t_0,t|\mathrm{m}}$ is differentiable with respect to t, equality in (25) holds if and only if $\wp_{t_0,t|\mathrm{m}}$ is the normal distribution with average zero and variance $J_{t_0}^{-1}$.

Proof of Theorem 2. When the integral $\int_{\mathbb{R}} \hat{t} \wp_{t_0,0|\mathrm{m}}(d\hat{t})$ does not converge, $V[\wp_{t_0,t|\mathrm{m}}]$ is infinite and satisfies (25). Hence, we can assume that the above integral converges. Further, we can assume that the outcome $\hat{t}$ satisfies the unbiasedness condition $\int_{\mathbb{R}} \hat{t} \wp_{t_0,t|\mathrm{m}}(d\hat{t}) = t$. Otherwise, we can replace $\hat{t}$ by $\hat{t}_0 := \hat{t} - \int_{\mathbb{R}} \hat{t}' \wp_{t_0,0|\mathrm{m}}(d\hat{t}')$ because the estimator $\hat{t}_0$ has a smaller MSE than $\hat{t}$ and satisfies the unbiasedness condition due to the covariance condition. Hence, Theorem 1 guarantees

$$V[\wp_{t_0,t|\mathrm{m}}] = V[\wp_{t_0,0|\mathrm{m}}] \geq \left(\liminf_{\epsilon \to 0} \frac{8 \left(1 - F(\wp_{t_0,0|\mathrm{m}} || \wp_{t_0,\epsilon|\mathrm{m}}) \right)}{\epsilon^2} \right)^{-1} . \tag{26}$$

Applying Lemma 20 to $\{\wp_{t_0,t|\mathrm{m}}\}$, we have

$$
\begin{aligned}
&\liminf_{\epsilon \to 0} \frac{8 \left(1 - F(\wp_{t_0,0|\mathrm{m}} || \wp_{t_0,\epsilon|\mathrm{m}}) \right)}{\epsilon^2} \\
&\stackrel{(a)}{\leq} \liminf_{\epsilon \to 0} \liminf_{n \to \infty} \frac{8 \left(1 - F \left(\wp_{t_0,0|M_n}^n || \wp_{t_0,\epsilon|M_n}^n \right) \right)}{\epsilon^2} \\
&\stackrel{(b)}{\leq} \liminf_{\epsilon \to 0} \liminf_{n \to \infty} \frac{8 \left(1 - F \left(\rho_{t_0}^{\otimes n} || \rho_{t_0 + \frac{\epsilon}{\sqrt{n}}}^{\otimes n} \right) \right)}{\epsilon^2} \\
&\stackrel{(c)}{=} \liminf_{\epsilon \to 0} \frac{8 \left(1 - e^{-\frac{J_{t_0} \epsilon^2}{8}} \right)}{\epsilon^2} = J_{t_0} .
\end{aligned}
\tag{27}
$$

The inequality (a) holds by Lemma 20 from "Appendix B", and the inequality (b) comes from the data-processing inequality of the fidelity. The equation (c) follows from Lemma 1. Finally, substituting Eq. (27) into Eq. (26), we have the desired bound (25).

Now, we denote the PDF of $\wp_{t_0,0|\mathrm{m}}$ by $\wp_{t_0,0|\mathrm{m},d}$. In "Appendix A" the proof of Lemma 2 shows that we can apply Lemma 19 to $\{\wp_{t_0,t|\mathrm{m}}\}_t$. Since the Fisher information

of $\{\wp_{t_0,t|\mathbf{m}}\}_t$ is upper bounded by J_{t_0}, this application guarantees that $\wp_{t_0,t|\mathbf{m},d}(x) \leq \sqrt{J_{t_0}}$ for $x \in \mathbb{R}$.

When the PDF $\wp_{t_0,t|\mathbf{m},d}$ is differentiable, to derive the equality condition in Eq. (25), we show (26) in a different way. Let $l_{t_0,t}(x)$ be the logarithmic derivative of $\wp_{t_0,t|\mathbf{m},d}(x)$, defined as $l_{t_0,t}(x) \cdot \wp_{t_0,t|\mathbf{m},d}(x) := \frac{\partial \log \wp_{t_0,t|\mathbf{m},d}(x)}{\partial t} = \frac{\partial \log \wp_{t_0,0|\mathbf{m},d}(x-t)}{\partial t}$. By Schwartz inequality, we have

$$V[\wp_{t_0,0|\mathbf{m}}] \geq \frac{\left|\int_{-\infty}^{\infty} \hat{x}\, l_{t_0,0}(\hat{x})\, \wp_{t_0,0|\mathbf{m}}(\hat{x}) d\hat{x}\right|^2}{\int_{-\infty}^{\infty} l_{t_0,0}^2(\hat{x}) \wp_{t_0,0|\mathbf{m}}(\hat{x}) d\hat{x}}. \tag{28}$$

The numerator on the right hand side of Eq. (28) can be evaluated by noticing that

$$\int_{-\infty}^{\infty} \hat{x}\, l_{t_0,0}(\hat{x})\, \wp_{t_0,0|\mathbf{m}}(\hat{x}) d\hat{x} = \frac{\partial}{\partial x} \int_{-\infty}^{\infty} \hat{x}\, \wp_{t_0,x|\mathbf{m}}(\hat{x}) d\hat{x}\Big|_{x=0}.$$

By local asymptotic covariance, this quantity can be evaluated as

$$\int_{-\infty}^{\infty} \hat{x}\, l_{t_0,0}(\hat{x})\, \wp_{t_0,0|\mathbf{m}}(\hat{x}) d\hat{x} = \frac{\partial}{\partial x} \left[\int_{-\infty}^{\infty} (\hat{x} + x)\, \wp_{t_0,0|\mathbf{m}}(\hat{x}) d\hat{x}\right]\Big|_{x=0} = 1. \tag{29}$$

Hence, (28) coincides with (26). The denominator on the right hand side of (28) equals the right hand side of (26). The equality in Eq. (28) holds if and only if $\int_{-\infty}^{\infty} \hat{x}^2 \wp_{t_0,0|\mathbf{m},d}(\hat{x}) d\hat{x} = J_{t_0}^{-1}$ and $\frac{d}{dx} \log \wp_{t_0,0|\mathbf{m},d}(\hat{x})$ is proportional to $\hat{x}$, which implies that $\wp_{t_0,0|\mathbf{m}}$ is the normal distribution with average zero and variance $J_{t_0}^{-1}$. $\quad\square$

The RHS of (25) can be regarded as the limiting distribution version of the SLD quantum Cramér–Rao bound. Note that, when the limiting PDF is differentiable and the bound is attained, the probability distribution $\wp_{t_0,t|M_n}^n$ is approximated (in the pointwise sense) by a normal distribution with average zero and variance $\frac{1}{nJ_{t_0}}$. Using this fact, we will show that there exists a sequence of POVMs that attains the equality (25) at all points uniformly. The optimal sequence of POVMs is constructed explicitly in Sect. 6.

2.4. Comparison between local asymptotic covariance and other conditions. We conclude the section by discussing the relation between asymptotic covariance and other conditions that are often imposed on measurements. This subsection is not necessary for understanding the technical results in the next sections and can be skipped at a first reading.

Let us start with the unbiasedness condition. Assuming unbiasedness, one can derive the quantum Cramér–Rao bound on the MSE [5]. Holevo showed the attainability of the quantum Cramér–Rao bound when estimating displacements in Gaussian systems [5].

The disadvantage of unbiasedness is that it is too restrictive, as it is satisfied only by a small class of measurements. Indeed, the unbiasedness condition for the estimator M requires the condition $\text{Tr}\left[E \frac{d^i \rho_t}{dt^i}\right]|_{t=t_0} = 0$ for $i \geq 2$ with $E := \int \hat{t} M(d\hat{t})$ as well as the condition $\text{Tr}\left[E \frac{d\rho_t}{dt}\right]|_{t=t_0} = 1$. In certain situations, the above conditions might be incompatible. For example, consider a family of qubit states $\rho_t := \frac{1}{2}(I + \boldsymbol{n}_t \cdot \boldsymbol{\sigma})$. When the Bloch vector $\boldsymbol{n}_t$ has a non-linear dependence on t and the set of higher order derivatives $\frac{d^i \rho_t}{dt^i}|_{t=t_0}$ with $i \geq 2$ spans the space of traceless Hermittian matrices, no

unbiased estimator can exist. In contrast, local asymptotic covariance is only related to the first derivative $\frac{d\rho_t}{dt}|_{t=t_0}$ because the contribution of higher order derivatives to the variable $\hat{t}_n$ has order $o\left(\frac{1}{\sqrt{n}}\right)$ and vanishes under the condition of the local asymptotic covariance.

One can see that the unbiasedness condition implies local asymptotic covariance with the parameterization $\rho_{t_0+\frac{t}{\sqrt{n}}}$ in the following sense. When we have n (more than one) input copies, we can construct unbiased estimator by applying a single-copy unbiased estimator M satisfying Eq. (9) to all copies as follows. For the i-th outcome x_i, we take the rescaled average $\sum_{i=1}^{n}\frac{x_i}{n}$, which satisfies the unbiasedness (9) for the parameter t as well. When the single-copy estimator M has variance v at t_0, which is lower bounded by the Cramér–Rao inequality, this estimator has variance v/n at t_0. In addition, the average (30) of the obtained data satisfies the local asymptotic covariance because the rescaled estimator $\sqrt{n}((\sum_{i=1}^{n}\frac{x_i}{n})-t_0)$ follows the Gaussian distribution with variance v in the large n limit by the central limit theorem; the center of the Gaussian distribution is pinned at the true value of the parameter by unbiasedness; the shape of the Gaussian is independent of the value t and depends only on t_0; thus locally asymptotic covariance holds.

The above discussion can be extended to the multiple-copy case as follows. Suppose that M_ℓ is an unbiased measurement for the ℓ-copy state $\rho_{t_0+\frac{t}{\sqrt{n}}}^{\otimes\ell}$ with respect to the parameter $t_0 + \frac{t}{\sqrt{n}}$, where ℓ is an arbitrary finite integer. From the measurement M_ℓ we can construct a measurement for the n-copy state with $n = k\ell + i$ and $i < \ell$ by applying the measurement M_ℓ k times and discarding the remaining i copies. In the following, we consider the limit where the total number n tends to infinity, while ℓ is kept fixed. When the variance of M_ℓ at t_0 is v/ℓ, the average

$$\sum_{i=1}^{k}\frac{x_i}{k} \tag{30}$$

of the k obtained data $x_1, \ldots, x_k$ satisfies local asymptotic covariance, i.e., the rescaled estimator $\sqrt{n}((\sum_{i=1}^{k}\frac{x_i}{k})-t_0)$ follows the Gaussian distribution with variance v in the large n limit. Therefore, for any unbiased estimator, there exists an estimator satisfying locally asymptotic covariance that has the same variance.

Another common condition, less restrictive than unbiasedness, is *local unbiasedness*. This condition depends on the true parameter t_0 and consists of the following two requirements

$$\int \hat{t}\,\mathrm{Tr}\left[\rho_t^{\otimes\ell}\,M_\ell(d\hat{t})\right]\bigg|_{t=t_0} = t_0 \tag{31}$$

$$\frac{d}{dt}\int \hat{t}\,\mathrm{Tr}\left[\rho_t^{\otimes\ell}\,M_\ell(d\hat{t})\right]\bigg|_{t=t_0} = 1\,, \tag{32}$$

where ℓ is a fixed, but otherwise arbitrary, integer. The derivation of the quantum Cramér–Rao bound still holds, because it uses only the condition (32). When the parametrization ρ_t is C^1 continuous, the first derivative $\frac{d}{dt}\int \hat{t}\,\mathrm{Tr}\left[\rho_t^{\otimes\ell}M_\ell(d\hat{t})\right]$ is continuous at $t = t_0$, and the locally unbiased condition at t_0 yields the local asymptotic covariance at t_0 in the way as Eq. (30).

Table 2. Alternative conditions for deriving MSE bounds

Condition	Definition		Limitation
Unbiasedness	$\int \hat{t}\, \mathrm{Tr}\left[\rho_t^{\otimes \ell} M(d\hat{t})\right] = t$	$\forall t \in \Theta$	Too restrictive (unbiased estimator may not exist)
Local unbiasedness	$\int \hat{t}\, \mathrm{Tr}\left[\rho_{t_0}^{\otimes \ell} M_\ell(d\hat{t})\right] = t_0$		Estimator depends on the true parameter t_0
	$\dfrac{d}{dt}\int \hat{t}\, \mathrm{Tr}\left[\rho_t^{\otimes \ell} M_\ell(d\hat{t})\right]\big|_{t=t_0} = 1$		
Asymptotic unbiasedness	$\lim\limits_{n\to\infty}\int \hat{t}\, \mathrm{Tr}\left[\rho_t^{\otimes n} M_n(d\hat{t})\right] = t$	$\forall t \in \Theta$	Attainability unknown for finite-dimensional systems
	$\lim\limits_{n\to\infty}\dfrac{d}{dt}\int \hat{t}\, \mathrm{Tr}\left[\rho_t^{\otimes n} M_n(d\hat{t})\right] = 1$	$\forall t \in \Theta$	
Weak asymptotic unbiasedness	$\lim\limits_{n\to\infty}\int \hat{t}\, \mathrm{Tr}\left[\rho_t^{\otimes n} M_n(d\hat{t})\right] = t$	$\forall t \in \Theta$	No lower bound to the MSE is known to hold at every point

Another relaxation of the unbiasedness condition is *asymptotic unbiasedness* [11]

$$\lim_{n\to\infty}\int \hat{t}\, \mathrm{Tr}\left[\rho_t^{\otimes n} M_n(d\hat{t})\right] = t \qquad\qquad \forall t \in \Theta \qquad (33)$$

$$\lim_{n\to\infty}\frac{d}{dt}\int \hat{t}\, \mathrm{Tr}\left[\rho_t^{\otimes n} M_n(d\hat{t})\right] = 1 \qquad\qquad \forall t \in \Theta. \qquad (34)$$

The condition of asymptotic unbiasedness leads to a precision bound on MSE [34, Chapter 6]. The bound is given by the SLD Fisher information, and therefore it is attainable for Gaussian states. However, no attainable bound for qudit systems has been derived so far under the condition of asymptotic unbiasedness. Interestingly, one cannot directly use the attainability for Gaussian systems to derive an attainability result for qudit systems, despite the asymptotic equivalence between Gaussian systems and qudit systems stated by *quantum local asymptotic normality (Q-LAN)* (see [16,17] and Sect. 4.1). The problem is that the error of Q-LAN goes to 0 for large n, but the error in the derivative may not go to zero, and therefore the condition (34) is not guaranteed to hold.

In order to guarantee attainability of the quantum Cramér–Rao bound, one could think of further loosening the condition of the asymptotic unbiasedness. An attempt to avoid the problem of the Q-LAN error could be to remove condition (34) and keep only condition (33). This leads to an enlarged class of estimators, called *weakly asymptotically unbiased*. The problem with these estimators is that no general MSE bound is known to hold at every point x. For example, one can find *superefficient* estimators [35,36], which violate the Cramér–Rao bound on a set of points. Such a set must be of zero measure in the limit $n \to \infty$, but the violation of the bound may occur in a considerably large set when n is finite. In contrast, local asymptotic covariance guarantees the MSE bound (25) at every point t where the local asymptotic convariance condition is satisfied. All these alternative conditions for deriving MSE bounds, discussed here in this subsection, are summarized in Table 2.

3. Holevo Bound and Gaussian States Families

3.1. Holevo bound.
When studying multiparameter estimation in quantum systems, we need to address the tradeoff between the precision of estimation of different parameters.

This is done using two types of quantum extensions of Fisher information matrix: the SLD and the right logarithmic derivative (RLD).

Consider a multiparameter family of density operators $\{\rho_t\}_{t\in\Theta}$, where Θ is an open set in $\mathbb{R}^k$, k being the number of parameters. Throughout this section, we assume that ρ_{t_0} is invertible and that the parametrization is C^1 in all parameters. Then, the SLD L_j and the RLD $\tilde{L}_j$ for the parameter t_j are defined through the following equations

$$\frac{\partial\rho_t}{\partial t_j} = \frac{1}{2}\left(\rho_t L_j + L_j \rho_t\right), \quad \frac{\partial\rho_t}{\partial t_j} = \rho_t \tilde{L}_j,$$

see e.g. [5,6] and [15, Sect. II]. It can be seen from the definitions that the SLD L_j can always be chosen to be Hermitian, while the RLD $\tilde{L}_j$ is in general not Hermitian.

The SLD quantum Fisher information matrix J_t and the RLD quantum Fisher information matrix $\tilde{J}_t$ are the $k \times k$ matrices defined as

$$(J_t)_{ij} := \mathrm{Tr}\left[\frac{\rho_t(L_i L_j + L_j L_i)}{2}\right], \quad \left(\tilde{J}_t\right)_{ij} := \mathrm{Tr}\left[\tilde{L}_j^\dagger \rho_t \tilde{L}_i\right]. \tag{35}$$

Notice that the SLD quantum Fisher information matrix J_t is a real symmetric matrix, but the RLD quantum Fisher information matrix $\tilde{J}_t$ is not a real matrix in general.

A POVM M is called an unbiased estimator for the family $\mathcal{S} = \{\rho_t\}$ when the relation

$$t = E_t(M) := \int x \,\mathrm{Tr}\left[\rho_t M(dx)\right]$$

holds for any parameter t. For a POVM M, we define the mean square error (MSE) matrix $V_t(M)$ as

$$\left(V_t(M)\right)_{i,j} := \int (x_i - t_i)(x_j - t_j)\,\mathrm{Tr}\left[\rho_t M(dx)\right]. \tag{36}$$

It is known that an unbiased estimator M satisfies the SLD type and RLD type of Cramer–Rao inequalities

$$V_t(M) \geq J_t^{-1}, \quad \text{and} \quad V_t(M) \geq \tilde{J}_t^{-1}, \tag{37}$$

respectively [5]. Since it is not always possible to minimize the MSE matrix under the unbiasedness condition, we minimize the weighted MSE $\mathrm{tr}\left[W V_t(M)\right]$ for a given weight matrix $W \geq 0$, where tr denotes the trace of $k \times k$ matrices. When a POVM M is unbiased, one has the *RLD bound* [5]

$$\mathrm{tr}\left[W V_t(M)\right] \geq C_{\mathrm{R},\mathcal{M}}(W, t) \tag{38}$$

with

$$\begin{aligned}C_{\mathrm{R},\mathcal{M}}(W, t) &:= \min\left\{\mathrm{tr}[VW] \,\middle|\, V \geq \tilde{J}_t^{-1}\right\} \\ &= \mathrm{tr}\left[W\,\mathsf{Re}(\tilde{J}_t)^{-1}\right] + \mathrm{tr}\left|\sqrt{W}\,\mathsf{Im}(\tilde{J}_t)^{-1}\sqrt{W}\right|.\end{aligned} \tag{39}$$

In particular, when $W > 0$, the lower bound (39) is attained by the matrix $V = \mathsf{Re}(\tilde{J}_t)^{-1} + \sqrt{W}^{-1}|\sqrt{W}\,\mathsf{Im}(\tilde{J}_t)^{-1}\sqrt{W}|\sqrt{W}^{-1}$.

The RLD bound has a particularly tractable form when the model is D-invariant:

Definition 1. The model $\{\rho_t\}_{t \in \Theta}$ is *D-invariant at* t when the space spanned by the SLD operators is invariant under the linear map $\mathcal{D}_t$. For any operator X, $\mathcal{D}_t(X)$ is defined via the following equation

$$\frac{\rho_t \mathcal{D}_t(X) + \mathcal{D}_t(X)\rho_t}{2} = i[X, \rho_t] \tag{40}$$

where $[A, B] = AB - BA$ denotes the commutator. When the model is D-invariant at any point, it is simply called D-invariant.

For a D-invariant model, the RLD quantum Fisher information can be computed in terms of the *D-matrix*, namely the skew-symmetric matrix defined as

$$(D_t)_{j,k} := i \, \text{Tr} \left[\rho_t [L_j, L_k] \right]. \tag{41}$$

Precisely, the RLD quantum Fisher information has the expression [5]

$$\left(\tilde{J}_t \right)^{-1} = (J_t)^{-1} + \frac{i}{2} (J_t)^{-1} D_t (J_t)^{-1}. \tag{42}$$

Hence, (39) becomes

$$\mathcal{C}_{\text{R},\mathcal{M}}(W, t) = \text{tr} \left[W (J_t)^{-1} \right] + \frac{1}{2} \, \text{tr} \left| \sqrt{W} (J_t)^{-1} D_t (J_t)^{-1} \sqrt{W} \right|. \tag{43}$$

For D-invariant models, the RLD bound is larger and thus it is a better bound than the bound derived by using the SLD Fisher information matrix (the SLD bound). However, in the one-parameter case, when the model is not D-invariant, the RLD bound is not tight, and it is common to use the SLD bound in the one-parameter case. Hence, both quantum extensions of the Cramér–Rao bound have advantages and disadvantages.

To unify both extensions, Holevo [5] derived the following bound, which improves the RLD bound when the model is not D-invariant. For a k-component vector X of operators, define the $k \times k$ matrix $Z_t(X)$ as

$$\left(Z_t(X) \right)_{ij} := \text{Tr} \left[\rho_t X_i X_j \right]. \tag{44}$$

Then, Holevo's bound is as follows: for any weight matrix W, one has

$$\inf_{M \in \text{UB}_{\mathcal{M}}} \text{tr} \left[W V_t(M) \right] \geq \mathcal{C}_{\text{H},\mathcal{M}}(W, t)$$

$$:= \min_X \min_V \left\{ \text{tr} \left[W V \right] \, \middle| \, V \geq Z_t(X) \right\} \tag{45}$$

$$= \min_X \text{tr} \left[W \text{Re}(Z_t(X)) \right] + \text{tr} \left| \sqrt{W} \text{Im}(Z_t(X)) \sqrt{W} \right|, \tag{46}$$

where $\text{UB}_{\mathcal{M}}$ denotes the set of all unbiased measurements under the model $\mathcal{M}$, V is a real symmetric matrix, and $X = (X_i)$ is a k-component vector of Hermitian operators satisfying

$$\text{Tr} \left[X_i \frac{\partial \rho_t}{\partial t_j} \right] \bigg|_{t=t_0} = \delta_{ij}, \ \forall i, j \leq k. \tag{47}$$

$\mathcal{C}_{\mathrm{H},\mathcal{M}}(W, t)$ is called the Holevo bound. When $W > 0$, there exists a vector X achieving the minimum in (45). Hence, similar to the RLD case, the equality in (45) holds for $W > 0$ only when

$$V_t(M) = \mathsf{Re}\,(Z_t(X)) + \sqrt{W}^{-1}|\sqrt{W}\mathsf{Im}\,(Z_t(X))\sqrt{W}|\sqrt{W}^{-1}. \qquad (48)$$

Moreover, we have the following proposition.

Proposition 1 ([15, Theorem 4]). *Let $\mathcal{S} = \{\rho_t\}_{t \in \Theta}$ be a generic k-parameter qudit model and let $\mathcal{S}' = \{\rho_{t,p}\}_{t,p}$ be a k'-parameter model containing $\mathcal{S}$ as $\rho_t = \rho_{(t,0)}$. When $\mathcal{S}'$ is D-invariant and the inverse $\tilde{J}_{t'}^{-1}$ of the RLD Fisher information matrix of the model $\mathcal{S}'$ exists, we have*

$$\mathcal{C}_{\mathrm{H},\mathcal{M}}(W, t) = \min_{X:\mathcal{M}'} \min_V \left\{ \mathrm{tr}\,[WV] \,\Big|\, V \geq Z_t(X) \right\} \qquad (49)$$

$$= \min_P \left\{ \mathrm{tr}\,\left[P^T W P (J_{t'}^{-1}) \right] + \frac{1}{2}\,\mathrm{tr}\,\left| \sqrt{P^T W P}\, J_{t'}^{-1} D_{t'} J_{t'}^{-1} \sqrt{P^T W P} \right| \right\}. \qquad (50)$$

In (49), $\min_{X:\mathcal{M}'}$ denotes the minimum for vector X whose components X_i are linear combinations of the SLDs operators in the model $\mathcal{M}'$. In (50), the minimization is taken over all $k \times k'$ matrices satisfying the constraint $(P)_{ij} := \delta_{ij}$ for $i, j \leq k$, $J_{t'}$ and $D_{t'}$ are the SLD Fisher information matrix and the D-matrix [cf. Eqs. (35) and (41)] for the extended model $\mathcal{S}'$ at $t' := (t, \mathbf{0})$.

The Holevo bound is always tighter than the RLD bound:

$$\mathcal{C}_{\mathrm{R},\mathcal{M}}(W, t) \leq \mathcal{C}_{\mathrm{H},\mathcal{M}}(W, t). \qquad (51)$$

The equality holds if and only if the model $\mathcal{M}$ is D-invariant [37].

In the above proposition, it is not immediately clear whether the Holevo bound depends on the choice of the extended model $\mathcal{S}'$. In the following, we show that there is a minimum D-invariant extension of $\mathcal{S}$, and thus the Holevo bound is independent of the choice of $\mathcal{S}'$. The minimum D-invariant subspace in the space of Hermitian matrices is given as follows. Let $\mathcal{V}$ be the subspace spanned SLDs $\{L_i\}$ of the original model $\mathcal{M}$ at ρ_t. Let $\mathcal{V}'$ be the subspace spanned by $\cup_{j=0}^{\infty} \mathcal{D}_t^j(\mathcal{V})$. Then, the subspace $\mathcal{V}'$ is D-invariant and contains $\mathcal{V}$. What remains is to show that $\mathcal{V}'$ is the minimum D-invariance subspace. Let $\mathcal{V}''$ be the orthogonal space with respect to $\mathcal{V}'$ for the inner product defined by $\mathrm{Tr}\,\rho X^\dagger Y$. We denote by P' and P'' the projections into $\mathcal{V}'$ and $\mathcal{V}''$ respectively. Each component X_i of a vector of operators X can be expressed as $X_i = P'X_i + P''X_i$. Then, the two vectors $X' := (P'X_i)$ and $X'' := (P''X_i)$ satisfy the inequality $Z_t(X) = Z_t(X') + Z_t(X'') \geq Z_t(X')$. Substituting Eq. (35) into Eq. (47) and noticing that $P''X_i$ has no support in $\mathcal{V}$, we get that only the part $P'X_i$ contributes the condition (47) and the minimum in (46) is attained when $X'' = 0$. Hence, the minimum is achieved when each component of the vector X is included in the minimum D-invariant subspace $\mathcal{V}'$. Therefore, since the minimum D-invariant subspace can be uniquely defined, the Holevo bound does not depend on the choice of the D-invariant model $\mathcal{S}'$ that extends $\mathcal{S}$.

3.2. Classical and quantum Gaussian states. For a classical system of dimension d^C, a Gaussian state is a d^C-dimensional normal distribution $N[\boldsymbol{\alpha}^C, \Gamma^C]$ with mean $\boldsymbol{\alpha}^C$ and covariance matrix Γ. The corresponding random variable will be denoted as $\boldsymbol{Z} = (Z_1, \ldots, Z_{d^C})$ and will take values $\boldsymbol{z} = (z_1, \ldots, z_{d^C}) \in \mathbb{R}^{d^C}$.

For quantum systems we will restrict our attention to a subfamily of Gaussian states, known as displaced thermal states. For a quantum system made of a single mode, the *displaced thermal states* are defined as

$$\rho_{\alpha,\beta} := T_\alpha^Q \, \rho_\beta^{\text{thm}} (T_\alpha^Q)^\dagger, \qquad T_\alpha^Q = \exp(\alpha \hat{a}^\dagger - \bar{\alpha}\hat{a}), \tag{52}$$

where $\alpha \in \mathbb{C}$ is the *displacement*, T_α^Q is the *displacement operator*, $\hat{a}$ is the annihilation operator satisfying the relation $[\hat{a}, \hat{a}^\dagger] = 1$, and ρ_β^{thm} is a *thermal state*, defined as

$$\rho_\beta^{\text{thm}} := (1 - e^{-\beta}) \sum_{j=0}^{\infty} e^{-j\beta} |j\rangle\langle j|, \tag{53}$$

where the basis $\{|j\rangle\}_{j \in \mathbb{N}}$ consists of the eigenvectors of $\hat{a}^\dagger\hat{a}$ and $\beta \in (0, \infty)$ is a real parameter, hereafter called the *thermal parameter*.

For a quantum system of d^Q modes, the products of single-mode displaced thermal states will be denoted as

$$\Phi[\boldsymbol{\alpha}^Q, \boldsymbol{\beta}^Q] := \bigotimes_{j=1}^{d^Q} \rho_{\alpha_j, \beta_j} \tag{54}$$

where $\boldsymbol{\alpha}^Q = (\alpha_j)_{j=1}^{d^Q}$ is the vector of displacements and $\boldsymbol{\beta}^Q = (\beta_j)_{j=1}^{d^Q}$ is the vector of thermal parameters. In the following we will regard $\boldsymbol{\alpha}$ as a vector in $\mathbb{R}^{2d^Q}$, using the notation $\boldsymbol{\alpha} = (\alpha_1^R, \alpha_1^I, \ldots, \alpha_{d^Q}^R, \alpha_{d^Q}^I)$, $\alpha_j^R := \mathsf{Re}(\alpha_j)$, $\alpha_j^I := \mathsf{Im}(\alpha_j)$.

For a hybrid system of d^C classical variables and d^Q quantum modes, we define the *canonical classical–quantum (c–q) Gaussian states* $G[\boldsymbol{\alpha}, \Gamma]$ as

$$G[\boldsymbol{\alpha}, \Gamma] := N[\boldsymbol{\alpha}^C, \Gamma^C] \otimes \Phi[\boldsymbol{\alpha}^Q, \boldsymbol{\beta}^Q], \tag{55}$$

with $\boldsymbol{\alpha} = \boldsymbol{\alpha}^C \oplus \boldsymbol{\alpha}^Q \in \mathbb{R}^{d^C + 2d^Q}$, $\boldsymbol{\alpha}^C \in \mathbb{R}^{d^C}$, $\boldsymbol{\alpha}^Q \in \mathbb{R}^{2d^Q}$, and $\Gamma = \Gamma^C \oplus \Gamma^Q$, where

$$\Gamma^Q := E_{d^Q}(\boldsymbol{N}) + \frac{i}{2}\Omega_{d^Q},$$

$$E_{d^Q}(\boldsymbol{N}) := \begin{pmatrix} N_1 & 0 & & & \\ 0 & N_1 & & & \\ & & \ddots & & \\ & & & N_{d^Q} & 0 \\ & & & 0 & N_{d^Q} \end{pmatrix} \qquad N_j := \frac{e^{-\beta_j}}{1 - e^{-\beta_j}} \tag{56}$$

$$\Omega_{d^Q} := \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{pmatrix}. \tag{57}$$

Equivalently, the canonical Gaussian states can be expressed as

$$G[\boldsymbol{\alpha}, \Gamma] = T_{\boldsymbol{\alpha}}\, G[\mathbf{0}, \Gamma]\, T_{\boldsymbol{\alpha}}^{\dagger}, \tag{58}$$

where $T_{\boldsymbol{\alpha}}$ is the Gaussian shift operator

$$T_{\boldsymbol{\alpha}} := \left(\bigotimes_{k=1}^{d^{C}} T_{\alpha_{k}^{C}}^{C} \right) \otimes \left(\bigotimes_{j=1}^{d^{Q}} T_{\alpha_{j}^{Q}}^{Q} \right), \tag{59}$$

$T_{\alpha_{k}^{Q}}^{Q}$ is given by Eq. (52), and $T_{\alpha_{j}^{C}}^{C}$ is the map $z_{j} \rightarrow z_{j} + \alpha_{j}^{C}$. For the classical part, we have adopt the notation

$$\mathrm{Tr}\left[N[\boldsymbol{\alpha}^{C}, \Gamma^{C}] \exp\left(\sum_{j=1}^{d^{C}} i\xi_{j} Z_{j} \right) \right] := \int dz_{1} \dots dz_{d^{C}}\, N[\boldsymbol{\alpha}^{C}, \Gamma^{C}](z) \exp\left(\sum_{j=1}^{d^{C}} i\xi_{j} z_{j} \right).$$

With this notation, the canonical Gaussian state $G[\boldsymbol{\alpha}, \Gamma]$ is uniquely identified by the characteristic equation [5]

$$\mathrm{Tr}\left[G[\boldsymbol{\alpha}, \Gamma] \exp\left(\sum_{j=1}^{d} i\xi_{j} R_{j} \right) \right] = \exp\left[i \sum_{j} \xi_{j}\alpha_{j} - \frac{1}{2} \sum_{j,k} \xi_{j}\xi_{k}\, \mathsf{Re}[\Gamma_{j,k}] \right], \tag{60}$$

with

$$d := d^{C} + 2d^{Q}$$
$$R_{j} := Z_{j}, \qquad \forall j \in \{1, \dots, d^{C}\}$$
$$R_{2j-1} := Q_{j} := \frac{a_{j} + a_{j}^{\dagger}}{\sqrt{2}}, \quad R_{2j} := P_{j} := \frac{a_{j} - a_{j}^{\dagger}}{\sqrt{2i}}, \quad \forall j \in \left\{ d^{C} + 1, \dots, d \right\}. \tag{61}$$

The formulation in terms of the characteristic equation (60) can be used to generalize the notion of canonical Gaussian state [38]. Given a d-dimensional Hermitian matrix (correlation matrix) $\Gamma = \mathsf{Re}(\Gamma) + i\,\mathsf{Im}(\Gamma)$ whose real part $\mathsf{Re}(\Gamma)$ is positive semi-definite, we define the operators $\boldsymbol{R} := (R_{1}, \dots, R_{d})$ via the commutation relation

$$[R_{k}, R_{j}] = i\,\mathsf{Im}(\Gamma_{k,j}). \tag{62}$$

We define the *general Gaussian state* $G[\boldsymbol{\alpha}, \Gamma]$ on the operators $\boldsymbol{R}$ as the linear functional on the operator algebra generated by $R_{1}, \dots, R_{d}$ satisfying the characteristic equation (60) [38]. Note that, although Γ is not necessarily positive semi-definite, its real part $\mathsf{Re}(\Gamma)$ is positive semi-definite. Hence, the right-hand-side of Eq. (60) is contains a negative semi-definite quadratic form, in the same way as for the standard Gaussian states.

For general Gaussian states, we have the following lemma.

Lemma 3. *Given a Hermitian matrix Γ, there exists an invertible real matrix T such that the Hermitian matrix $T\Gamma T^{T}$ is the correlation matrix of a canonical Gaussian state. In particular, when $\mathrm{Im}(\Gamma)$ is invertible, $T\Gamma T^{T} = E_{d^{Q}}(N) + \frac{i}{2}\Omega_{d^{Q}}$ and the vector β is unique up to the permutation. Further, when two matrices T and $\tilde{T}$ satisfy the above condition, the canonical Gaussian states $G[T^{-1}\boldsymbol{\alpha}, T\Gamma T^{T}]$ and $G[\tilde{T}^{-1}\boldsymbol{\alpha}, \tilde{T}\Gamma\tilde{T}^{T}]$ are unitarily equivalent.*

The proof is provided in "Appendix C".

In the above lemma, we can transform Γ into the block form $\Gamma^C \oplus \Gamma^Q$ where Γ^C is real by applying orthogonal transformation. The unitary operation on the classical part is given as a scale conversion. Hence, an invertible real matrix T can be realized by the combination of a scale conversion and a linear conversion, which can be implemented as a unitary on the Hilbert space. Hence, a general Gaussian state can be given as the resultant linear functional on the operator algebra after the application of the linear conversion to a canonical Gaussian state. This kind of construction is unique up to unitarily equivalence. Indeed, Petz [38] showed a similar statement by using Gelfand–Naimark–Segal (GNS) construction. Our derivation directly shows the uniqueness without using the GNS construction.

Lemma 4. *The Gaussian states family* $\{G[\boldsymbol{\alpha}, \Gamma]\}_{\boldsymbol{\alpha}}$ *is D-invariant. The SLDs are calculated as* $L_{\boldsymbol{\alpha}, j} = \sum_{k=1}^d ((\mathsf{Re}(\Gamma))^{-1})_{j,k} R_k$. *The D-operator at any point* $\boldsymbol{\alpha}$ *is given as* $\mathcal{D}(R_j) = \sum_k 2\mathsf{Im}(\Gamma)_{j,k} R_k$. *The inverse of the RLD Fisher information matrix* $\tilde{J}_{\boldsymbol{\alpha}}$ *is calculated as*

$$(\tilde{J}_{\boldsymbol{\alpha}})^{-1} = \Gamma. \tag{63}$$

This lemma shows the inverse of the RLD Fisher information matrix is given by the correlation matrix.

Proof. Due to the coordinate conversion give in Lemma 3, it is sufficient to show the relation (63) for the canonical Gaussian states family. In that case, the desired statement has already been shown by Holevo in [5]. □

Therefore, as shown in "Appendix D", a *D*-invariant Gaussian model can be characterized as follows:

Lemma 5. *Given an* $d \times d$ *strictly positive-definite Hermitian matrix* $\Gamma = A + iB$ *(A, B are real matrices) and a* $d \times k$ *real matrix* T *with* $k \leq d$, *the following conditions are equivalent.*

(1) *The linear submodel* $\mathcal{M} := \{G[T\boldsymbol{t}, \Gamma]\}_{\boldsymbol{t} \in \mathbb{R}^k}$ *(with displacement* $T\boldsymbol{t}$*) is D-invariant.*
(2) *The image of the linear map* $A^{-1}T$ *is invariant for the application of* B.
(3) *There exist a unitary operator* U *and a Hermitian matrix* Γ_0 *such that*

$$UG[T\boldsymbol{t}, \Gamma]U^{\dagger} = G[\boldsymbol{t}, \Gamma_T] \otimes G[0, \Gamma_0], \tag{64}$$

where $\Gamma_T := (T^T A^{-1} T)^{-1} + i(T^T A^{-1} T)^{-1}(T^T B T)(T^T A^{-1} T)^{-1}$.

3.3. Measurements on Gaussian states family. We discuss the stochastic behavior of the outcome of the measurement on the c–q system generated by $\boldsymbol{R} = (R_j)_{j=1}^d$ when the state is given as a general Gaussian state $G[\boldsymbol{\alpha}, \Gamma]$. To this purpose, we introduce the notation $\wp_{\boldsymbol{\alpha}|M}(B) := \mathrm{Tr}\left[G[\boldsymbol{\alpha}, \Gamma]M(B)\right]$ for a POVM M. Then, we have the following lemma.

Lemma 6. *Let* $X = (X_i)_{i=1}^k$ *be the vector defined by* $X_i := \sum_{j=1}^d P_{i,j} R_j$, *where* P *is a real* $k \times d$ *matrix. For a weighted matrix* $W > 0$, *there exists a POVM* $M_{P|W}^{\Gamma}$ *with*

outcomes in $\mathbb{R}^k$ such that $\int x_i M^{\Gamma}_{P|W}(d\mathbf{x}) = X_i$ and $\wp_{\boldsymbol{\alpha}|M^{\Gamma}_{P|W}}$ is the normal distribution with average $(\sum_{j=1}^{d} P_{i,j}\alpha_j)^k_{i=1}$ and covariance matrix

$$\mathsf{Re}(Z_{\boldsymbol{\alpha}}(X)) + \sqrt{W}^{-1}|\sqrt{W}\mathsf{Im}(Z_{\boldsymbol{\alpha}}(X))\sqrt{W}|\sqrt{W}^{-1}.$$

In this case, the weighted covariance matrix is

$$\mathrm{tr}\left[W\mathsf{Re}(Z_{\boldsymbol{\alpha}}(X))\right] + \mathrm{tr}\left|\sqrt{W}\mathsf{Im}(Z_{\boldsymbol{\alpha}}(X))\sqrt{W}\right|.$$

The proof is provided in "Appendix E".

In the above lemma, when $X = \mathbf{R}$, we simplify $M^{\Gamma}_{P|W}$ to M^{Γ}_W. This lemma is useful for estimation in the Gaussian states family $\mathcal{M}' := \{G[t, \Gamma]\}_{t \in \mathbb{R}^d}$. In this family, we consider the covariant condition.

Definition 2. A POVM M is a *covariant* estimator for the family $\{G[t, \Gamma]\}_{t \in \mathbb{R}^k}$ when the distribution $\wp_{t|M}(\mathsf{B}) := \mathrm{Tr}\, G[t, \Gamma]M(\mathsf{B})$ satisfies the condition $\wp_{0|M}(\mathsf{B}) = \wp_{t|M}(\mathsf{B}+t)$ for any t. This condition is equivalent to

$$M(\mathsf{B}+t) = T_t M(\mathsf{B})T_t^{\dagger} \qquad \forall t \in \mathbb{R}^k .$$

Then, we have the following lemma for this Gaussian states family.

Corollary 1 ([5]). *For any weight matrix $W \geq 0$ and the above Gaussian states family $\mathcal{M}'$, we have*

$$\inf_{M \in \mathsf{UB}_{\mathcal{M}'}} \mathrm{tr}\left[WV_t(M)\right] = \inf_{M \in \mathsf{CUB}_{\mathcal{M}'}} \mathrm{tr}\left[WV_t(M)\right]$$
$$= \mathcal{C}_{\mathrm{R},\mathcal{S}'}(W, t) = \mathrm{tr}\left[\mathsf{Re}(\Gamma)W\right] + \mathrm{tr}\left|\sqrt{W}\mathsf{Im}(\Gamma)\sqrt{W}\right|, \qquad (65)$$

where $\mathsf{CUB}_{\mathcal{M}'}$ are the sets of covariant unbiased estimators for the model $\mathcal{M}'$, respectively. Further, when $W > 0$, the above infimum is attained by the covariant unbiased estimators M^{Γ}_W whose output distribution is the normal distribution with average t and covariance matrix $\mathsf{Re}((\Gamma) + \sqrt{W}^{-1}|\sqrt{W}\mathsf{Im}(\Gamma)\sqrt{W}|\sqrt{W}^{-1}$.

This corollary can be shown as follows. Due to Lemma 4, the lower bound (43) of the weighted MSE $\mathrm{tr}\, WV_t(M)$ of unbiased estimator M is calculated as the RHS of (65). Lemma 6 guarantees the required performance of M^{Γ}_W. To discuss the case when W is not strictly positive definite, we consider $W_\epsilon := W + \epsilon I$. Using the above method, we can construct an unbiased and covariant estimator whose output distribution is the $2d^Q$-dimensional distribution of average t and covariance $\mathsf{Re}((\Gamma) + \sqrt{W_\epsilon}^{-1}|\sqrt{W_\epsilon}\mathsf{Im}(\Gamma)\sqrt{W_\epsilon}|\sqrt{W_\epsilon}^{-1}$. The weighted MSE matrix is $\mathrm{tr}\left[W\mathsf{Re}(\Gamma)\right] + \mathrm{tr}\left[\sqrt{W_\epsilon}^{-1}W\sqrt{W_\epsilon}^{-1}|\sqrt{W_\epsilon}\mathsf{Im}(\Gamma)\sqrt{W_\epsilon}|\right]$, which converges to the bound (65).

By combining Proposition 1, this corollary can be extended to a linear subfamily of k'-dimensional Gaussian family $\{G[t', \Gamma]\}_{t' \in \mathbb{R}^{k'}}$. Consider a linear map T from $\mathbb{R}^k$ to $\mathbb{R}^{k'}$. We have the following corollary for the subfamily $\mathcal{M} := \{G[T(t), \Gamma]\}_{t \in \mathbb{R}^k}$.

Corollary 2. *For any weight matrix $W \geq 0$, we have*

$$\inf_{M \in \mathsf{UB}_{\mathcal{M}}} \mathrm{tr}\left[W V_t(M)\right] = \inf_{M \in \mathsf{CUB}_{\mathcal{M}}} \mathrm{tr}\left[W V_t(M)\right] = \mathcal{C}_{\mathrm{H},\mathcal{M}}(W, t). \quad (66)$$

Further, when $W > 0$, we choose a vector X to realize the minimum in (49). The above infimum is attained by the covariant unbiased estimators M_W whose output distribution is the normal distribution with average t and covariance matrix $\mathrm{Re}((Z_t(X)) + \sqrt{W}^{-1}|\sqrt{W}\mathrm{Im}(Z_t(X))\sqrt{W}|\sqrt{W}^{-1}$.

Proposition 1 guarantees that $\mathcal{C}_{\mathrm{H},\mathcal{M}}(W, t)$ with (49) can be given when the components X are given a linear combination of $R_1, \ldots, R_{k'}$. Hence, the latter part of the corollary with $W > 0$ follows from (45) and Lemma 6, implies this corollary for $W > 0$. The case with non strictly positive W can be shown by considering W_ϵ in the same way as Corollary 1.

4. Local Asymptotic Normality

The extension from one-parameter estimation to multiparameter estimation is quite non-trivial. Hence we first develop the concept of local asymptotic normality which is the key tool to constructing the optimal measurement in multiparameter estimation. Since we could derive the tight bound of MSE for the Gaussian states family, it is a natural idea to approximate the general case by Gaussian states family, and local asymptotic normality will serve as the bridge between these general qudit families and Gaussian state families.

4.1. Quantum local asymptotic normality with specific parametrization. For a quantum system of dimension $d < \infty$, also known as qudit, we consider *generic states*, described by density matrices with full rank and non-degenerate spectrum. To discuss quantum local asymptotic normality, we need to define a specific coordinate system. For this aim, we consider the neighborhood of a fixed density matrix ρ_{θ_0}, assumed to be diagonal in the canonical basis of $\mathbb{C}^d$, and parametrized as

$$\rho_{\theta_0} = \sum_{j=1}^{d} \theta_{0,j} \, |j\rangle\langle j|$$

with spectrum ordered as $\theta_{0,1} > \cdots > \theta_{0,d-1} > \theta_{0,d} > 0$. In the neighborhood of ρ_{θ_0}, we parametrize the states of the system as

$$\rho_{\theta_0 + \frac{\theta}{\sqrt{n}}} = U_{\theta^R, \theta^I} \, \rho_0(\theta^C) \, U^\dagger_{\theta^R, \theta^I} \quad (67)$$

for $\theta := (\theta^C, \theta^R, \theta^I)$ with $(\theta^R, \theta^I) \in \mathbb{R}^{d(d-1)}$ and $\theta^C \in \mathbb{R}^{d-1}$, where $\rho_0(\theta^C)$ is the diagonal density matrix

$$\rho_0(\theta^C) = \sum_{j=1}^{d} \left(\theta_{0,j} + \frac{\theta^C_j}{\sqrt{n}}\right) |j\rangle\langle j|, \qquad \theta^C_d := -\sum_{k=1}^{d-1} \theta^C_k, \quad (68)$$

and U_{θ^R, θ^I} is the unitary matrix defined by

$$U_{\theta^R, \theta^I} = \exp\left[\sum_{1 \le j < k \le d} \frac{i\left(\theta_{j,k}^{\mathrm{I}} F_{j,k}^{\mathrm{I}} + \theta_{j,k}^{\mathrm{R}} F_{k,j}^{\mathrm{R}}\right)}{\sqrt{n(\theta_{0,j} - \theta_{0,k})}}\right]. \tag{69}$$

Here θ^R and θ^I are vectors of real parameters $(\theta_{j,k}^R)_{1 \le j < k \le d}$ and $(\theta_{j,k}^I)_{1 \le j < k \le d}$, and F^{I} (F^{R}) is the matrix defined by $(F^{\mathrm{I}})_{j,k} := i\delta_{j,k} - i\delta_{k,j}$ $((F^{\mathrm{R}})_{k,j} := \delta_{k,j} + \delta_{k,j})$, where $\delta_{j,k}$ is the delta function. We note that by this definition the components of θ^R and θ^I are in one-to-one correspondence. The parameter $\theta = (\theta^C, \theta^R, \theta^I)$ will be referred to as *the Q-LAN coordinate*, and the state with this parametrization, which was used by Khan and Guta in [16,17,39], will be denoted by $\rho_\theta^{\mathrm{KG}}$.

Q-LAN establishes an asymptotic correspondence between multicopy qudit states and Gaussian shift models. Using the parameterization $\theta = (\theta^C, \theta^R, \theta^I)$, we have the multicopy qudit models and Gaussian shift models are equivalent in terms of the RLD quantum Fisher information matrix:

Lemma 7. *The RLD quantum Fisher information matrices of the qudit model and the corresponding Gaussian model in Eq.* (75) *are both equal to*

$$\left(\tilde{\mathcal{J}}_\theta^Q\right)^{-1} = E_{d(d-1)/2}\left(\beta'\right) + \frac{i}{2}\Omega_{d(d-1)/2} \qquad e^{-\beta_i'} = \frac{1}{4}\coth\frac{\beta_i}{2}. \tag{70}$$

The calculations can be found in "Appendix F". The quantum version of local asymptotic normality has been derived in several different forms [16,17,39] with applications in quantum statistics [12,40], benchmarks [41] and data compression [42]. Here we use the version of [17], which states that n identical copies of a qudit state can be locally approximated by a c–q Gaussian state in the large n limit. The approximation is in the following sense:

Definition 3 (*Compact uniformly asymptotic equivalence of models*). For every $n \in \mathbb{N}^*$, let $\{\rho_{t,n}\}_{t \in \Theta_n}$ and $\{\widetilde{\rho}_{t,n}\}_{t \in \Theta_n}$ be two models of density matrices acting on Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ respectively where the set of parameters Θ_n may depend on n. We say that the two families are *asymptotically equivalent* for $t \in \Theta_n$, denoted as $\rho_{t,n} \cong \widetilde{\rho}_{t,n}$ ($t \in \Theta_n$), if there exists a quantum channel $\mathcal{T}_n$ (i.e. a completely positive trace preserving map) mapping trace-class operators on $\mathcal{H}$ to trace-class operators on $\mathcal{K}$ and a quantum channel $\mathcal{S}_n$ mapping trace-class operators on $\mathcal{K}$ to trace-class operators on $\mathcal{H}$, which are independent of t and satisfy the conditions

$$\sup_{t \in \Theta_n} \left\|\mathcal{T}_n\left(\rho_{t,n}\right) - \widetilde{\rho}_{t,n}\right\|_1 \xrightarrow{n \to \infty} 0 \tag{71}$$

$$\sup_{t \in \Theta_n} \left\|\rho_{t,n} - \mathcal{S}_n\left(\widetilde{\rho}_{t,n}\right)\right\|_1 \xrightarrow{n \to \infty} 0. \tag{72}$$

Next, we extend asymptotic equivalence to compact uniformly asymptotic equivalence. In this extension, we also describe the order of the convergence.

Given a sequence $\{a_n\}$ converging to zero, for every t' in a compact set K consider two models $\{\rho_{t,t',n}\}_{t \in \Theta_n}$, and $\{\widetilde{\rho}_{t,t',n}\}_{t \in \Theta_n}$. We say that they are *asymptotically equivalent* for $t \in \Theta_n$ compact uniformly with respect to t' with order a_n, denoted as $\rho_{t,t',n} \overset{t'}{\cong} \widetilde{\rho}_{t,t',n}$ ($t \in \Theta_n, a_n$), if for every $t' \in \mathsf{K}$ there exists a quantum channel $\mathcal{T}_{n,t'}$ mapping

trace-class operators on $\mathcal{H}$ to trace-class operators on $\mathcal{K}$ and a quantum channel $\mathcal{S}_{n,t'}$ mapping trace-class operators on $\mathcal{K}$ to trace-class operators on $\mathcal{H}$ such that

$$\sup_{t' \in K} \sup_{t \in \Theta_n} \|\mathcal{T}_{n,t'}(\rho_{t,t',n}) - \widetilde{\rho}_{t,t',n}\|_1 = O(a_n) \tag{73}$$

$$\sup_{t' \in K} \sup_{t \in \Theta_n} \|\rho_{t,t',n} - \mathcal{S}_{n,t'}(\widetilde{\rho}_{t,t',n})\|_1 = O(a_n). \tag{74}$$

Notice that the channels $\mathcal{T}_{n,t'}$ and $\mathcal{S}_{n,t'}$ depend on t' and are independent of t.

In the above terminology, Q-LAN establishes an asymptotic equivalence between families of n copy qudit states and Gaussian shift models. Precisely, one has the following.

Proposition 2 (Q-LAN for a fixed parameterization; Kahn and Guta [16,17]). *For any* $x < 1/9$, *we define the set* $\Theta_{n,x}$ *of* θ *as*

$$\Theta_{n,x} := \left\{ \theta \mid \|\theta\| \leq n^x \right\}$$

($\|\cdot\|$ denotes the vector norm). Then, we have the following compact uniformly asymptotic equivalence;

$$(\rho^{\mathrm{KG}}_{\theta_0 + \theta/\sqrt{n}})^{\otimes n} \overset{\theta_0}{\cong} G[\theta, \Gamma_{\theta_0}] := N[\theta^C, \Gamma^C_{\theta_0}] \otimes \Phi[(\theta^R, \theta^I), \beta_{\theta_0}] \; (\theta \in \Theta_{n,x}, n^{-\kappa}), \tag{75}$$

where κ is a parameter to satisfy $\kappa \geq 0.027$, and $N[\theta^C, \Gamma_{\theta_0}]$ is the multivariate normal distribution with mean θ^C and covariance matrix $\Gamma_{\theta_0,k,l} := (J^{-1}_{\theta_0})_{k,l}$ for $k, l = 1, \ldots, d - 1$, and $(\beta)_{\theta_0,j,k} := \frac{(\rho_{\theta_0})_{k,k}}{(\rho_{\theta_0})_{j,j}}$.

The conditions (73) and (74) are not enough to translate precision limits for one family into precision limits for the other. This is because such limits are often expressed in terms of the derivatives of the density matrix, whose asymptotic behaviour is not fixed by (73) and (74). In the following we will establish an asymptotic equivalence in terms of the RLD quantum Fisher information.

4.2. Quantum local asymptotic normality with generic parametrization. In the following, we explore to which extent can we extend Q-LAN in Proposition 2. Precisely, we derive a Q-LAN equivalence as in Eq. (75) which is not restricted to the parametrization of Eqs. (68) and (69).

In the previous subsection, we have discussed the specific parametrization given in (67). In the following, we discuss a generic parametrization. Given an arbitrary D-invariant model $\rho^{\otimes n}_{t_0 + \frac{t}{\sqrt{n}}}$ with vector parameter t, we have the following theorem.

Theorem 3 (Q-LAN for an arbitrary parameterization). *Let $\{\rho_t\}_{t \in \Theta}$ be a k-parameter D-invariant qudit model. Assume that ρ_{t_0} is a non-degenerate state, the parametrization is C^2 continuous, and $\tilde{J}^{-1}_{t_0}$ exists. Then, there exist a constant $c(t_0)$ such that the set*

$$\Theta_{n,x,c(t_0)} := \left\{ t \mid \|t\| \leq c(t_0)n^x \right\} \tag{76}$$

with $x < 1/9$ satisfies

$$\rho_{t_0 + \frac{t}{\sqrt{n}}}^{\otimes n} \stackrel{t_0}{\cong} G[t, \tilde{J}_{t_0}^{-1}] \ (t \in \Theta_{n,x,c(t_0)} \cap \mathbb{R}^k, n^{-\kappa}), \tag{77}$$

where $\tilde{J}_{t_0}^{-1}$ is the RLD Fisher information at t_0 and κ is a parameter to satisfy $\kappa \geq 0.027$.

Proof. We choose the basis $\{|i\rangle\}_{i=1}^d$ to diagonalize the state ρ_{t_0}. We denote the Q-LAN parametrization based on this basis by $\rho_{\boldsymbol{\theta}}^{\mathrm{KG}|t_0}$, where this parametrization depends on t_0. It is enough to consider the neighborhood $U(t_0)$ of t_0. There exists a map f_{t_0} on $U(t_0)$ such that $\rho_{t_0+t} = \rho_{\boldsymbol{\theta}_0(t_0)+f_{t_0}(t)}^{\mathrm{KG}|t_0}$, where $\boldsymbol{\theta}_0(t_0)$ is the parameter to describe the diagonal elements of ρ_{t_0}. Since the parametrization ρ_t is C^2-continuous, the function f is also C^2-continuous. Proposition 2 guarantees that

$$\begin{aligned}
\rho_{t_0+t/\sqrt{n}}^{\otimes n} &= \left(\rho_{\boldsymbol{\theta}_0(t_0)+f_{t_0}(t/\sqrt{n})}^{\mathrm{KG}|t_0} \right)^{\otimes n} \\
&\stackrel{\boldsymbol{\theta}_0}{\cong} G[\sqrt{n} f_{t_0}(t/\sqrt{n}), \Gamma_{\boldsymbol{\theta}_0(t_0)}] \ (t \in \Theta_{n,x,c(t_0)} \cap \mathbb{R}^k, n^{-\kappa})
\end{aligned} \tag{78}$$

with suitable choice of the constant $c(t_0)$. Denoting by f_{t_0}' the Jacobian matrix of f_{t_0}, since f is C^2-continuous and $f_{t_0}(0) = 0$, the norm $\|\sqrt{n} f_{t_0}(t/\sqrt{n}) - f_{t_0}'(0)t\|_1$ is evaluated as $O(\frac{\|t\|^2}{\sqrt{n}})$. Hence, the trace norm $\|G[\sqrt{n} f_{t_0}(t/\sqrt{n}), \Gamma_{\boldsymbol{\theta}_0(t_0)}] - G[f_{t_0}'(0)t, \Gamma_{\boldsymbol{\theta}_0(t_0)}]\|_1$ is also $O(\frac{\|t\|^2}{\sqrt{n}})$, which is at most $O(n^{-5/18})$ because $t \in \Theta_{n,x,c(t_0)}$. Since $O(n^{-5/18})$ is smaller than $n^{-\kappa}$, the combination of this evaluation and (78) yields

$$\rho_{t_0+t/\sqrt{n}}^{\otimes n} \stackrel{\boldsymbol{\theta}_0}{\cong} G[f_{t_0}'(0)t, \Gamma_{\boldsymbol{\theta}_0(t_0)}] \ (t \in \Theta_{n,x,c(t_0)} \cap \mathbb{R}^k, n^{-\kappa}). \tag{79}$$

The combination of Lemma 5 and (79) implies (77). ∎

5. The ϵ-Difference RLD Fisher Information Matrix

In Sect. 2.1 we evaluated the limiting distribution in the one-parameter case, using the fidelity as a discretized version of the SLD Fisher information. In order to tackle the multiparameter case, we need to develop a similar discretization for the RLD Fisher information matrix, which is the relevant quantity for the multiparameter setting (cf. Sect. 3). In this section we define a discretized version of the RLD Fisher information matrix, extending to the multiparameter case the single-parameter definition introduced by Tsuda and Matsumoto [25], who in turn extended the corresponding classical notion [43,44].

5.1. Definition. Let $\mathcal{M} = \{\rho_t\}_{t \in \Theta}$ be a k-parameter model, with the property that ρ_{t_0} is invertible. If the parametrization ρ_t is differentiable, the RLD quantum Fisher information matrix $\tilde{J}_t$ can be rewritten as the following $k \times k$ matrix

$$\left(\tilde{J}_{t_0} \right)_{ij} = \mathrm{Tr} \left[\tilde{L}_j^\dagger \rho_{t_0} \tilde{L}_i \right] = \mathrm{Tr} \left[\left. \frac{\partial \rho_t}{\partial t_j} \right|_{t=t_0} \rho_{t_0}^{-1} \left. \frac{\partial \rho_t}{\partial t_i} \right|_{t=t_0} \right]. \tag{80}$$

The ϵ-difference RLD quantum Fisher information matrix $\tilde{J}_{t_0,\epsilon}$ is defined by replacing the partial derivatives with finite increments:

$$\left(\tilde{J}_{t_0,\epsilon}\right)_{i,j} := \mathrm{Tr}\left[\left(\frac{\rho_{t_0+\epsilon e_i} - \rho_{t_0}}{\epsilon}\right)\rho_{t_0}^{-1}\left(\frac{\rho_{t_0+\epsilon e_j} - \rho_{t_0}}{\epsilon}\right)\right]$$

$$= \frac{\mathrm{Tr}\left[\rho_{t_0+\epsilon e_i}\,\rho_{t_0}^{-1}\,\rho_{t_0+\epsilon e_j}\right] - 1}{\epsilon^2}, \tag{81}$$

where e_j is the unit vector with 1 in the j-th entry and zero in the other entries. Notice that one has

$$\left(\tilde{J}_{t_0,\epsilon}\right)_{i,i} = \frac{\exp\left[D_2(\rho_{t_0+\epsilon e_i}||\rho_{t_0})\right] - 1}{\epsilon^2}, \tag{82}$$

where $D_2(\rho||\sigma) := \log\mathrm{Tr}\left[\rho^2\sigma^{-1}\right]$ is the (Petz's) α-Renyi entropy for $\alpha = 2$.

When the parametrization ρ_t is differentiable, one has

$$\lim_{\epsilon\to 0}\tilde{J}_{t_0,\epsilon} = \tilde{J}_{t_0}, \tag{83}$$

where $\tilde{J}_{t_0}$ is the RLD quantum Fisher information matrix (80).

When the parametrization is not differentiable, we define the RLD Fisher information matrix $\tilde{J}_{t_0}$ to be the limit (83), provided that the limit exists. All throughout this section, we impose no condition on the parametrization ρ_t, except for the requirement that ρ_{t_0} be invertible.

5.2. The ϵ-difference RLD Cramér–Rao inequality.
A discrete version of the RLD quantum Cramér–Rao inequality can be derived under the assumption of ϵ-*locally unbiasedness*, defined as follows:

Definition 4. A POVM M with outcomes in $\mathbb{R}^k$ is ϵ-locally unbiased at t_0 if the expectation value $E_{t_0}(M)$ satisfies the conditions

$$E_{t_0}(M) = t_0, \quad \text{and} \quad E_{t_0+\epsilon e_j}(M) = t_0 + \epsilon e_j \quad \forall j \in \{1, \dots, k\}.$$

Under the ϵ-locally unbiasedness condition, Tsuda et al. [25] derived a lower bound on the MSE for the one-parameter case. In the following theorem, we extend the bound to the multiparameter case.

Theorem 4 (ϵ-difference RLD Cramér–Rao inequality). *The MSE matrix for an ϵ-locally unbiased POVM M at t_0 satisfies the bound*

$$V_{t_0}(M) \geq (\tilde{J}_{t_0,\epsilon})^{-1}. \tag{84}$$

Proof. For simplicity, we assume that $t_0 = 0$. For two vectors $a \in \mathbb{C}^k$ and $b \in \mathbb{C}^k$, we define the two observables $X := \int\sum_i a_i x_i\, M(dx)$ and $Y := \sum_j b_j\frac{\rho_{t_0+\epsilon e_j} - \rho_{t_0}}{\epsilon}\rho_{t_0}^{-1}$.

Then, the Cauchy-Schwartz inequality implies

$$\mathrm{Tr}\left[X^{\dagger}X\,\rho_{t_0}\right]\mathrm{Tr}\left[Y^{\dagger}Y\,\rho_{t_0}\right]\geq\left|\mathrm{Tr}\left[X^{\dagger}Y\rho_{t_0}\right]\right|^2$$

$$=\left|\sum_{i,j}\overline{a}_i b_j\int x_i\,\mathrm{Tr}\left[M(d\boldsymbol{x})\frac{\rho_{t_0+\epsilon\boldsymbol{e}_j}-\rho_{t_0}}{\epsilon}\right]\right|^2$$

$$=|\langle\boldsymbol{a}|\boldsymbol{b}\rangle|^2\,,\tag{85}$$

the second equality following from ϵ-locally unbiasedness at t_0. Note that one has $\mathrm{Tr}[Y^{\dagger}Y\rho_{t_0}]=\langle\boldsymbol{b}|\,\tilde{J}_{t_0,\epsilon}\,|\boldsymbol{b}\rangle$ and

$$\langle\boldsymbol{a}|V_{t_0}(M)|\boldsymbol{a}\rangle-\mathrm{Tr}\left[X^{\dagger}X\rho_{t_0}\right]=\int\mathrm{Tr}\left[\left(\sum_i a_i x_i-X\right)^{\dagger}\left(\sum_j a_j x_j-X\right)M(d\boldsymbol{x})\rho_{t_0}\right]$$

$$\geq 0\,.\tag{86}$$

Choosing $\boldsymbol{b}:=(\tilde{J}_{t_0,\epsilon})^{-1}\boldsymbol{a}$, we have

$$\langle\boldsymbol{a}|V_{t_0}(M)|\boldsymbol{a}\rangle\,\langle\boldsymbol{a}|\tilde{J}_{t_0,\epsilon}^{-1}|\boldsymbol{a}\rangle\geq\mathrm{Tr}\left[X^{\dagger}X\,\rho_{t_0}\right]\mathrm{Tr}\left[Y^{\dagger}Y\,\rho_{t_0}\right]$$

$$\geq\left|\mathrm{Tr}\left[X^{\dagger}Y\rho_{t_0}\right]\right|^2$$

$$=\left|\langle\boldsymbol{a}|(\tilde{J}_{t_0,\epsilon})^{-1}|\boldsymbol{a}\rangle\right|^2\,,\tag{87}$$

which implies $\langle\boldsymbol{a}|V_{t_0}(M)|\boldsymbol{a}\rangle\geq\langle\boldsymbol{a}|(\tilde{J}_{t_0,\epsilon})^{-1}|\boldsymbol{a}\rangle$. Since $\boldsymbol{a}$ is arbitrary, the last inequality implies (84).

We will call (84) the ϵ-*difference RLD Cramér–Rao inequality*.

The ϵ-difference RLD Cramér–Rao inequality can be used to derive an information processing inequality, which states that the ϵ-difference RLD Fisher information matrix is non-increasing under the application of measurements. For a family of probability distributions $\{P_t\}_{t\in\Theta}$, we assume that $P_{t+\epsilon\boldsymbol{e}_j}$ is absolutely continuous with respect to P_t for every j. Then, the ϵ-difference RLD Fisher information is defined as

$$(J_{t,\epsilon})_{ij}:=\int\left(\frac{p_{t+\epsilon\boldsymbol{e}_j}(x)-1}{\epsilon}\right)\left(\frac{p_{t+\epsilon\boldsymbol{e}_i}(x)-1}{\epsilon}\right)P_t(dx)\tag{88}$$

where $p_{t+\epsilon\boldsymbol{e}_j}$ and $p_{t+\epsilon\boldsymbol{e}_i}$ are the Radon–Nikodým derivatives of $P_{t+\epsilon\boldsymbol{e}_j}$ and $P_{t+\epsilon\boldsymbol{e}_i}$ with respect to P_t, respectively. We note that the papers [43,44] defined its one-parameter version when the distributions are absolutely continuous with respect to the Lebesgue measure. Hence, when an estimator $\hat{t}$ for the distribution family $\{P_t\}_{t\in\Theta}$ is ϵ-locally unbiased at t_0, in the same way as (84), we can show the ϵ-difference Cramér–Rao inequality;

$$V_{t_0}[\hat{t}]\geq(J_{t_0,\epsilon})^{-1}.\tag{89}$$

For a family of quantum states $\{\rho_t\}_{t\in\Theta}$ and a POVM M, we denote by $J_{t,\epsilon}^M$ the ϵ-difference Fisher information matrix of the probability distribution family $\{P_t^M\}_{t\in\Theta}$ defined by $P_t^M:=\mathrm{Tr}\left[M\rho_t\right]$. With this notation, we have the following lemma:

Lemma 8. *For every family of quantum states $\{\rho_t\}_{t\in\Theta}$ and every POVM M, one has the information processing inequality*

$$\tilde{J}_{t_0,\epsilon} \geq J^M_{t_0,\epsilon}\,, \tag{90}$$

where $\tilde{J}_{t_0,\epsilon}$ is the ϵ-difference RLD Fisher information of the model $\{\rho_t\}_{t\in\Theta}$.

Proof. Consider the estimation of t from the probability distribution family $\{P^M_t\}_{t\in\Theta}$. Following the same arguments used for the achievability of the Cramér–Rao bound with locally unbiased estimators (see, for instance, Chapter 2 of Ref. [34]), it is possible to show that there exists an ϵ-locally unbiased estimator $\hat{t}$ at t_0 such that

$$V_{t_0}\left(\hat{t}\right) = (J^M_{t_0,\epsilon})^{-1}\,. \tag{91}$$

Combining the POVM M with the ϵ-locally unbiased estimator $\hat{t}$ we obtain a new POVM M', which is ϵ-locally unbiased. Applying Theorem 4 to the POVM M' we obtain

$$(\tilde{J}_{t_0,\epsilon})^{-1} \leq V_{t_0}\left(M'\right) = V_{t_0}\left(\hat{t}\right) = (J^M_{t_0,\epsilon})^{-1}\,, \tag{92}$$

which implies (90).   □

We stress that (90) is a matrix inequality for *Hermitian* matrices: in general, $\tilde{J}_{t_0,\epsilon}$ has complex entries. Also note that any classical process can be regarded as a POVM. Hence, in the same way as (90), using the ϵ-difference Cramér–Rao inequality (89), we can show the inequality

$$J_\epsilon \geq J'_\epsilon \tag{93}$$

for an classical process $\mathcal{E}$ when J_ϵ is the ϵ-difference Fisher information matrix on the distribution family $\{P_t\}_{t\in\Theta}$ and J'_ϵ is the ϵ-difference Fisher information matrix on the distribution family $\{\mathcal{E}(P_t)\}_{t\in\Theta}$.

5.3. Extended models. The lemmas in the previous subsection can be generalized to the case where an extended model $\mathcal{M}' := \{\rho_{t'}\}_{t'=(t,p)}$ contains the original model $\mathcal{M}$ as $\rho_t = \rho_{(t,0)}$. Choosing $t'_0 = (t_0, 0)$, we denote the ϵ-difference RLD Fisher information matrix at t'_0 for the family $\mathcal{M}'$ by $\tilde{J}_{t'_0,\epsilon}$.

Lemma 9. *For an ϵ-locally unbiased estimator M at t'_0, there exists a $k \times k'$ matrix P such that $P_{ij} = \delta_{ij}$ for $i, j \leq k$ and*

$$V_{t_0}(M) \geq P\tilde{J}^{-1}_{t'_0,\epsilon}P^T\,.$$

Proof of Lemma 9. For an ϵ-locally unbiased estimator M at t_0, there exists a $k \times k'$ matrix P such that

$$P_{ij} = \delta_{ij} \text{ for } i, j \leq k \tag{94}$$

$$\epsilon P_{ij} = \int x_i \operatorname{Tr} M(dx)(\rho_{t'_0+\epsilon e_j} - \rho_{t'_0}) \text{ for } i \leq k, k+1 \leq j \leq k'. \tag{95}$$

Now, we introduce a new parametrization $\tilde{\rho}_\eta := \rho_{t_0' + \sum_{i,j} \eta_i P_{j,i}^{-1} e_j}$. Since $\frac{\partial \theta_j}{\partial \eta_i} = P_{j,i}^{-1}$, the ϵ-difference RLD quantum Fisher information under the parameter η is $(P^{-1})^T \tilde{J}_{t_0',\epsilon} P^{-1}$. Applying Theorem 4 to the parameter η, we obtain

$$V_{t_0}(M) \geq P \tilde{J}_{t_0',\epsilon}^{-1} P^T. \tag{96}$$

Combining (94) and (96), we obtain the desired statement.

In the same way as Lemmas 8, 9 yields the following lemma.

Lemma 10. *For any POVM M, there exists a $k \times k'$ matrix P such that $P_{ij} = \delta_{ij}$ for $i, j \leq k$ and*

$$(J_{t_0,\epsilon}^M)^{-1} \geq P \tilde{J}_{t_0',\epsilon}^{-1} P^T. \tag{97}$$

5.4. Asymptotic case. We denote by $\tilde{J}_{t_0,\epsilon}^n$ the ϵ-difference RLD Fisher information matrix of the n-copy states $\{\rho_t^{\otimes n}\}_{t \in \Theta}$.

In the following we provide the analogue of Lemma 1 for the RLD Fisher information matrix.

Lemma 11. *When the parametrization is C^1 continuous, the following relations hold*

$$\lim_{n \to \infty} \frac{1}{n} \tilde{J}_{t_0, \frac{\epsilon}{\sqrt{n}}}^n = \tilde{J}_{t_0}^{[\epsilon]} \tag{98}$$

$$\lim_{\epsilon \to 0} \tilde{J}_{t_0}^{[\epsilon]} = \tilde{J}_{t_0}, \tag{99}$$

where $J_{t_0}^{[\epsilon]}$ is the matrix defined by $\left(\tilde{J}_{t_0}^{[\epsilon]}\right)_{i,j} := \frac{1}{\epsilon^2}\left[e^{\epsilon^2\left(\tilde{J}_{t_0}\right)_{i,j}} - 1\right]$.

Proof of Lemma 11. Eq. (99) holds trivially. Using (83), we have

$$\lim_{n \to \infty} \frac{1}{n} \left(\tilde{J}_{t_0, \frac{\epsilon}{\sqrt{n}}}^n\right)_{i,j} = \lim_{n \to \infty} \frac{1}{\epsilon^2} \left(\text{Tr}\left[\rho_{t_0+\frac{\epsilon}{\sqrt{n}}e_j}^{\otimes n} (\rho_{t_0}^{\otimes n})^{-1} \rho_{t_0+\frac{\epsilon}{\sqrt{n}}e_i}^{\otimes n}\right] - 1\right)$$

$$= \lim_{n \to \infty} \frac{1}{\epsilon^2}\left[\left(1 + \frac{\epsilon^2}{n}\left(\tilde{J}_{t_0}\right)_{i,j} + O\left(\frac{1}{n^2}\right)\right)^n - 1\right]$$

$$= \lim_{n \to \infty} \frac{1}{\epsilon^2}\left[\left(1 + \frac{\epsilon^2}{n}\left(\tilde{J}_{t_0}\right)_{i,j} + O\left(\frac{1}{n^2}\right)\right)^n - 1\right]$$

$$= \frac{1}{\epsilon^2}\left[e^{\epsilon^2\left(\tilde{J}_{t_0}\right)_{i,j}} - 1\right],$$

which implies (98). □

6. Precision Bounds for Multiparameter Estimation

6.1. Covariance conditions. First, we introduce the condition for our estimators. The correspondence between qudit states and Gaussian states also extends to the estimator level. We consider a generic state family $\mathcal{M} = \{\rho_t\}_{t \in \Theta}$, with the parameter space Θ being an open subset of $\mathbb{R}^k$. Similar to the single-parameter case, given a point $t_0 \in \Theta$, we consider a local model $\rho^n_{t_0,t} := \rho^{\otimes n}_{t_0 + t/\sqrt{n}}$. Throughout this section, we assume that ρ_{t_0} is invertible. For a sequence of POVM $\mathfrak{m} := \{M_n\}$, we introduce the condition of local asymptotic covariance as follows:

Condition 2 (Local asymptotic covariance). *We say that a sequence of measurements* $\mathfrak{m} := \{M_n\}$ *satisfies local asymptotic covariance at* $t_0 \in \Theta$ *under the state family* $\mathcal{M}$, *if the probability distribution*

$$\wp^n_{t_0,t|M_n}(\mathsf{B}) := \mathrm{Tr}\, \rho^{\otimes n}_{t_0 + \frac{t}{\sqrt{n}}} M_n \left(\frac{\mathsf{B}}{\sqrt{n}} + t_0 \right) \tag{100}$$

converges to a limiting distribution

$$\wp_{t_0,t|\mathfrak{m}}(\mathsf{B}) := \lim_{n \to \infty} \wp^n_{t_0,t|M_n}(\mathsf{B}), \tag{101}$$

the relation

$$\wp_{t_0,t|\mathfrak{m}}(\mathsf{B} + t) = \wp_{t_0,0|\mathfrak{m}}(\mathsf{B}) \tag{102}$$

holds for any $t \in \mathbb{R}^k$.[2] *When we need to express the outcome of* $\wp^n_{t_0,t|M_n}$ *or* $\wp_{t_0,t|\mathfrak{m}}$, *we denote it by* $\hat{t}$.

Further, we say that a sequence of measurements $\mathfrak{m} := \{M_n\}$ *satisfies local asymptotic covariance under the state family* $\mathcal{M}$ *when it satisfies local asymptotic covariance at any element* $t_0 \in \Theta$ *under the state family* $\mathcal{M}$.

Under these preparations, we obtain the following theorem by using Theorem 3.

Theorem 5. *Let* $\{\rho^{\otimes n}_t\}_{t \in \Theta}$ *be a k-parameter D-invariant qudit model with* C^2 *continuous parametrization. Assume that* $\tilde{J}^{-1}_{t_0}$ *exists,* ρ_{t_0} *is a non-degenerate state, and a sequence of measurements* $\mathfrak{m} := \{M_n\}$ *satisfies local asymptotic covariance at* $t_0 \in \Theta$. *Then there exists a covariant POVM* $\widetilde{M}^G$ *such that*

$$\mathrm{Tr}\, \widetilde{M}^G(\mathsf{B}) G[t, \tilde{J}^{-1}_{t_0}] = \wp_{t_0,t|\mathfrak{m}}(\mathsf{B}) \tag{103}$$

for any vector t *and any measurable subset* B. *Here* $\tilde{J}_{t_0}$ *is the RLD Fisher information of the qudit model at* t_0.

To show Theorem 5, we will use the following lemma.

Lemma 12. *For a function* f, *an operator* F, *and a c–q Gaussian state in the canonical form* $G[\boldsymbol{\alpha}, \Gamma]$, *the relation*

$$\mathrm{Tr}\, F G[\boldsymbol{\alpha}, \Gamma] = f(\boldsymbol{\alpha}) \tag{104}$$

---

[2] The range of t is determined via the constraint $t_0 + t/\sqrt{n} \in \Theta$. Just as in the one-parameter case, t can take any value in $\mathbb{R}^k$ when n is large enough. The range of the local parameter is then $t \in \mathbb{R}^k$.

holds for any vector $\boldsymbol{\alpha}$ if and only if

$$F = \int d\boldsymbol{y}\, \mathcal{F}^{-1}_{\boldsymbol{\xi} \to \boldsymbol{y}} \left(\sqrt{\pi^k} e^{\frac{1}{4} \sum_j \frac{\xi_j^2}{1-\gamma_j}} \mathcal{F}_{\boldsymbol{\alpha} \to \boldsymbol{\xi}} \left(f(\boldsymbol{\alpha}) \right) \right) |\boldsymbol{y}\rangle\langle\boldsymbol{y}|. \tag{105}$$

Here $\boldsymbol{\xi}$ and $\boldsymbol{y}$ are k-dimensional vectors, $|\boldsymbol{y}\rangle$ is a (multimode) coherent state, γ_j are thermal parameters of the Gaussian, and $\mathcal{F}^{-1}_{\boldsymbol{\xi} \to \boldsymbol{y}}(g)$ denotes the inverse of the Fourier transform $\mathcal{F}_{\boldsymbol{\xi} \to \boldsymbol{y}}(g) := \int d\boldsymbol{\xi}\, e^{i\boldsymbol{\xi} \cdot \boldsymbol{y}} g$. Therefore, for a given function $f(\boldsymbol{\alpha})$, there uniquely exists an operator F to satisfy (104).

The proof can be found in "Appendix G". Now, we are ready to prove Theorem 5.

Proof of Theorem 5. We consider without loss of generality $G[\boldsymbol{t}, \tilde{J}^{-1}_{\boldsymbol{t}_0}]$ to be in the canonical form, noticing that any Gaussian state is unitarily equivalent to a Gaussian state in the canonical form as shown by Lemma 3. For any measurable set B, we define the operator $\widetilde{M}^G(\mathsf{B})$ as

$$\widetilde{M}^G(\mathsf{B}) := \int d\boldsymbol{y}\, \mathcal{F}^{-1}_{\boldsymbol{\xi} \to \boldsymbol{y}} \left(\sqrt{\pi^k} e^{\frac{1}{4} \sum_j \frac{\xi_j^2}{1-\gamma_j}} \mathcal{F}_{\boldsymbol{t} \to \boldsymbol{\xi}} \left(\wp_{\boldsymbol{t}_0, \boldsymbol{t}|\mathsf{m}}(\mathsf{B}) \right) \right) |\boldsymbol{y}\rangle\langle\boldsymbol{y}|. \tag{106}$$

From the above definition, it can be verified that $\widetilde{M}^G(\mathsf{B})$ satisfies the definition of a POVM: first, it is immediate to see that the term $\mathcal{F}^{-1}_{\boldsymbol{\xi} \to \boldsymbol{y}} \left(\sqrt{\pi^k} e^{\frac{1}{4} \sum_j \frac{\xi_j^2}{1-\gamma_j}} \mathcal{F}_{\boldsymbol{t} \to \boldsymbol{\xi}} \left(\wp_{\boldsymbol{t}_0, \boldsymbol{t}|\mathsf{m}}(\mathsf{B}) \right) \right)$

equals the convolution of $\wp_{\boldsymbol{t}_0, \boldsymbol{t}|\mathsf{m}}(\mathsf{B})$ and a Gaussian function by employing the convolution theorem. Since both functions are non-negative, the outcome of convolution is still non-negative, which implies that $\widetilde{M}^G(\mathsf{B}) \geq 0$. Second, by linearity we have $\widetilde{M}^G(\mathsf{B}_1 \cup \mathsf{B}_2) = \widetilde{M}^G(\mathsf{B}_1) + \widetilde{M}^G(\mathsf{B}_2)$ for any disjoint sets B_1 and B_2. Finally, the equality $\widetilde{M}^G(\mathsf{B})$ can be shown by combining the linearity with the fact that $\wp_{\boldsymbol{t}_0, \boldsymbol{t}|\mathsf{m}}$ is a probability distribution function.

Moreover, Lemma 12 guarantees that

$$\operatorname{Tr} \widetilde{M}^G(\mathsf{B}) G[\boldsymbol{t}, \tilde{J}^{-1}_{\boldsymbol{t}_0}] = \wp_{\boldsymbol{t}_0, \boldsymbol{t}|\mathsf{m}}(\mathsf{B}). \tag{107}$$

What remains to be shown is that the POVM $\{\widetilde{M}^G(\mathsf{B})\}$ satisfies the covariance condition. Eq. (107) guarantees that

$$\operatorname{Tr} T_{\boldsymbol{t}'} \widetilde{M}^G(\mathsf{B}) T^\dagger_{\boldsymbol{t}'} G[\boldsymbol{t}, \tilde{J}^{-1}_{\boldsymbol{t}_0}] = \operatorname{Tr} \widetilde{M}^G(\mathsf{B}) G[\boldsymbol{t} - \boldsymbol{t}', \tilde{J}^{-1}_{\boldsymbol{t}_0}] = \wp_{\boldsymbol{t}_0, \boldsymbol{t} - \boldsymbol{t}'|\mathsf{m}}(\mathsf{B}),$$

and

$$\operatorname{Tr} \widetilde{M}^G(\mathsf{B} + \boldsymbol{t}') G[\boldsymbol{t}, \tilde{J}^{-1}_{\boldsymbol{t}_0}] = \wp_{\boldsymbol{t}_0, \boldsymbol{t}|\mathsf{m}}(\mathsf{B} + \boldsymbol{t}') = \wp_{\boldsymbol{t}_0, \boldsymbol{t} - \boldsymbol{t}'|\mathsf{m}}(\mathsf{B}).$$

The uniqueness of the operator to satisfy the condition (104) implies the covariance condition

$$\widetilde{M}^G(\mathsf{B} + \boldsymbol{t}') = T_{\boldsymbol{t}'} \widetilde{M}^G(\mathsf{B}) T^\dagger_{\boldsymbol{t}'}.$$

$\square$

6.2. MSE bound for the D-invariant case. Next, we derive the lower bound of MSE of the limiting distribution for any D-invariant model. As an extension of the mean square error, we introduce the *mean square error matrix (MSE matrix)*, defined as

$$V_{i,j}[\wp] := \int x_i x_j \wp \, (dx) \tag{108}$$

for a generic probability distribution $\wp$. Since the set of symmetric matrices is not totally ordered, we will consider the minimization of the expectation value tr $WV[\wp_{t_0,t|\mathrm{m}}]$ for a certain weight matrix $W \geq 0$. For short, we will refer to the quantity tr $WV[\wp_{t_0,t|\mathrm{m}}]$ as the *weighted MSE*.

Under local asymptotic covariance, one can derive lower bounds on the covariance matrix of the limiting distribution and construct optimal measurements to achieve them. In general, the attainability of the conventional quantum Cramér–Rao bounds is a challenging issue. For instance, a well-known bound is the symmetric logarithmic derivative (SLD) Fisher information bound

$$\text{tr } WV[\wp_{t_0,t|\mathrm{m}}] \geq \text{tr } W J_{t_0}^{-1}, \tag{109}$$

where J_{t_0} is the SLD Fisher information. The SLD bound is attainable in the single-parameter case, i.e. when $k = 1$, yet it is in general not attainable for multiparameter estimation (see, for instance, later in Sect. 10.1 for a concrete example).

In the following, we derive an *attainable* lower bound on the weighted MSE. To this purpose, we define the set $\mathsf{LAC}(t_0)$ of local asymptotic covariant sequences of measurements at the point $t_0 \in \Theta$. For a model $\mathcal{M}$, we focus on the minimum value

$$\mathcal{C}_{\mathcal{S}}(W, t_0) := \min_{\mathrm{m}' \in \mathsf{LAC}(t_0)} \text{tr } WV[\wp_{t_0,t|\mathrm{m}'}]. \tag{110}$$

When $k \geq 2$, a better choice is the RLD quantum Fisher information bound. The main result of this section is an attainable bound on the weighted MSE, relying on the RLD quantum Fisher information.

Theorem 6 (Weighted MSE bound for D-invariant models). *Assume that $\tilde{J}_{t_0}^{-1}$ exists. Consider any sequence of locally asymptotically covariant measurements $\mathrm{m} := \{M_n\}$. The limiting distribution is evaluated as*

$$V[\wp_{t_0,t|\mathrm{m}}] \geq (\tilde{J}_{t_0})^{-1}, \tag{111}$$

where $\tilde{J}_{t_0}$ is the RLD quantum Fisher information. When the model is C^1 continuous and D-invariant, we have the bound for the weighted MSE with weight matrix $W \geq 0$ of the limiting distribution as

$$\text{tr } WV[\wp_{t_0,t|\mathrm{m}}] \geq \text{tr } W J_{t_0}^{-1} + \frac{1}{2} \text{tr } \left| \sqrt{W} J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} \sqrt{W} \right|, \tag{112}$$

where J_{t_0} is the SLD quantum Fisher information (35) and D_{t_0} is the D-matrix (41). When $\mathcal{S}$ is a D-invariant qudit model and the state ρ_{t_0} is not degenerate, we have

$$\mathcal{C}_{\mathcal{S}}(W, t_0) = \text{tr } W J_{t_0}^{-1} + \frac{1}{2} \text{tr } \left| \sqrt{W} J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} \sqrt{W} \right|. \tag{113}$$

Moreover, if $W > 0$ and $\wp_{t_0,0|\mathrm{m}}$ has a differentiable PDF, the equality in (112) *holds if and only if $\wp_{t_0,t|\mathrm{m}}$ is the normal distribution with average zero and covariance*

$$V_{t_0|W} := J_{t_0}^{-1} + \frac{1}{2}\sqrt{W}^{-1} \left| \sqrt{W} J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} \sqrt{W} \right| \sqrt{W}^{-1}. \tag{114}$$

Further, when $\{\rho_t\}_{t\in\Theta}$ is a qudit-model with C^2 continuous parametrization, the equality in (112) *holds, i.e., there exist a sequence of POVMs $M_W^{t_0,n}$, a compact set K, and constant $c(t_0)$ such that*

$$\limsup_{n\to\infty} \sup_{t_0\in K} \sup_{t\in\Theta_{n,x,c(t_0)}} n^\kappa \|\wp_{t_0,t|M_W^{t_0,n}}^n - N[t, V_{t_0|W}]\|_1 < \infty, \tag{115}$$

where κ is a parameter to satisfy $\kappa \geq 0.027$.

In the following, we prove Theorem 6 following three steps. The first step is to derive the bound (112). The second step is to show that, to achieve the equality, the limiting distribution needs to be a Gaussian with certain covariance. The last step is to find a measurement attaining the equality. In this way, when the state is not degenerate, we can construct the measurement using Q-LAN.[3]

Proof of Theorem 6. Impossibility part[4] (Proofs of (111) *and* (112)*):*
To give a proof, we focus on the ϵ-difference RLD Fisher information matrix $\tilde{J}_{t_0,\epsilon}$ at t_0 for a quantum states family $\{\rho_t\}_{t\in\Theta}$. We denote the ϵ-difference Fisher information matrices for the distribution family $\{\wp_{t_0,t|M_n}^n\}_t$ and $\{\wp_{t_0,t|\mathrm{m}}\}_t$ by $J_{t,\epsilon}^n$ and $J_{t,\epsilon}^\mathrm{m}$, respectively. Also, we employ the notations given Sect. 5.4.

Applying (90) to the POVM M_n, we have

$$\frac{1}{n}\tilde{J}_{t_0,\epsilon/\sqrt{n}}^n \geq J_{0,\epsilon}^n, \tag{116}$$

where $\tilde{J}_{t_0,\epsilon/\sqrt{n}}^n$ denotes the $(\epsilon/\sqrt{n})$-difference RLD Fisher information matrix of the n-copy states $\{\rho_t^{\otimes n}\}_{t\in\Theta}$. Combining the above with (98) of Lemma 11, we find that $J_{0,\epsilon}^n < (\tilde{J}_{t_0}^{[\epsilon]} + I)$ for sufficiently large n. Hence, we can apply Lemma 21 (see "Appendix B") to the distribution family $\{\wp_{t_0,t|M_n}^n\}_t$. Then, for any complex vector a, we have

$$\liminf_{n\to\infty} \langle a | J_{0,\epsilon}^n - J_{0,\epsilon}^\mathrm{m} | a \rangle \geq 0. \tag{117}$$

By taking the limit $n \to \infty$, the combination of (116), (98) of Lemma 11, and (117) implies

$$\tilde{J}_{t_0}^{[\epsilon]} \geq J_{0,\epsilon}^\mathrm{m}. \tag{118}$$

Here, in the same way as the proof of Theorem 2, we can assume that the outcome $\hat{t}$ satisfies the unbiasedness condition. Hence, the ϵ-difference Carmér–Rao inequality (89) implies that

$$V[\wp_{t_0,t|\mathrm{m}}] \geq \left(J_{0,\epsilon}^\mathrm{m}\right)^{-1} \geq \left(\tilde{J}_{t_0}^{[\epsilon]}\right)^{-1}. \tag{119}$$

---

[3] We note that state estimation involving degenerate states was only solved in a few special cases (see, e.g., [45] for the case of maximally mixed qubits).

[4] Note that when the state is not degenerate the proof of the impossibility part of Theorem 6 can be simplified using the correspondence between n-copy qudit states and Gaussian states as established by the Q-LAN.

By taking the limit $\epsilon \to 0$, (99) of Lemma 11 implies

$$V[\wp_{t_0,t|\mathrm{m}}] \geq \left(\tilde{J}_{t_0}\right)^{-1}. \tag{120}$$

When the model is C^1 continuous and D-invariant, adding the conventional discussion for MSE bounds (see, e.g., Chapter 6 of [5]) to (119), we obtain (112).

Achievability part (Proof of (113)):

Next, we discuss the attainability of the bound when $W > 0$ and $\wp_{t_0,0|\mathrm{m}}$ has a differentiable PDF. In this case, we have the Fisher information matrix $J_{\mathbf{0}}^{\mathrm{m}}$ of the location shift family $\{\wp_{t_0,t|\mathrm{m}}\}_t$. Taking limit $\epsilon \to 0$ in (119), we have

$$V[\wp_{t_0,t|\mathrm{m}}] \geq \left(J_{\mathbf{0}}^{\mathrm{m}}\right)^{-1} \geq \left(\tilde{J}_{t_0}\right)^{-1}. \tag{121}$$

The equality of (112) holds if and only if $V[\wp_{t_0,t|\mathrm{m}}] = V_{t_0|W}$ and the equality in the first inequality of (121) holds. Due to the same discussion as the proof of Theorem 2, the equality in the first inequality of (121) holds only when all the components of the logarithmic derivative of the distribution family $\{\wp_{t_0,t|\mathrm{m}}\}_t$ equal the linear combinations of the estimate of t_i. This condition is equivalent to the condition that the distribution family $\{\wp_{t_0,t|\mathrm{m}}\}_t$ is a distribution family of shifted normal distributions. Therefore, when $W > 0$, the equality condition of Eq. (112) is that $\wp_{t_0,t|\mathrm{m}}$ is the normal distribution with average zero and covariance matrix $V_{t_0|W}$.

Now, we assume that the state ρ_{t_0} is not degenerate. Then, we use Q-LAN to show that there always exists a sequence of POVM $\mathrm{m} = \{M_n\}$ satisfying the above property. We rewrite Eq. (77) of Theorem 3 as follows.

$$\limsup_{n \to \infty} \sup_{t_0 \in K} \sup_{t \in \Theta_{n,x,c(t_0)}} n^\kappa \left\| \mathcal{T}_n^{\mathrm{QLAN}}\left(\rho_{t_0+t/\sqrt{n}}^{\otimes n}\right) - G[t, \tilde{J}_{t_0}^{-1}] \right\|_1 < \infty, \tag{122}$$

where the notation is the same as Theorem 3. Then, we choose the covariant POVM $M_W^{\tilde{J}_{t_0}^{-1}}$ on the Gaussian system given in Lemma 6. When the POVM $M_W^{\tilde{J}_{t_0}^{-1}}$ applied to the system with the state $G[\boldsymbol{\theta}, \Gamma]$, the outcome obeys the distribution $N[t, V_{t_0|W}]$. Therefore, due to (122), when we choose m to be the sequence of POVMs $M_n(\mathsf{B}) := (\mathcal{T}_n^{\mathrm{QLAN}})^\dagger (M_W^{\tilde{J}_{t_0}^{-1}}(\mathsf{B}))$ satisfies the condition (115).

Notice that when W has null eigenvalues, $\sqrt{W}^{-1}$ is not properly defined. In this case, we consider $W_\epsilon := W + \epsilon \cdot P_W^\perp$ for $\epsilon > 0$, where $P_W^\perp$ is the projection on the kernel of W, i.e. $P_W^\perp \boldsymbol{x} = \boldsymbol{x}$ if and only if $W\boldsymbol{x} = 0$. Denoting by $J_\epsilon^{-1} := J_{t_0}^{-1} + \frac{1}{2}\sqrt{W_\epsilon}^{-1} \left|\sqrt{W_\epsilon} J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} \sqrt{W_\epsilon}\right| \sqrt{W_\epsilon}^{-1}$ the

$$\mathrm{tr}\, W J_\epsilon^{-1} \leq \mathrm{tr}\, W_\epsilon J_\epsilon^{-1} - \epsilon\, \mathrm{tr}\, J_\epsilon^{-1}.$$

Meanwhile, since $W_\epsilon > 0$ we can repeat the above argument to find a qudit measurement that attains $\mathrm{tr}\, W_\epsilon J_\epsilon^{-1}$. Taking the limit $\epsilon \to 0$ the quantity $\mathrm{tr}\, W J_\epsilon^{-1}$ converges to the equality of Eq. (113). Therefore, we can still find a sequence of measurements with Fisher information $\{J_\epsilon\}$ that approaches the bound. $\quad\square$

6.3. Precision bound for the estimation of generic models. In the previous subsection, we established the precision bound for D-invariant models, where the bound is attainable and has a closed form. Here we extend the bound to any n-copy qudit models. The main idea is to *extend* the model to a larger D-invariant model by introducing additional parameters.

When estimating parameters in a generic model $\mathcal{S}$ (consisting of states generated by noisy evolutions, for instance), the bound (112) may not hold. It is then convenient to *extend* the model to a D-invariant model $\mathcal{S}'$ which contains $\mathcal{S}$. Since the bound (112) holds for the new model $\mathcal{S}'$, a corresponding bound can be derived for the original model $\mathcal{S}$. The new model $\mathcal{S}'$ has some additional parameters other than those of $\mathcal{S}$, which are fixed in the original model $\mathcal{S}$. Therefore, a generic quantum state estimation problem can be regarded as an estimation problem in a D-invariant model with *fixed parameters*. The task is to estimate parameters in a model $\mathcal{S}'$ (globally) parameterized as $t'_0 = (t_0, p_0) \in \Theta'$, where p_0 is a fixed vector and Θ' is an open subset of $\mathbb{R}^{k'}$ that equals Θ when restricted to $\mathbb{R}^k$. In the neighborhood of t'_0, since the vector p_0 is fixed, we have $t' = (t, 0)$ with 0 being the null vector of $\mathbb{R}^{k'-k}$ and $t \in \mathbb{R}^k$ being a vector of free parameters. For this scenario, only the parameters in t need to be estimated and we know the parameters p_0. Hence, the MSE of t' is of the form

$$V[\wp_{t'_0, t'|\mathfrak{m}}] = \begin{pmatrix} V[\wp_{t_0, t|\mathfrak{m}}] & 0 \\ 0 & 0 \end{pmatrix}$$

for any local asymptotic covariant measurement sequence $\mathfrak{m}$. Due to the block diagonal form of the MSE matrix, to discuss the weight matrix W in the original model $\mathcal{S}$, we consider the weight matrix $W' = P^T W P$ in the D-invariant model $\mathcal{S}'$, where P is any $k \times k'$ matrix satisfying the constraint $(P)_{ij} := \delta_{ij}$ for $i, j \leq k$ in the following way.

Theorem 7 (MSE bound for generic models). *The models $\mathcal{S}$ and $\mathcal{S}'$ are C^1 continuous and are given in the same way as Proposition 1, and the notations are the same as Proposition 1. Also, we assume that $\tilde{J}_{t_0}^{-1}$ exists. Consider any sequence of locally asymptotically covariant measurements $\mathfrak{m} := \{M_n\}$. Then, the MSE matrx of the limiting distribution is evaluated as $\wp_{t_0, t|\mathfrak{m}}$. There exists a $k \times k'$ matrix P such that*

$$P_{ij} = \delta_{ij} \text{ for } i, j \leq k \tag{123}$$

$$V[\wp_{t_0, t|\mathfrak{m}}] \geq P(\tilde{J}_{t'_0})^{-1} P^T, \tag{124}$$

where $\tilde{J}_{t'_0}$ is the RLD quantum Fisher information of S'. When the model $\mathcal{M}'$ is a D-invariant model, we have the bound for the weighted MSE with weight matrix $W \geq 0$ of the limiting distribution as

$$\operatorname{tr} W V[\wp_{t_0, t|\mathfrak{m}}] \geq \mathcal{C}_{\mathrm{H}, \mathcal{M}}(W, t_0) \tag{125}$$

with $\mathcal{C}_{\mathrm{H}, \mathcal{M}}(W, t_0)$ defined in Eq. (46). When the model $\mathcal{M}'$ is a D-invariant qudit model and the state ρ_{t_0} is not degenerate, we have

$$\mathcal{C}_{\mathcal{M}}(W, t_0) = \mathcal{C}_{\mathrm{H}, \mathcal{M}}(W, t_0). \tag{126}$$

Moreover, if $W > 0$ and $\wp_{t_0, 0|\mathfrak{m}}$ has a differentiable PDF, the equality in (125) holds if and only if $\wp_{t_0, t|\mathfrak{m}}$ is the normal distribution with average zero and covariance matrix $V_{t_0|W} := \mathsf{Re}(Z_{t_0}(X)) + \sqrt{W}^{-1} |\sqrt{W} \mathsf{Im}(Z_{t_0}(X)) \sqrt{W}| \sqrt{W}^{-1}$, where X is the vector to

realize the minimum (49). *Further, when the models* $\mathcal{S}$ *and* $\mathcal{S}'$ *are qudit-models with* C^2
continuous parametrization, the equality in (125) *holds, i.e., there exist a sequence of
POVMs* $M_W^{t_0,n}$, *a compact set* K, *and constant* $c(t_0)$ *such that*

$$\limsup_{n \to \infty} \sup_{t_0 \in K} \sup_{t \in \Theta_{n,x,c(t_0)}} n^\kappa \left\| \wp^n_{t_0,t|M_W^{t_0,n}} - N[t, V_{t_0|W}] \right\|_1 < \infty. \tag{127}$$

Theorem 7 determines the ultimate precision limit for generic qudit models. Now, we
compare it with the most general existing bound on quantum state estimation, namely
Holevo's bound [5]. Let us define the ultimate precision of *unbiased* measurements as

$$\mathcal{C}_{\mathsf{UB}_\mathcal{M}}(W, t_0) := \lim_{n \to \infty} \min_{M_n \in \mathsf{UB}_\mathcal{M}} \operatorname{tr} W V \left[\wp^n_{t_0,t|M_n} \right].$$

Since the Holevo bound still holds with the n-copy case, (see [15, Lemma 4]) we have

$$\mathcal{C}_{\mathsf{UB}_\mathcal{M}}(W, t_0) \geq \mathcal{C}_{\mathsf{H},\mathcal{M}}(W, t_0). \tag{128}$$

There are a couple of differences between our results and existing results: The Holevo
bound is derived under unbiasedness assumption, which, as mentioned earlier, is more
restrictive than local asymptotic covariance. Our bound (125) thus applies to a wider
class of measurements than the Holevo bound.

Furthermore, Yamagata et al. [19] showed a similar statement as (127) of Theorem 7 in
a local model scenario. They did not show the compact uniformity of the convergence and
had no order estimation of the convergence. However, our evaluation (127) guarantees
the compact uniformity with the order estimation. Then, they did not discuss an estimator
to attain the bound globally. Later, we will construct an estimator to attain our bound
globally based on the estimator given in Theorem 7. Our detailed evaluation with the
compact uniformity and the order estimation enables us to evaluate the performance of
such an estimator globally.

Proof of Theorem 7. Impossibility part (Proofs of (124) *and* (125)):
We denote the ϵ-difference Fisher information matrices for the distribution family
$\{\wp^n_{t_0,t|M_n}\}_t$ and $\{\wp_{t_0,t|\mathsf{m}}\}_t$ by $J^n_{t,\epsilon}$ and $J^{\mathsf{m}}_{t,\epsilon}$, respectively. Also, we denote the ϵ-difference
type RLD Fisher information matrix at $t'_0 = (t_0, \mathbf{0})$ of the family $\{\rho^{\otimes n}_{t'}\}_{t'}$ by $\tilde{J}^n_{t'_0,\epsilon}$. Then,
we have (117) in the same way.

Applying (97) of Lemma 10 with $\epsilon \to \epsilon/\sqrt{n}$, there exist $k \times k'$ matrices P_n such
that

$$(P_n)_{ij} = \delta_{ij} \qquad \text{for } i, j \leq k \tag{129}$$

$$\frac{1}{n} \left(P_n (\tilde{J}^n_{t'_0,\epsilon/\sqrt{n}})^{-1} P_n^T \right)^{-1} \geq J^n_{t_0,\epsilon}. \tag{130}$$

Hence, the combination of (98) of Lemma 11, (130), and (117) implies that there exists
a $k \times k'$ matrices P such that

$$P_{ij} = \delta_{ij} \qquad \text{for } i, j \leq k \tag{131}$$

$$\left(P (\tilde{J}^{[\epsilon]}_{t'_0})^{-1} P^T \right)^{-1} \geq J^{\mathsf{m}}_{t_0,\epsilon}. \tag{132}$$

Due to the same reason as (119), we have

$$V[\wp_{t_0, t|\mathrm{m}}] \geq (J_{t_0, \epsilon}^{\mathrm{m}})^{-1} \geq P(\tilde{J}_{t_0'}^{[\epsilon]})^{-1} P^T. \tag{133}$$

By taking the limit $\epsilon \to 0$, the combination of (99) of Lemma 11 and (133) implies (124).

When the model $\mathcal{M}'$ is D-invariant, since

$$\mathrm{tr}\, W V[\wp_{t_0, t|\mathrm{m}}] \geq \mathrm{tr}\, P^T W P \tilde{J}_{t_0'}^{-1},$$

we obtain (125) by using the expression (50) in the same way as (112):

Achievability part (Proof of (126)):

Since $\rho_{t_0'}$ is not degenerate, we can show the achievability in the same way as Theorem 6 because we can apply Q-LAN (Theorem 3) for the model $\mathcal{M}'$. The difference is the following. Choosing the matrix P to achieve the minimum (50), we employ the covariant POVM $M_{P|W}^{\tilde{J}_{t_0'}^{-1}}$ instead of the covariant POVM $M_W^{\tilde{J}_{t_0}^{-1}}$. Then, we obtain the desired statement. □

7. Nuisance Parameters

For state estimation in a noisy environment, the strength of noise is not a parameter of interest, yet it affects the precision of estimating other parameters. In this scenario, the strength of noise is a *nuisance parameter* [46,47]. To illustrate the difference between nuisance parameters and fixed parameters that are discussed in the previous section, let us consider the case of a qubit clock state under going a noisy time evolution. To estimate the duration of the evolution, we introduce the strength of the noise as an additional parameter and consider the estimation problem in the extended model parameterized by the duration and the noise strength. The strength of the noise is usually unknown. Although it is not a parameter of interest, its value will affect the precision of our estimation, and thus it should be treated as a *nuisance* parameter.

7.1. Precision bound for estimation with nuisance parameters. In this subsection, we consider state estimation of an arbitrary $(k + s)$-parameter model $\{\rho_{t,p}\}_{(t,p)\in\tilde{\Theta}}$, where t and p are k-dimensional and s-dimensional parameters, respectively. Our task is to estimate only the parameters t and it is not required to estimate the other parameters p, which is called nuisance parameters. Hence, our estimate is k-dimensional. We say that a parametric family of a structure of nuisance parameters is a nuisance parameter model, and denote it by $\tilde{S} = \{\rho_{t,p}\}_{(t,p)\in\tilde{\Theta}}$. We simplify (t, p) by $\tilde{t}$.

The concept of local asymptotic covariance can be extended to a model with nuisance parameters by considering a local model $\rho_{\tilde{t}_0, \tilde{t}}^n := \rho_{\tilde{t}_0 + \tilde{t}/\sqrt{n}}^{\otimes n}$. Throughout this section, we assume that $\rho_{\tilde{t}_0}$ is invertible and all the parametrizations are at least C^1 continuous.

Condition 3 (Local asymptotic covariance with nuisance parameters). *We say that a sequence of measurements $\mathrm{m} := \{M_n\}$ to estimate the k-dimensional parameter t satisfies local asymptotic covariance at $\tilde{t}_0 = (t_0, p_0) \in \tilde{\Theta}$ under the nuisance parameter model $\tilde{\mathcal{M}}$ when the probability distribution*

$$\wp_{\tilde{t}_0, \tilde{t}|M_n}^n(\mathsf{B}) := \mathrm{Tr}\, \rho_{\tilde{t}_0 + \frac{\tilde{t}}{\sqrt{n}}}^{\otimes n} M_n\left(\frac{\mathsf{B}}{\sqrt{n}} + t_0\right) \tag{134}$$

converges to a limiting distribution

$$\wp_{\tilde{t}_0,\tilde{t}|\mathrm{m}}(\mathsf{B}) := \lim_{n\to\infty} \wp^n_{\tilde{t}_0,\tilde{t}|M_n}(\mathsf{B}), \tag{135}$$

the relation

$$\wp_{\tilde{t}_0,\tilde{t}|\mathrm{m}}(\mathsf{B}+t) = \wp_{\tilde{t}_0,(\mathbf{0},\mathbf{0})|\mathrm{m}}(\mathsf{B}) \tag{136}$$

holds for any $\tilde{t} = (t, p) \in \mathbb{R}^{k+s}$.

Further, we say that a sequence of measurements $\mathrm{m} := \{M_n\}$ *satisfies local asymptotic covariance under the nuisance parameter model* $\tilde{\mathcal{M}}$ *when it satisfies local asymptotic covariance at any element* $\tilde{t}_0 \in \tilde{\Theta}$ *under the nuisance parameter model* $\tilde{\mathcal{M}}$.

The quantity we want to bound is $\mathrm{tr}\, V[\wp_{t_0,t|\mathrm{m}}]W$, where $\wp_{t_0,t|\mathrm{m}}$ is the limiting distribution of a sequence m of locally asymptotically covariant measurements and W is a weight matrix. Since nuisance parameters are not of interest, the weight matrix of the model $\tilde{\mathcal{S}}$ is a $(k+s) \times (k+s)$ matrix of the form

$$\tilde{W} = \begin{pmatrix} W & 0 \\ 0 & 0 \end{pmatrix}. \tag{137}$$

Lemma 13. *Let* $\tilde{\mathcal{S}} = \{\rho_{\tilde{t}}\}_{\tilde{t}\in\tilde{\Theta}}$ *be a* $(k+s)$-*parameter nuisance parameter model and let* $\mathcal{S}' = \{\rho_{t'}\}_{t'=(\tilde{t},q)}$ *be a* k'-*parameter model containing* $\tilde{\mathcal{S}}$ *as* $\rho_{\tilde{t}} = \rho_{(\tilde{t},\mathbf{0})}$. *When* $\mathcal{S}'$ *is* D-*invariant and the inverse* $\tilde{J}^{-1}_{\tilde{t}_0}$ *exists, we have*

$$\mathcal{C}_{\mathrm{NH},\mathcal{M}}(W,\tilde{t}_0) := \min_X \min_V \{\mathrm{tr}\, WV | V \geq Z_{\tilde{t}_0}(X)\} \tag{138}$$

$$= \min_P \left\{ \mathrm{tr}\, P^T W P(J^{-1}_{t'_0}) + \frac{1}{2}\mathrm{tr}\left| \sqrt{P^T W P} J^{-1}_{t'_0} D_{t'_0} J^{-1}_{t'_0} \sqrt{P^T W P} \right| \right\}. \tag{139}$$

In (138), V *is a real symmetric matrix and* $X = (X_i)$ *is a* k-*component vector of operators to satisfy*

$$\mathrm{Tr}\, X_i \frac{\partial \rho_{\tilde{t}}}{\partial \tilde{t}_j}\bigg|_{\tilde{t}=\tilde{t}_0} = \delta_{ij}, \quad \forall i \leq k, \quad \forall j \leq k+s. \tag{140}$$

In (139), *the minimization is taken over all* $k \times (k+s)$ *matrices satisfying the constraint* $(P)_{ij} := \delta_{ij}$ *for* $i \leq k$, $j \leq k+s$, *and,* $J_{t'_0}$ *and* $D_{t'_0}$ *are the SLD Fisher information matrix and the D-matrix [cf. Eqs.* (35) *and* (41)*] for the extended model* $\mathcal{S}'$ *at* $t'_0 := (\tilde{t}_0, \mathbf{0})$.

This lemma is a different statement from [15, Theorem 4]. However, using the method of [15, Theorem 4], we can show this lemma.

In the following, we derive an *attainable* lower bound on the weighted MSE. To this purpose, we define the set $\mathsf{LAC}(\tilde{t}_0)$ of local asymptotic covariant sequences of measurements at the point $\tilde{t}_0 \in \tilde{\Theta}$ for the nuisance parameter model $\tilde{\mathcal{M}}$, and focus on the minimum value

$$\mathcal{C}_{\mathrm{N},\tilde{\mathcal{M}}}(W,\tilde{t}_0) := \min_{\mathrm{m}\in\mathsf{LAC}(\tilde{t}_0)} \mathrm{tr}\, WV[\wp_{\tilde{t}_0,\tilde{t}|\mathrm{m}}]. \tag{141}$$

Theorem 8 (Weighted MSE bound with nuisance parameters). *The models $\tilde{\mathcal{S}}$ and $\mathcal{S}'$ are given in the same way as Lemma 13, and the notations is the same as Lemma 13. Also, we assume that $(\tilde{J}_{t'_0})^{-1}$ exists. Consider any sequence of locally asymptotically covariant measurements $\mathfrak{m} := \{M_n\}$ for the nuisance parameter model $\tilde{\mathcal{S}}$. Then, the MSE matrx of the limiting distribution is evaluated as follows. There exists a $k \times k'$ matrix P such that*

$$P_{ij} = \delta_{ij} \quad \text{for } 1 \le i \le k, \, 1 \le j \le k' \tag{142}$$

$$V[\wp_{\tilde{t}_0, \tilde{t}|\mathfrak{m}}] \ge P(\tilde{J}_{t'_0})^{-1} P^T. \tag{143}$$

When the model $\mathcal{S}'$ is D-invariant, we have the bound for the weighted MSE with weight matrix $W \ge 0$ of the limiting distribution as

$$\operatorname{tr} W V[\wp_{\tilde{t}_0, \tilde{t}|\mathfrak{m}}] \ge \mathcal{C}_{\mathrm{NH}, \mathcal{M}}(W, \tilde{t}_0). \tag{144}$$

When the model $\mathcal{S}'$ is a D-invariant qudit model and the state ρ_{t_0} is not degenerate, we have

$$\mathcal{C}_{\mathrm{N}, \tilde{\mathcal{M}}}(W, \tilde{t}_0) = \mathcal{C}_{\mathrm{NH}, \tilde{\mathcal{M}}}(W, \tilde{t}_0). \tag{145}$$

Moreover, if $W > 0$ and $\wp_{\tilde{t}_0, \tilde{t}|\mathfrak{m}}$ has a differentiable PDF, the equality in (144) holds if and only if $\wp_{t_0, t|\mathfrak{m}}$ is the normal distribution with average zero and covariance

$$V_{t_0|W} := \mathsf{Re}(Z_{t_0}(X)) + \sqrt{W}^{-1} |\sqrt{W} \mathsf{Im}(Z_{t_0}(X)) \sqrt{W}| \sqrt{W}^{-1}, \tag{146}$$

where X is the vector to realize the minimum (138). Further, when the models $\tilde{\mathcal{S}}$ and $\mathcal{S}'$ are qudit-models with C^2 continuous parametrization, the equality in (144) holds, i.e., there exist a sequence of POVMs $M_W^{t_0, n}$, a compact set K, and constant $c(t_0)$ such that

$$\limsup_{n \to \infty} \sup_{t_0 \in K} \sup_{t \in \Theta_{n,x,c(t_0)}} n^\kappa \|\wp_{t_0, t|M_W^{t_0, n}}^n - N[t, V_{t_0|W}]\|_1 < \infty. \tag{147}$$

Here κ is a parameter to satisfy $\kappa \ge 0.027$.

Before proving Theorem 8, we discuss a linear subfamily of k'-dimensional Gaussian family $\{G[t', \gamma]\}_{t' \in \mathbb{R}^{k'}}$. Consider a linear map T from $\mathbb{R}^{(k+s)}$ to $\mathbb{R}^{k'}$. We have the subfamily $\tilde{\mathcal{M}} := \{G[T(t, p), \gamma]\}_{(t, p) \in \mathbb{R}^{k+s}}$ as a nuisance parameter model. Then, the covariance condition is extended as follows.

Definition 5. A POVM M is *unbiased* for the nuisance parameter model $\{\rho_{(t, p)}\}$ when

$$t = E_{t, p}(M) := \int x \operatorname{Tr} \rho_{(t, p)} M(dx)$$

holds for any parameter (t, p). A POVM M is a *covariant* estimator for the nuisance parameter model $\{G[T(t, p), \gamma]\}$ when the distribution $\wp_{(t, p)|M}$ satisfies the condition $\wp_{0, 0|M}(\mathsf{B}) = \wp_{t, p|M}(\mathsf{B} + t)$.

Then, we have the following corollary of Lemma 6.

Corollary 3. *For any weight matrix $W \geq 0$, the nuisance parameter model $\tilde{\mathcal{M}} = \{G[T(t, p), \gamma]\}$ with C^1 continuous parametrization satisfies*

$$\inf_{M \in \mathsf{UB}_{\tilde{\mathcal{M}}}} \operatorname{tr} W V_t(M) = \inf_{M \in \mathsf{CUB}_{\tilde{\mathcal{M}}}} \operatorname{tr} W V_t(M) = \mathcal{C}_{\mathrm{NH},\tilde{\mathcal{M}}}(W, t), \quad (148)$$

where $\mathsf{UB}_{\tilde{\mathcal{M}}}$ and $\mathsf{CUB}_{\tilde{\mathcal{M}}}$ are the sets of unbiased estimators and covariant unbiased estimators of the nuisance parameter model $\tilde{\mathcal{M}}$, respectively. Further, when $W > 0$, we choose a vector X to realize the minimum in (49). The above infimum is attained by the covariant unbiased estimators M_W whose output distribution is the normal distribution with average t and covariance matrix $\operatorname{Re}((Z_t(X)) + \sqrt{W}^{-1} |\sqrt{W} \operatorname{Im}(Z_t(X)) \sqrt{W}| \sqrt{W}^{-1}$.

This corollary can be shown as follows. The inequality $\inf_{M \in \mathsf{UB}_{\tilde{\mathcal{M}}}} \operatorname{tr} W V_t(M) \geq \mathcal{C}_{\mathrm{NH},\tilde{\mathcal{M}}}(W, t)$ follows from the condition (140). Similar to Corollary 2, Proposition 1 guarantees that the latter part of the corollary with $W > 0$ follows from (138) and Lemma 6. Hence, we obtain this corollary for $W > 0$. The case with non strictly positive W can be shown by considering W_ϵ in the same way as Corollary 1.

Proof of Theorem 8. Impossibility part (Proofs of (143) and (144)):
We denote the ϵ-difference Fisher information matrix of $\{\wp_{\tilde{t}_0, \tilde{t}|\mathrm{m}}\}_{\tilde{t}}$ by $J_{\tilde{t}_0, \epsilon}^{\mathrm{m}}$. Due to (132), there exists a $(k + s) \times k'$ matrix $\tilde{P}$ satisfying the following conditions.

$$\tilde{P}_{ij} = \delta_{ij} \text{ for } 1 \leq i, j \leq k + s, \quad \left(P(\tilde{J}_{t_0'}^{[\epsilon]})^{-1} P^T \right)^{-1} \geq J_{t_0, \epsilon}^{\mathrm{m}}. \quad (149)$$

We define the $k \times (k + s)$ matrix $\bar{P}$ by

$$\bar{P}_{ij} = \delta_{ij} \text{ for } 1 \leq i \leq k, 1 \leq j \leq k + s.$$

Now, we extend Theorem 4. Let $E_{(0,0)}[\hat{t}]$ be the expectation of $\hat{t}$ under the distribution $\wp_{t_0, (0,0)|\mathrm{m}}$. We denote the Radon–Nikodým derivative of $\wp_{\tilde{t}_0, \epsilon e_i|\mathrm{m}}$ with respect to $\wp_{\tilde{t}_0, 0|\mathrm{m}}$ by p_i. Then, for two vectors $a \in \mathbb{R}^k$ and $b \in \mathbb{R}^{k+s}$, we apply Schwartz inequality to the two variables $X := \sum_j b_j (\hat{t} - E_{(0,0)}[\hat{t}])_j$ and $Y := \sum_i \frac{a_i}{\epsilon}(p_j(x) - 1)$. Using $\tilde{X} := \sum_j b_j t_j$, we obtain

$$\langle b | V[\wp_{\tilde{t}_0, (0,0)|\mathrm{m}}]|b \rangle \langle a | J_{t_0, \epsilon}^{\mathrm{m}} | a \rangle = \int \tilde{X}(x)^2 \wp_{\tilde{t}_0, (0,0)|\mathrm{m}}(dx) \int Y(x)^2 \wp_{\tilde{t}_0, (0,0)|\mathrm{m}}(dx)$$

$$\geq \int X(x)^2 \wp_{\tilde{t}_0, (0,0)|\mathrm{m}}(dx) \int Y(x)^2 \wp_{\tilde{t}_0, (0,0)|\mathrm{m}}(dx)$$

$$\geq \left| \int X(x) Y(x) \wp_{\tilde{t}_0, (0,0)|\mathrm{m}}(dx) \right|^2 = \left| \langle b | \bar{P} | a \rangle \right|^2, \quad (150)$$

where the final equation follows from the fact that the expectation of the variable $\hat{t} - E_{(0,0)}[\hat{t}]$ equals $t = \bar{P} \tilde{t}$ under the distribution $\wp_{\tilde{t}_0, \tilde{t}|\mathrm{m}}$, which can be shown by the covariance condition for the distribution family $\{\wp_{\tilde{t}_0, \tilde{t}|\mathrm{m}}\}_{\tilde{t}}$.

Choosing $a := (J_{\tilde{t}_0, \epsilon}^{\mathrm{m}})^{-1} \bar{P}^T b$, we have

$$\langle b | V[\wp_{\tilde{t}_0, (0,0)|\mathrm{m}}]|b \rangle \geq \langle b | \bar{P} (J_{\tilde{t}_0, \epsilon}^{\mathrm{m}})^{-1} \bar{P}^T | b \rangle, \quad (151)$$

which implies

$$V[\wp_{\tilde{t}_0, \tilde{t}|\mathrm{m}}] = V[\wp_{\tilde{t}_0, (\mathbf{0},\mathbf{0})|\mathrm{m}}] \geq \bar{P}(J^{\mathrm{m}}_{\tilde{t}_0, \epsilon})^{-1} \bar{P}^T. \tag{152}$$

Combining the above with Eq. (149), since $P := \bar{P}\tilde{P}$ satisfies the condition (142), we obtain (143).

When the model $\mathcal{M}'$ is D-invariant, since

$$\mathrm{tr}\, W V[\wp_{\tilde{t}_0, \tilde{t}|\mathrm{m}}] \geq \mathrm{tr}\, P^T W P \tilde{J}^{-1}_{t'_0},$$

we obtain (144) by using the expression (139) in the same way as (125):

Achievability part (Proof of (145));

Since ρ_{t_0} is not degenerate, we can show the achievability part in the same way as Theorem 6 because we can apply Q-LAN (Theorem 3) for the model $\mathcal{M}'$. The difference is the following. Instead of Corollary 1, we employ Corollary 3 to choose the covariant POVM $M^{\tilde{J}^{-1}_{t'_0}}_{P|W}$. Then, we obtain the desired statement.   □

7.2. Nuisance parameter with D-invariant model. Next, we discuss the nuisance parameters when the model is D-invariant.

Lemma 14. *When* $\tilde{\mathcal{S}} = \{\rho_{(\mathbf{t}, \mathbf{p})}\}_{(\mathbf{t}, \mathbf{p}) \in \tilde{\Theta}}$ *is a D-invariant $k + s$-parameter nuisance parameter model and $J^{-1}_{t_0}$ exists, we have*

$$\mathcal{C}_{\mathrm{NH}, \tilde{\mathcal{M}}}(W, \tilde{t}_0) = \mathrm{tr}\, \tilde{W}(J^{-1}_{\tilde{t}_0}) + \frac{1}{2} \mathrm{tr}\, \left| \sqrt{\tilde{W}} J^{-1}_{\tilde{t}_0} D_{\tilde{t}_0} J^{-1}_{\tilde{t}_0} \sqrt{\tilde{W}} \right|. \tag{153}$$

A few comments are in order. First, the nuisance parameter bound (144) reduces to the bound (112), when the parameters to estimate are *orthogonal* to the nuisance parameters in the sense that the RLD Fisher information matrix $\tilde{J}_{\tilde{t}_0}$ is block-diagonal. This orthogonality is equivalent to the condition that the SLD Fisher information matrix $J_{\tilde{t}_0}$ and the D-matrix take the block diagonal forms

$$J_{\tilde{t}_0} = \begin{pmatrix} J_{t_0} & 0 \\ 0 & J_{\mathrm{N}} \end{pmatrix} \quad D_{\tilde{t}_0} = \begin{pmatrix} D_{t_0} & 0 \\ 0 & D_{\mathrm{N}} \end{pmatrix}. \tag{154}$$

This is the case, for instance, of simultaneous estimation of the spectrum and the Hamiltonian-generated phase of a two-level system. Under such circumstances, the inverse of the Fisher information matrix can be done by inverting J_{t_0} and J_{N} independently. The same precision bound is thus obtained with or without introducing nuisance parameters, and we have the following lemma.

Lemma 15. *When all nuisance parameters are orthogonal to the parameters of interest, the bound with nuisance parameters (144) coincides with the D-invariant MSE bound (112).*

In the case of orthogonal nuisance parameters, the estimation of nuisance parameters does not affect the precision of estimating the parameters of interest, which does not hold for the generic case of non-orthogonal nuisance parameters. Thanks to this fact, one can achieve the bound (144) by first measuring the nuisance parameters and then constructing the optimal measurement based on the estimated value of the nuisance

parameters. On the other hand, an RLD bound [cf. Eq. (39)] can be attained if and only if its model is D-invariant. Combining these arguments with Lemma 15, we obtain a characterization of the attainability of RLD bounds as follows.

Corollary 4. *An RLD bound can be achieved if and only if it has an orthogonal nuisance extension, i.e. Eq. (154) holds for some choice of nuisance parameters.*

The above corollary offers a simple criterion for the important problem of the attainability of RLD bounds. In Sect. 10.3, we will illustrate the application of this criterion with a concrete example.

The bound (144) can be straightforwardly computed even for complex models; for D-invariant models, the SLD operators have an uniform entry-wise expression and one only needs to shot it into a program to yield the bound (144). Moreover, the bound does not rely on the explicit choice of nuisance parameters. To see this, one can consider another parameterization x' of the D-invariant model. The bound (144) comes from the RLD bound for the D-invariant model, and the RLD quantum Fisher information matrices $\widetilde{J}_{t'_0}$ and $\widetilde{J}_{x'_0}$ for two parameterizations are connected by the equation $\widetilde{J}_{t'_0} = A\widetilde{J}_{x'_0}A^T$, where A is the Jacobian $(\partial x'/\partial t')$ at t'_0. Since both parameterizations are extensions of the same model $\mathcal{S}$ satisfying $P_0 t'_0 = P'_0 x'_0 = t_0$, the Jacobian takes the form

$$A = \begin{pmatrix} I_k & A' \\ 0 & A'' \end{pmatrix}.$$

Then we have $\widetilde{J}_{x'_0}^{-1} = A^T \widetilde{J}_{t'_0}^{-1} A$, which implies that the upper-left $k \times k$ blocks of $\widetilde{J}_{x'_0}$ and $\widetilde{J}_{t'_0}$ are equal. The bound (144) thus remains unchanged.

7.3. Precision bound for joint measurements. A useful implication of Theorem 8 is a bound on MSEs of several jointly measured observables. Consider a set $\{O_i\}$ of k bounded observables. The goal is to jointly measure their expectations

$$o_i := \text{Tr} \, \rho \, O_i \qquad i = 1, \dots, k. \tag{155}$$

The main result of this subsection is the following corollary:

Corollary 5. *Define the SLD gap of o_i as*

$$\Delta_{o_i} := \mathbf{MSE}_{o_i} - \left(J^{-1}\right)_{ii}, \tag{156}$$

where $\mathbf{MSE}_{o_i}$ denotes the MSE of o_i under joint measurement and J is the SLD quantum Fisher information. The sum of the SLD gaps for all observables satisfies the attainable bound:

$$\sum_{i=1}^{d} \Delta_{o_i} \geq \frac{1}{2} \text{tr} \left| \left(J^{-1}DJ^{-1}\right)_{k \times k} \right|, \tag{157}$$

where D is the D-matrix.

The right hand side of Eq. (157) is exactly the gap between the SLD bound and the ultimate precision limit. It shows a typical example where the SLD bound is not attainable.

Proof. Substituting W' in Eq. (144) by the projection into the subspace $\mathbb{R}^k$, we obtain a bound for the MSE $\{\mathbf{MSE}_{o_i}\}$ of the limiting distributions:

$$\sum_{i=1}^{d} \mathbf{MSE}_{o_i} \geq \sum_{i=1}^{d} \left(J^{-1} \right)_{ii} + \frac{1}{2} \operatorname{tr} \left| \left(J^{-1} D J^{-1} \right)_{k \times k} \right|. \tag{158}$$

Here J and D are the SLD Fisher information and D-matrix for the extended model, and $(A)_{k \times k}$ denotes the upper-left $k \times k$ block of a matrix A. Substituting the above definition into Eq. (158), we obtain Corollary 5. □

Specifically, for the case of two parameters, the bound (157) reduces to

$$\Delta_{o_1} + \Delta_{o_2} \geq \left| \operatorname{Tr} \rho_{\boldsymbol{\theta}} [\hat{L}_1, \hat{L}_2] \right|, \tag{159}$$

where $\hat{L}_j := \sum_{j=1}^{k'} \left(J_{\boldsymbol{\theta}'} \right)_{ji}^{-1} L_i$ are the SLD operators in the dual space. Next, taking partial derivative with respect to o_j on both sides of Eq. (155) and substituting in the definition of RLD operators, the observables satisfy the orthogonality relation with the SLD operators as

$$\frac{1}{2} \operatorname{Tr} \left(\rho L_j + L_j \rho \right) O_i = \delta_{ij}.$$

By uniqueness of the dual space, we have

$$\hat{L}_i = O_i - o_i I \qquad i = 1, \dots, k'$$

and the bound becomes

$$\Delta_{o_1} + \Delta_{o_2} \geq |\langle [O_1, O_2] | \rangle|. \tag{160}$$

Another bound expressing the tradeoff between Δ_{o_1} and Δ_{o_2} was obtained by Watanabe et al. [48] as

$$\Delta_{o_1} \Delta_{o_2} \geq |\langle [O_1, O_2] \rangle|^2 / 4. \tag{161}$$

Now, substituting O_2 by αO_2 for a variable $\alpha \in \mathbb{R}$ in Eq. (160), we have

$$\Delta_{o_1} + \Delta_{\alpha o_2} = \Delta_{o_1} + \alpha^2 \Delta_{o_2} \geq \alpha |\langle [O_1, O_2] | \rangle|.$$

For the above quadratic inequality to hold for any $\alpha \in \mathbb{R}$, its discriminant must be non-positive, which immediately implies the bound (161). Notice that the bound (161) was derived under asymptotic unbiasedness [48], and thus it was not guaranteed to be attainable. Here, instead, since our bound (160) is always attainable, the bound (161) can also be achieved in any qudit model under the asymptotically covariant condition.

7.4. Nuisance parameters versus fixed parameters. It is intuitive to ask what is the relationship between the nuisance parameter bound (144) and the general bound (125). To see it, let $\mathcal{S} = \{\rho_t\}_{t \in \Theta}$ be a generic k-parameter qudit model and let $\tilde{\mathcal{S}}$ be a $(k + s)$-parameter D-invariant model containing $\mathcal{S}$. When ρ_{t_0} is non-degenerate, we notice that the QCR bound with nuisance parameters (144) can be rewritten as

$$
\mathcal{C}_{\mathcal{M}}(W, t_0) = \text{tr } P_0^T W P_0 (J_{\tilde{t}_0}^{-1}) + \frac{1}{2} \text{tr} \left| \sqrt{P_0^T W P_0} J_{\tilde{t}_0}^{-1} D_{\tilde{t}_0} J_{\tilde{t}_0}^{-1} \sqrt{P_0^T W P_0} \right|, \quad (162)
$$

where P_0 is a $k \times (k+s)$ matrix satisfying the constraint $(P_0)_{ij} := \delta_{ij}$ for any $i, j \leq k+s$. By definition, P_0 is a special case of P, and it follows straightforwardly from comparing Eq. (162) with Eq. (125) that the general MSE bound is upper bounded by the MSE bound for the nuisance parameter case. This observation agrees with the obvious intuition that having additional information on the system is helpful for (or at least, not detrimental to) estimation. At last, since $J_{\tilde{t}_0}$ and $D_{\tilde{t}_0}$ are block-diagonal in the case of orthogonal nuisance parameters, we have

$$
P(J_{\tilde{t}_0})^{-1} P^T = J_{t_0}^{-1} \quad P D_{\tilde{t}_0} P^T = D_{t_0}
$$

for any $k \times (k+s)$ matrix satisfying the constraint $(P)_{ij} := \delta_{ij}$ for $i, j \leq k$. This implies that the general bound (125) coincides with the nuisance parameter bound (144) when the nuisance parameters are orthogonal.

8. Tail Property of the Limiting Distribution

In previous discussions, we focused on the MSE of the limiting distribution. Here, instead, we consider the behavior of the limiting distribution itself. The characteristic property is the tail property: Given a weight matrix $W \geq 0$ and a constant c, we define the tail region $\mathsf{T}_{W,c}(t)$ as

$$
\mathsf{T}_{W,c}(t) := \left\{ x \mid (x^T - t^T) W (x - t) \geq c \right\}.
$$

For a measurement $\mathfrak{m} = \{M_n(\hat{t}_n)\}$, the probability that the estimate $\hat{t}_n$ is in the tail region can be approximated by the tail probability of the limiting distribution, i.e.

$$
\mathbf{Prob}\left(\hat{t}_n \in \mathsf{T}_{W,c}(t) \right) = \wp_{t_0, t \mid \mathfrak{m}} \left(\mathsf{T}_{W,c}(t) \right) + \epsilon_n,
$$

up to ϵ_n being a term vanishing in n. The tail property is usually harder to characterize than the MSE. Nevertheless, here we show that, under certain conditions, there exists a good bound on the tail property of the limiting distribution.

8.1. Tail property of Gaussian shift models. Just like in the previous sections, the tail property of n-copy qudit models can be analyzed by studying the tail property of Gaussian shift models. In this subsection, we first derive a bound on the tail probability of Gaussian shift models. The result has an interest in its own and can be used for further analysis of qudit models using Q-LAN.

Consider a Gaussian shift model $\{G[\boldsymbol{\alpha}, \boldsymbol{\gamma}]\}$ with $G[\boldsymbol{\alpha}, \boldsymbol{\gamma}] = N[\boldsymbol{\alpha}^C, \Gamma] \otimes \Phi$ $[\boldsymbol{\alpha}^Q, \boldsymbol{\beta}]$ and a measurement $M^G(\hat{\alpha})$. Then, define the probability $\wp_{\boldsymbol{\alpha}|M^G}\left(\mathsf{T}_{W,c}(\boldsymbol{\alpha})\right)$, where $\mathsf{T}_{W,c}(\boldsymbol{\alpha})$ is the tail region around $\boldsymbol{\alpha}$ defined as

$$\mathsf{T}_{W,c}(\boldsymbol{\alpha}) := \left\{ x \mid (x^T - \boldsymbol{\alpha}^T)W(x - \boldsymbol{\alpha}) \geq c \right\}.$$

Then, for covariant POVMs, the tail probability is independent of $\boldsymbol{\alpha}$ and is given by:

$$\wp_{\boldsymbol{\alpha}|M^G}\left(\mathsf{T}_{W,c}(\boldsymbol{\alpha})\right) = \operatorname{Tr} N[\mathbf{0}, \Gamma] \otimes \Phi[\mathbf{0}, \boldsymbol{\beta}]M\left(\mathsf{T}_{W,c}(\mathbf{0})\right).$$

When the measurement is covariant, we have the following bound on the tail probability, which can be attained by a certain covariant POVM:

Lemma 16. *Consider a Gaussian model $G[\boldsymbol{\alpha}, \boldsymbol{\gamma}] = N[\boldsymbol{\alpha}^C, \Gamma] \otimes \Phi[\boldsymbol{\alpha}^Q, \boldsymbol{\beta}]$ with s' classical parameters and s pairs of quantum parameters. Assume that a POVM $\{M^G(\mathsf{B})\}_{\mathsf{B} \subset \mathbb{R}^{s'} \times \mathbb{R}^{2s}}$ is covariant and the weight matrix W has the following form;*

$$W = \begin{pmatrix} W^C & & & \\ & w_{s'+1}I_2 & & \\ & & \ddots & \\ & & & w_{s'+s}I_2 \end{pmatrix} \tag{163}$$

with $W^C \geq 0$. Then, the tail probability of the limiting distribution is bounded as

$$\wp_{\boldsymbol{\alpha}|M^G}\left(\mathsf{T}_{W,c}(\boldsymbol{\alpha})\right) \geq N\left[\mathbf{0}, \Gamma \otimes E_s\left(e^{-\boldsymbol{\beta}} + \boldsymbol{e}/2\right)\right]\left(\mathsf{T}_{W,c}(\mathbf{0})\right), \tag{164}$$

where $\boldsymbol{e}$ is the $2s$-dimensional vector with all entries equal to 1. For the definition of $E_s\left(e^{-\boldsymbol{\beta}} + \boldsymbol{e}/2\right)$, see (56). When the POVM M^G is given as $M^G(B) = \int_B |\alpha_1, \ldots, \alpha_s\rangle \langle \alpha_1, \ldots, \alpha_s| d\boldsymbol{\alpha}$, the equality in (164) holds.

The proof can be found in "Appendix H". When the model has a group covariance, similar evaluation might be possible. For example, similar evaluation was done in the n-copy of full pure states family [49] and in the n-copy of squeezed states family [50, Sect. 4.1.3].

8.2. Tail property of D-invariant qudit models. For a k-parameter D-invariant model $\{\rho_t\}$, using Lemma 16 and Q-LAN, we have the following theorem.

Theorem 9. *Let $\{\rho_t\}_{t \in \Theta}$ be a k-parameter D-invariant model. Assume that $(\tilde{J}_{t_0'})^{-1}$ exists, ρ_{t_0} is a non-degenerate state, and a sequence of measurements $\mathfrak{m} := \{M_n\}$ satisfies local asymptotic covariance at $t_0 \in \Theta$. When $J_{t_0}^{-1/2}W J_{t_0}^{-1/2}$ commutes with $J_{t_0}^{-1/2}D_{t_0}J_{t_0}^{-1/2}$, we have*

$$\wp_{t_0,t|\mathfrak{m}}\left(\mathsf{T}_{W,c}(t)\right) \geq N\left[0, W^{1/2}J_{t_0}^{-1}W^{1/2} + \frac{1}{2}\left|W^{1/2}J_{t_0}^{-1}D_{t_0}J_{t_0}^{-1}W^{1/2}\right|\right](\mathsf{T}_c) \tag{165}$$

for $\mathsf{T}_c := \{x \in \mathbb{R}^k \mid \|x\| \geq c\}$. The equality holds if and only if $\wp_{t_0,t|\mathfrak{m}}$ is the normal distribution with average zero and covariance $V_{t_0|W}$ as defined in Eq. (114).

We note that bounds on the probability distributions are usually more difficult to obtain and more informative than the MSE bounds, as the MSE can be determined by the probability distribution. Theorem 9 provides an attainable bound of the tail probability, which can be used to determine the maximal probability that the estimate falls into a confidence region $\mathsf{T}_{W,c}$ as well as the optimal measurement.

Our proof of Theorem 9 needs some preparations. First, we introduce the concept of simultaneous diagonalization in the sense of symmetric transformation. Two $2k \times 2k$ real symmetric matrices A_1 and A_2 are called *simultaneously symplectic diagonalizable* when there exist a symplectic matrix S and two real vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ such that such that

$$S^T A_1 S = E_k(e^{-\boldsymbol{\beta}_1}), \quad S^T A_2 S = E_k(e^{-\boldsymbol{\beta}_2}) \tag{166}$$

with E_k defined in Eq. (56). Regarding the simultaneous diagonalization, we have the following property, whose proof can be found in "Appendix I":

Lemma 17. *The following conditions are equivalent for two $2k \times 2k$ real symmetric matrices A_1 and A_2.*

(i) A_1 *and* A_2^{-1} *are simultaneously symplectic diagonalizable.*
(ii) $\Omega_k A_2 A_1 = A_1 A_2 \Omega_k$, *where Ω_k is defined in Eq. (57).*

Using Lemma 17, we obtain the following lemma.

Lemma 18. *Let A_1 be $\left| J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} \right|^{1/2}$. Assume that $A_1^{-1} J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} A_1^{-1} = \Omega_k$. When $J_{t_0}^{-1/2} W J_{t_0}^{-1/2}$ commutes with $J_{t_0}^{-1/2} D_{t_0} J_{t_0}^{-1/2}$, $(A_1 W A_1)^{-1}$ and $A_1^{-1} J_{t_0}^{-1} A_1^{-1}$ are simultaneously symplectic diagonalizable.*

Proof. From $[J_{t_0}^{-1/2} W J_{t_0}^{-1/2}, J_{t_0}^{-1/2} D_{t_0} J_{t_0}^{-1/2}] = 0$, we get $[J_{t_0}^{-1} W, (J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1})] = 0$. Next, noticing that $(J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1}) = A_1 \Omega_k A_1$, we have

$$J_{t_0}^{-1} W A_1 \Omega_k A_1 = J_{t_0}^{-1} W (J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1}) = (J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1}) W J_{t_0}^{-1} = A_1 \Omega_k A_1 W J_{t_0}^{-1}.$$

The above equalities show that $J_{t_0}^{-1} W$ commutes with $A_1 \Omega_k A_1$, which further implies that

$$A_1^{-1} J_{t_0}^{-1} A_1^{-1} A_1 W A_1 \Omega_k = \Omega_k A_1 W A_1 A_1^{-1} J_{t_0}^{-1} A_1^{-1}.$$

Using Lemma 17, we get $(A_1 W A_1)^{-1}$ and $A_1^{-1} J_{t_0}^{-1} A_1^{-1}$ are simultaneously symplectic diagonalizable. $\square$

Proof of Theorem 9. Define $A_1 := \left| J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} \right|^{1/2}$ and $A_2 := A_1^{-1} J_{t_0}^{-1} D_{t_0} J_{t_0}^{-1} A_1^{-1}$. Applying a suitable orthogonal matrix, we assume that $A_2 = \Omega_k$ without loss of generality.

Step 1 For simplicity, we assume that there is no classical part. First, we choose an orthogonal matrix S' such that $S'^T A_2 S' = D$. Using Lemma 18 guarantees that the condition $[J_{t_0}^{-1/2} W J_{t_0}^{-1/2}, J_{t_0}^{-1/2} D_{t_0} J_{t_0}^{-1/2}] = 0$ allows us to simultaneously diagonalize W and D_{t_0}. That is, we can choose symplectic matrix S such that $S^T S'^T (A_1 W A_1)^{-1} S' S$ and $S^T S'^T A_1^{-1} J_{t_0}^{-1} A_1^{-1} S' S$ are diagonal matrixces $E_k(\boldsymbol{w})^{-1}$ and $E_k(\boldsymbol{\beta})$ for $k \in \mathbb{N}^*$. We introduce the local parameter $\boldsymbol{t}' := S^T S'^T A_1^{-1} \boldsymbol{t}$. Then, $\boldsymbol{t} \cdot W \boldsymbol{t} = \boldsymbol{t}' \cdot E_k(\boldsymbol{w}) \boldsymbol{t}'$.

For a sequence of measurement $\mathfrak{m} := \{M_n\}$ to satisfy local asymptotic covariance at $t_0 \in \Theta$, according to Theorem 5, we choose a covariant POVM $\tilde{M}^G$ to satisfy (103). Applying Lemma 16 to the POVM $\tilde{M}^G$, we obtain the desired statement.

Step 2 We consider the general case. Now, we choose the local parameter $t' := J_{t_0}^{-1/2} t$. In this coordinate, The inverse of the RLD quantum Fisher information is $I + J_{t_0}^{-1/2} D_{t_0} J_{t_0}^{-1/2}$. Since $J_{t_0}^{-1/2} D_{t_0} J_{t_0}^{-1/2}$ commutes with $J_{t_0}^{-1/2} W J_{t_0}^{-1/2}$, the weight matrix has no cross term between the classical and quantum parts. Using the above discussion and Lemma 16, we obtain the desired statement. □

9. Extension to Global Estimation and Generic Cost Functions

In the previous sections, we focused on local models and cost functions of the form $\mathrm{tr}\, W V[\wp_{t_0,t|\mathfrak{m}}]$. In this section, our treatment will be extended to global models $\{\rho_t\}_{t \in \Theta}$. (where the parameter to be estimated is not restricted to a local neighborhood) and to generic cost functions.

9.1. Optimal global estimation via local estimation. Our optimal global estimation is given by combining the two-step method and local optimal estimation. That is, the first step is the application of full tomography proposed in [26] on $n^{1-x/2}$ copies with the outcome $\hat{t}_0$ for a constant $x \in (0, 2/9)$, and the second step is the local optimal estimation at $\hat{t}_0$, given in Sect. 6.3, on

$$a_{n,x} := n - n^{1-x/2}$$

copies. Before its full description, we define the neighborhood $\Theta_{n,x}(t)$ of $t \in \Theta$ as

$$\Theta_{n,x}(t) := \left\{ y \mid \|y - t\| \leq n^{-\frac{1-x}{2}} \right\}. \tag{167}$$

Given a generic model $\mathcal{M} = \{\rho_t\}_{t \in \Theta}$ that does not contain any degenerate state and a weight matrix $W > 0$, we describe the full protocol as follows.

(A1) *Localization:* Perform full tomography proposed in [26] on $n^{1-x/2}$ copies, which is described by a POVM $\{M_{n^{1-x/2}}^{\mathrm{tomo}}\}$, for a constant $x \in (0, 2/9)$. The tomography outputs the first estimate $\hat{t}_0$ so that

$$\mathrm{Tr}\, \rho_t^{\otimes n^{1-x/2}} M_{n^{1-x/2}}^{\mathrm{tomo}} \left(\Theta_{n,x}(t_{\mathrm{g}}) \right) = 1 - O\left(e^{-n^{x/2}} \right) \tag{168}$$

for any true parameter t.

(A2) *Local estimation:* Based on the first estimate $\hat{t}_0$, apply the optimal local measurement $M_W^{\hat{t}_0, a_{n,x}}$ given in Theorem 7 with the weight matrix W. If the measurement outcome $\hat{t}_1$ of $M_W^{\hat{t}_0, a_{n,x}}$ is in $\Theta_{n,x}(\hat{t}_0)$, output the outcome $\hat{t}_1$ as the final estimate; otherwise output $\hat{t}_0$ as the final estimate.

Denoting the POVM of the whole process by $\mathfrak{m}_W = \{M_W^n\}$, we obtain the following theorem.

Theorem 10. *Assume that a qudit-model $\mathcal{M} = \{\rho_t\}_{t \in \Theta}$ does not contain any degenerate state, the parametrization is C^2 continuous, Θ is an open set, and $J_{t_0}^{-1}$ exists. (i) The sequence $\mathfrak{m}_W$ satisfies local asymptotic covariance at any point t_0 in the parameter space. (ii) The equation*

$$\operatorname{tr} W V[\wp_{t_0, t | \mathfrak{m}_W}] = \mathcal{C}(W, t_0) \tag{169}$$

holds for any point $t_0 \in \Theta$ and any $t \in \Theta_{n,x,c(t_0)}$ corresponding to a non-degenerate state, where $\mathcal{C}_\mathcal{S}(W, t_0)$ is the minimum weighted MSE as defined in Eq. (110). More precisely, we have

$$\limsup_{n \to \infty} \sup_{t_0 \in K} \sup_{t \in \Theta_{n,x,c(t_0)}} n^\kappa \left\| \wp_{t_0, t | M_W^n}^n - N[t, V_{t_0 | W}] \right\|_1 < \infty \tag{170}$$

for a compact set $K \subset \Theta$, where $V_{t_0 | W}$ is defined in Eq. (146) and $\Theta_{n,x,c(t_0)}$ is defined in Eq. (76). Further, when the parameter set Θ is bounded and $x < \kappa$, we have the following relation.

$$\lim_{n \to \infty} \sup_{t_0 \in K} \sup_{t \in \Theta_{n,x,c(t_0)}} \left\| V[\wp_{t_0, t | M_W^n}^n] - V_{t_0 | W} \right\|_1 = 0. \tag{171}$$

Here, we should remark the key point of the derivation. The existing papers [8,11] addressed the achievability of $\min_M \operatorname{tr} W J_{t | M}^{-1}$ with the two-step method, where $J_{t | M}$ is the Fisher information matrix of the distribution family $\{\wp_{t | M}\}_t$, which expresses the bounds among separable measurement [34, Exercise 6.42]. Hence it can be called the separable bound. In the one-parameter case, the separable bound equals the Holevo bound. To achieve the separable bound, we do not consider the sequence of measurement. Hence, we do not handle a complicated convergence. The global achievability of the separable bound can be easily shown by the two-step method [8,11]. However, in our setting, we need to handle the sequence of measurement to achieve the local optimality. Hence, we need to carefully consider the compact uniformity and the order estimate of the convergence in Theorem 7. In the following proof, we employ our evaluation with such detailed analysis as in Eq. (127).

Proof. Step 1 Define by $t_g := t_0 + \frac{t}{\sqrt{n}}$ the true value of the parameters. By definition, we have $\|t_g - \hat{t}_0\| \le n^{-\frac{1-x}{2}}$ with probability $1 - O(e^{-n^{x/2}})$ and $\|t_g - t_0\| \le c(t_0) n^{-\frac{1}{2} + x}$ by definition. Since the error probability vanishes exponentially, it would not affect the scaling of MSE. In this step, we will show

$$\left\| \wp_{t_0, t | M_W^{\hat{t}_0, a_{n,x}}}^n - N[t, V_{t_0 | W}] \right\|_1 = O(n^{-\kappa}). \tag{172}$$

Since $\|t_g - \hat{t}_0\| \le n^{-\frac{1-x}{2}}$ and $\|t_g - t_0\| \le c(t_0) n^{-\frac{1}{2} + x}$ imply $\|t_0 - \hat{t}_0\| \le 2c(t_0) n^{-\frac{1}{2} + x}$, we have

$$\|N[0, V_{t_0 | W}] - N[0, V_{\hat{t}_0 | W}]\|_1 = O(n^{-\frac{1}{2} + x}). \tag{173}$$

Eq. (127) of Theorem 7 implies

$$\left\| \wp_{t_0, t | M_W^{\hat{t}_0, a_{n,x}}}^{a_{n,x}} - N[0, V_{\hat{t}_0 | W}] \right\|_1 = O(n^{-\kappa}). \tag{174}$$

Since $\wp^n_{t_0,t|M_W^{\hat{t}_0,a_n,x}}(\mathsf{B}) = \wp^{a_{n,x}}_{t_0,t|M_W^{\hat{t}_0,a_{n,x}}}\left(\frac{\sqrt{a_{n,x}}\mathsf{B}}{\sqrt{n}}\right)$,

$$\left\|\wp^n_{t_0,t|M_W^{\hat{t}_0,a_n,x}} - N[t, \frac{n}{a_{n,x}}V_{t_0|W}]\right\|_1 = \left\|\wp^{a_{n,x}}_{t_0,t|M_W^{\hat{t}_0,a_{n,x}}} - N[t, V_{t_0|W}]\right\|_1$$

$$\leq \left\|\wp^{a_{n,x}}_{t_0,t|M_W^{\hat{t}_0,a_{n,x}}} - N[\mathbf{0}, V_{\hat{t}_0|W}]\right\|_1 + \|N[t, V_{t_0|W}] - N[t, V_{\hat{t}_0|W}]\|_1$$

$$= O(n^{-\kappa}) + O(n^{-\frac{1}{2}+x}) = O(n^{-\kappa}). \tag{175}$$

As we have

$$\|N[\mathbf{0}, V_{t_0|W}] - N[\mathbf{0}, \frac{n}{a_{n,x}}V_{t_0|W}]\|_1 = O(\frac{n}{a_{n,x}} - 1)$$

$$= O((1 - n^{x/2})^{-1} - 1) = O(n^{x/2}), \tag{176}$$

we obtain

$$\|\wp^n_{t_0,t|M_W^{\hat{t}_0,a_n,x}} - N[t, V_{t_0|W}]\|_1$$

$$\leq \|N[t, V_{t_0|W}] - N[t, \frac{n}{a_{n,x}}V_{t_0|W}]\|_1 + \|\wp^n_{t_0,t|M_W^{\hat{t}_0,a_n,x}} - N[t, \frac{n}{a_{n,x}}V_{t_0|W}]\|_1$$

$$= O(n^{x/2}) + O(n^{-\kappa}). \tag{177}$$

Step 2 We will show (170). First, we discuss two exceptional cases $\|t_g - \hat{t}_0\| > n^{-\frac{1-x}{2}}$ and $\|\hat{t}_1 - \hat{t}_0\| > n^{-\frac{1-x}{2}}$. Eq. (168) guarantees that

$$\mathrm{Tr}\, \rho_{t_g}^{\otimes n^{1-x/2}} M_{n^{1-x/2}}^{\mathrm{tomo}}\left(\left\{\hat{t}_0 \middle| \|t_g - \hat{t}_0\| > n^{-\frac{1-x}{2}}\right\}\right) = O\left(e^{-n^{x/2}}\right). \tag{178}$$

Eq. (175) and the property of normal distribution implies

$$\mathrm{Tr}\, \rho_{t_g}^{\otimes a_{n,x}} M_W^{\hat{t}_0,a_{n,x}}\left(\left\{\hat{t}_0 \middle| \|\hat{t}_1 - \hat{t}_0\| > n^{-\frac{1-x}{2}}\right\}\right)$$

$$= O(n^{-\kappa}) + N[t, \frac{n}{a_{n,x}}V_{t_0,|W}](\{t|\ \|t\| > n^{x/2}\})$$

$$= O(n^{-\kappa}) + O(e^{-O(n^{x/2})}) = O(n^{-\kappa}). \tag{179}$$

When $\|t_g - \hat{t}_0\| \leq n^{-\frac{1-x}{2}}$ and $\|\hat{t}_1 - \hat{t}_0\| \leq n^{-\frac{1-x}{2}}$, Eq. (172) holds under the condition $\|t_g - t_0\| \leq c(t_0)n^{-\frac{1}{2}+x}$, which implies that

$$\sup_{t\in\Theta_{n,x,c(t_0)}} \|\wp^n_{t_0,t|M_W^n} - N[t, V_{t_0|W}]\|_1 = O(n^{-\kappa}). \tag{180}$$

Since the above evaluation is compactly uniform with respect to t_0, we have (170).
Step 3 We will show

$$\lim_{n\to\infty} \sup_{t_0\in K} \sup_{t\in\Theta_{n,x,c(t_0)}} \left\|V[\wp^n_{t_0,t|M_W^n}] - \frac{n}{a_{n,x}}V_{t_0|W}\right\|_1 = 0 \tag{181}$$

because Eq. (181) implies (171) due to the convergence $\frac{n}{a_{n,x}} \to 1$. There are three cases.
(1) $\|t_g - \hat{t}_0\| > n^{-\frac{1-x}{2}}$, (2) $\|\hat{t}_1 - \hat{t}_0\| > n^{-\frac{1-x}{2}}$ and $\|t_g - \hat{t}_0\| \le n^{-\frac{1-x}{2}}$, and (3) the remaining case.

The compactness of Θ guarantees that the error $n(\hat{t} - t_g)(\hat{t} - t_g)^T$ is bounded by nC with a constant C. Due to (178), the contribution of the first case is bounded by $nC \cdot O(e^{-n^{x/2}})$, which goes to zero.

In the second case, since $\hat{t}_0 = \hat{t}$, the error $n(\hat{t} - t_g)(\hat{t} - t_g)^T$ is bounded by $n(n^{-\frac{1-x}{2}})^2 = n^x$. Due to (179), the contribution of the second case is bounded by $n^x \cdot O(n^{-\kappa}) = O(n^{x-\kappa})$, which goes to zero.

In the third case, since $\|t_g - \hat{t}_1\| \le \|t_g - \hat{t}_0\| + \|\hat{t}_0 - \hat{t}_1\| \le 2n^{-\frac{1-x}{2}}$, the error $n(\hat{t} - t_g)(\hat{t} - t_g)^T$ is bounded by $2n(n^{-\frac{1-x}{2}})^2 = 2n^x$. Due to (175), the contribution of the second case is bounded by $2n^x \cdot O(n^{-\kappa}) = O(n^{x-\kappa})$, which goes to zero. Therefore, we obtain (181). $\square$

9.2. Generic cost functions.

Finally, we show that results in this work hold also for any cost function $c(\hat{t}, t)$, which is bounded and has a symmetric expansion, in the sense of satisfying the following two conditions:

(B1) $c(\hat{t}, t)$ has a continuous third derivative, so that it can be expanded as $c(\hat{t}, t) = (\hat{t}^T - t^T)W_t(\hat{t} - t) + O(\|\hat{t} - t\|^3)$ as $\hat{t}$ is close to t, where the matrix $W_t \ge 0$ is a continuous function of t.
(B2) $c(\hat{t}, t) \le C$ for a constant $C > 0$ and for any $\hat{t}, t \in \mathbb{R}^k$.

To adopt this situation, we replace the step (A2) by the following step (A2)':

(A2)' Based on the first estimate $\hat{t}_0$, apply the optimal local measurement $M_{W_{\hat{t}_0}}^{\hat{t}_0, a_{n,x}}$ given in Theorem 7 with the weight matrix $W_{\hat{t}_0}$. If the measurement outcome $\hat{t}_1$ of $M_{W_{\hat{t}_0}}^{\hat{t}_0, a_{n,x}}$ is in $\Theta_{n,x}(\hat{t}_0)$, output the outcome $\hat{t}_1$ as the final estimate; otherwise output $\hat{t}_0$ as the final estimate $\hat{t}$.

Denoting the POVM of the whole process by $\mathfrak{m}_c = \{M_c^n\}$, we have the following result:

Theorem 11. *Assume the same assumption for the model $\mathcal{M}$ as Theorem 10. (i) When a sequence of measurements $\mathfrak{m} := \{M_n(d\hat{t})\}$ satisfies local asymptotic covariance at $t_0 \in \Theta$ and a cost function c satisfies condition (B1), the inequality*

$$\lim_{n \to \infty} n \cdot c_{t_0 + \frac{t}{\sqrt{n}}}(\mathfrak{m}) \ge \mathcal{C}(W_{t_0}, t_0) \tag{182}$$

holds, where $\mathcal{C}_{\mathcal{S}}(W_{t_0}, t_0)$ is defined in Eq. (110) and

$$c_{t_0 + \frac{t}{\sqrt{n}}}(\mathfrak{m}) := \int c\left(\hat{t}_n, t_0 + \frac{t}{\sqrt{n}}\right) \operatorname{Tr} \rho_{t_0 + \frac{t}{\sqrt{n}}}^{\otimes n} M_n\left(d\hat{t}_n\right).$$

(ii) In addition, if c also satisfies Condition (B2), $\mathfrak{m}_c = \{M_c^n\}$ is locally asymptotically covariant and attains the equality in (182) at any point $t_0 \in \Theta$ corresponding to a non-degenerate state.

Theorem 11 is reduced to a bound for the (actual) MSE when $c(\hat{t}, t) = (\hat{t}^T - t^T) W (\hat{t} - t)$ for $W \geq 0$. Therefore, bounds in this work, Eqs. (125) and (144) for instance, are also attainable bounds for the MSE of any locally asymptotically unbiased measurement.

Proof. Step 1 We prove (1). Consider any sequence of asymptotically covariant measurements $\mathfrak{m}_{t_0} := \{M_{n,t_0}\}$ at t_0. Denote by $t_g := t_0 + \frac{t}{\sqrt{n}}$ the true value of the parameters. For a cost function c satisfying (ii), we have

$$\lim_{n \to \infty} n \cdot c_{t_0 + \frac{t}{\sqrt{n}}}(\mathfrak{m})$$

$$= \lim_{n \to \infty} n \cdot \int c(\hat{t}_n, t_g) \, \mathrm{Tr} \, \rho_{t_g}^{\otimes n} M_{n,t_0}(d\hat{t}_n)$$

$$= \lim_{n \to \infty} \int \left(t'^T W_{t_g} t' + \frac{1}{\sqrt{n}} O(\|t'\|^3) \right) \wp_{t_0, t | M_{n,t_0}}^n(dt') \quad (t' := \sqrt{n}(\hat{t}_n - t_g))$$

$$= \int t'^T W_{t_0} t' \, \wp_{t_0, t | \mathfrak{m}_{t_0}}(dt') \geq \mathcal{C}(W_{t_0}, t_0).$$

Step 2 We prove (2). We replace W by W_t in the proof of Theorem 10. In this replacement, (173) is replaced by

$$\|N[\mathbf{0}, V_{t_0 | W_{t_0}}] - N[\mathbf{0}, V_{\hat{t}_0 | W_{\hat{t}_0}}]\|_1 = O(n^{-\frac{1}{2} + x}) \tag{183}$$

where $x \in (0, 2/9)$. Hence, the contributions of the first and second cases of Step 3 of the proof of Theorem 10 go to zero.

In the third case of Step 3 of the proof, we have $\|t_g - \hat{t}_1\| \leq 2n^{-\frac{1-x}{2}}$, Hence,

$$nc(\hat{t}_1, t_g) - n(\hat{t}_1^T - t_g^T) W_{t_g}(\hat{t}_1 - t_g) = nO(\|\hat{t}_1 - t_g\|^3) = O(n^{-\frac{1-3x}{2}}) \to 0. \tag{184}$$

Hence, in the contribution of the third case, we can replace the expectation of $nc(\hat{t}_1, t_g)$ by the weighted MSE with weight W_{t_g}. Hence, we obtain the part (2). □

10. Applications

In this section, we show how to evaluate the MSE bounds in several concrete examples.

10.1. Joint measurement of observables. Here we consider the fundamental problem of the joint measurement of two observables. For simplicity we choose to analyze qubit systems, although the approach can be readily generalized to arbitrary dimension. The task is to simultaneously estimate the expectation of two observables A and B in a qubit system. The observables can be expressed as $A = \mathbf{a} \cdot \boldsymbol{\sigma}$ and $B = \mathbf{b} \cdot \boldsymbol{\sigma}$ with $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ being the vector of Pauli matrices. We assume without loss of generality that $|\mathbf{a}| = |\mathbf{b}| = 1$ and $\mathbf{a} \cdot \mathbf{b} \in [0, 1)$. The state of an arbitrary qubit system can be expressed as

$$\rho := \frac{1}{2}(I + \mathbf{n} \cdot \boldsymbol{\sigma}),$$

where $\mathbf{n}$ is the Bloch vector.

With this notation, the task is reduced to estimate the parameters

$$x := \boldsymbol{a} \cdot \boldsymbol{n}, \qquad y := \boldsymbol{b} \cdot \boldsymbol{n}.$$

It is also convenient to introduce a third unit vector $\boldsymbol{c}$ orthogonal to $\boldsymbol{a}$ and $\boldsymbol{b}$ so that $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ form a (non-orthogonal) normalized basis of $\mathbb{R}^3$. In terms of this vector, we can define the parameter $z := \boldsymbol{c} \cdot \boldsymbol{n}$. In this way, we extend the problem to the full model containing all qubit states, where x, y are the parameters of interest and z is a nuisance parameter. Under this parameterization, we can evaluate the SLD operators for x, y, and z, as well as the SLD Fisher information matrix and the D matrix (see "Appendix J" for details), substituting which into the bound (144) yields:

$$\operatorname{tr} VW \geq \operatorname{tr} W \begin{pmatrix} \frac{1-|\boldsymbol{n}|^2+y'^2(1-s^2)+z^2}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & -\frac{x'y'(1-s^2)+(1-|\boldsymbol{n}|^2+z^2)s}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} \\[2mm] -\frac{x'y'(1-s^2)+(1-|\boldsymbol{n}|^2+z^2)s}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & \frac{1-|\boldsymbol{n}|^2+x'^2(1-s^2)+z^2}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} \end{pmatrix}$$

$$+ \frac{1}{2} \operatorname{tr} \left| \sqrt{W} \begin{pmatrix} 0 & -2z\sqrt{1-s^2} \\ 2z\sqrt{1-s^2} & 0 \end{pmatrix} \sqrt{W} \right|, \tag{185}$$

where $s := \boldsymbol{a} \cdot \boldsymbol{b}$, $x' = \frac{x-ys}{1-s^2}$, and $y' = \frac{y-xs}{1-s^2}$.

The tradeoff between the measurement precisions for the two observables is of fundamental interest. Substituting the expressions of D-matrix and the SLD Fisher information matrix (see "Appendix J") into Eq. (159), we obtain

$$\Delta_A + \Delta_B \geq 2|z|\sqrt{1-s^2},$$

which characterizes the precision tradeoff in joint measurements of qubit observables.

10.2. Direction estimation in the presence of noise. Consider the task of estimating a pure qubit state $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$, which can also be regarded as determining a direction in space, as qubits are often realized in spin-1/2 systems. In a practical setup, it is necessary to take into account the effect of noise, under which the qubit becomes mixed. For noises with strong symmetry, like depolarization, the usual MSE bound produces a good estimate of the error. For other kind of noises, it is essential to introduce nuisance parameters, and to use the techniques introduced in this paper.

As an illustration, we consider the amplitude damping noise as an example, which can be formulated as the channel

$$\mathcal{A}_\eta(\rho) := A_0 \rho A_0^\dagger + A_1 \rho A_1^\dagger$$

where $A_0 = |0\rangle\langle 0| + \sqrt{\eta}|1\rangle\langle 1|$ and $A_1 = \sqrt{1-\eta}|0\rangle\langle 1|$ are the Kraus operators. After the noisy evolution, the qubit state can be expressed as

$$\rho := \frac{1}{2}(I + \boldsymbol{n} \cdot \boldsymbol{\sigma})$$

with $\boldsymbol{n} := (\sqrt{\eta}\sin\theta\sin\varphi, \sqrt{\eta}\sin\theta\cos\varphi, 1 - \eta + \eta\cos\theta)$. Now we can regard η as a nuisance parameter, while θ and φ are the parameters of interest.
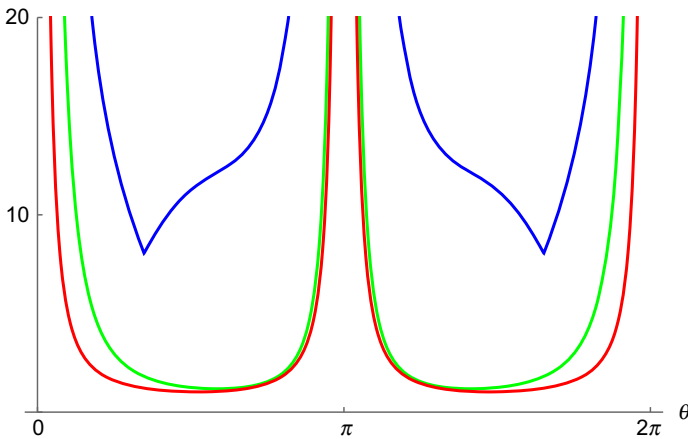
Fig. 2. Amplitude damping model. Plot of $\mathbf{MSE}_\theta + \mathbf{MSE}_\varphi$ as a function of θ (the value is the same for all φ) for $\eta = 0.9$ (red), $\eta = 0.5$ (green), and $\eta = 0.1$ (blue). Here $\mathbf{MSE}_x$ denotes the (x, x)-th element of the MSE matrix

Defining the derivative vector through the equation $\boldsymbol{p}_x \cdot \boldsymbol{\sigma} = \partial\rho/\partial x$, we can calculate the vectors

$$\boldsymbol{p}_\theta = (\sqrt{\eta}\cos\theta\sin\varphi, \sqrt{\eta}\cos\theta\cos\varphi, -\eta\sin\theta)$$
$$\boldsymbol{p}_\varphi = (\sqrt{\eta}\sin\theta\cos\varphi, -\sqrt{\eta}\sin\theta\sin\varphi, 0)$$
$$\boldsymbol{p}_\eta = (\sin\theta\sin\varphi/(2\sqrt{\eta}), \sin\theta\cos\varphi/(2\sqrt{\eta}), -1+\cos\theta)$$

In terms of the derivative vector, the SLD for the parameter $x \in \{\theta, \varphi, \eta\}$ takes the form

$$L_x = -\frac{2\boldsymbol{p}_x \cdot \boldsymbol{n}}{1 - |\boldsymbol{n}|^2}I + \left(2\boldsymbol{p}_x + \frac{2\boldsymbol{p}_x \cdot \boldsymbol{n}}{1 - |\boldsymbol{n}|^2}\boldsymbol{n}\right) \cdot \boldsymbol{\sigma} .$$

After some straightforward calculations, we get

$$J = \begin{pmatrix} 4\eta & 0 & 2\sin\theta \\ \\ 0 & 4\eta\sin^2\theta & 0 \\ \\ 2\sin\theta & 0 & \frac{(1-\eta)[\sin^2\theta+4\eta(1-\cos\theta)^2]+(2\eta-1)^2}{\eta(1-\eta)} \end{pmatrix}$$

and

$$D_{\theta,\varphi} = -D_{\varphi,\theta} = 8\eta\sin\theta(\eta - \eta\cos\theta + \cos\theta)$$
$$D_{\eta,\varphi} = -D_{\varphi,\eta} = 4\sin^2\theta(1+\eta-\eta\cos\theta)$$
$$D_{\theta,\eta} = -D_{\eta,\theta} = 0.$$

Then we have the MSE bound with nuisance parameter η. An illustration can be found in Fig. 2 with $W = I$ in Eq. (144). The minimum of the sum of the (x, x)-th matrix element of the MSE matrix for $x = \theta, \varphi$ is independent of φ, which is a result of the symmetry of the problem: the D-matrix does not depend on φ, and thus an estimation of φ can be obtained without affecting the precisions of other parameters. Notice that
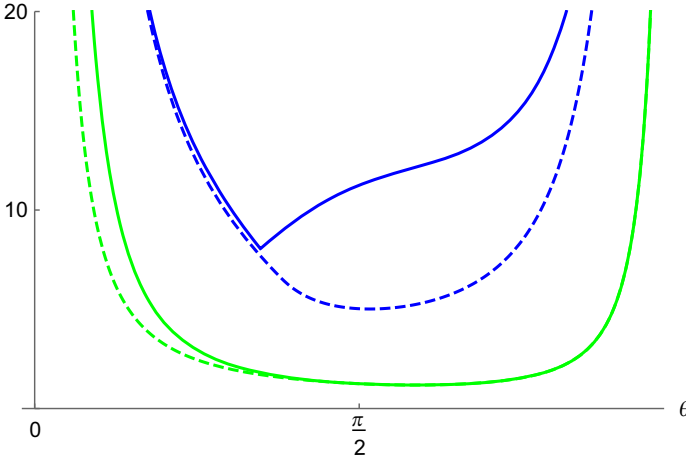
Fig. 3. The nuisance parameter bound versus the fixed parameter bound. $\mathbf{MSE}_\theta + \mathbf{MSE}_\varphi$ as a function of θ (the value is the same for all φ) for $\eta = 0.5$ (green) and $\eta = 0.1$ (blue) are plotted. The solid curves correspond to the case when η is a nuisance parameter, while the dashed curves correspond to the case when η is a fixed parameter. Here $\mathbf{MSE}_x$ denotes the (x, x)-th element of the MSE matrix

when the state is close to $|0\rangle$ or $|1\rangle$, it is insensitive to the change of θ, resulting in the cup-shape curves in Fig. 2.

Next, we evaluate the sum of MSEs of φ and θ when η is a (known) fixed parameter using Eq. (125) and compare it to the nuisance parameter case. The result of the numerical evaluation is plotted in Fig. 3. It is clear from the plot that the variance sum is strictly lower when η is treated as a fixed parameter, compared to the nuisance parameter case. This is a good example of how knowledge on a parameter (η) can assist the estimation of other parameters (φ and θ). It is also observed that, when the noise is larger (i.e. when η is smaller), the gain of precision by knowing η is also bigger.

10.3. Multiphase estimation with noise. Here we consider a noisy version of the multiphase estimation setting [20,51]. This problem was first studied by [20], where the authors derived a lower bound for the quantum Fisher information and conjectured that it was tight. Under local asymptotic covariance, we can now derive an attainable bound and show its equivalence to the SLD bound using the orthogonality of nuisance parameters, which proves the conjecture.

Our techniques also allow to resolve an open issue about the result of Ref. [20], where it was unclear whether or not the best precision depended on the knowledge of the noise. Using Corollary 4, we will also see that knowing *a priori* the strength of the noise does not help to decrease the estimation error.

The setting is illustrated in Fig. 4. Due to photon loss, the phase-shift operation is no longer unitary. Instead, it corresponds to a noisy channel with the following Kraus form:

$$\hat{\Pi}_{l,t} = \sqrt{\frac{(1-\eta)^l}{l!}} e^{i\hat{n}t} \eta^{\frac{\hat{n}}{2}} \hat{a}^l$$

$$\mathcal{N}_t(\rho) := \sum_l \hat{\Pi}_{l,t} \rho \hat{\Pi}_{l,t}^\dagger, \qquad \hat{\Pi}_{l,t} := \hat{\Pi}_{l_j,0} \otimes \left(\bigotimes_{j=1}^d \hat{\Pi}_{l_j,t_j} \right).$$
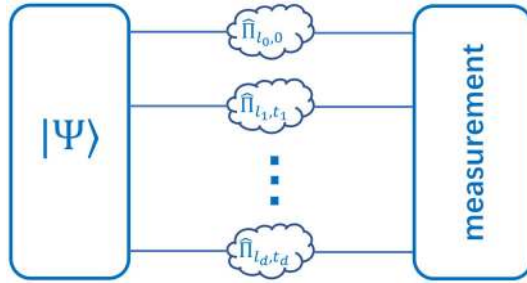
Fig. 4. Setup of noisy multiphase estimation. A input state $|\Psi\rangle$ passes through a noisy channel $\mathcal{N}_t$ consisting of $d + 1$ modes and carrying d parameters. The resultant state is then measured to obtain an estimate of t

Note that $\eta = 0$ corresponds to the noiseless scenario. We consider a pure input state with N photons and in the "generalized NOON form" as

$$|\Psi\rangle := a|N\rangle_0 + \frac{b}{\sqrt{d}} \sum_{j=1}^{d} |N\rangle_j \qquad |N\rangle_j := |0\rangle^{\otimes j} \otimes |N\rangle \otimes |0\rangle^{\otimes(d-j)}.$$

The output state from the noisy multiphase evolution would be

$$\mathcal{I} \otimes \mathcal{N}_t(|\Psi\rangle\langle\Psi|) = p_\eta |\psi_{\eta,t}\rangle\langle\psi_{\eta,t}| + (1 - p_\eta)\rho_\eta$$

where

$$|\psi_{\eta,t}\rangle = \cos\alpha_\eta|N\rangle_0 + \sin\alpha_\eta \sum_j \frac{e^{iNt}|N\rangle_j}{\sqrt{d}}, \qquad \cos\alpha_\eta = a/\sqrt{1 - b^2(1 - \eta^N)},$$

$p_\eta = 1 - b^2(1 - \eta^N)$, and ρ_η is independent of t. Notice that the output state is supported by the finite set of orthonormal states $\{|n\rangle_j : j = 0, \ldots, d, n = 0, \ldots, N\}$, and thus it is in the scope of this work.

In this case, $\{t_j\}$ are the parameters of interest, while α_η and p_η can be regarded as nuisance parameters. The SLD operators for these parameters can be calculated as

$$L_{t_j} = 2i \left[N|N\rangle_j\langle N|_j, |\psi_{\eta,t}\rangle\langle\psi_{\eta,t}| \right]$$

$$L_{\alpha_\eta} = 2 \left(|\psi_{\eta,t}\rangle\langle\psi^\perp_{\eta,t}| + |\psi^\perp_{\eta,t}\rangle\langle\psi_{\eta,t}| \right)$$

$$L_{p_\eta} = \frac{1}{p_\eta}|\psi_{\eta,t}\rangle\langle\psi_{\eta,t}| - \frac{1}{1 - p_\eta}\wp_{\mathcal{H}_\perp},$$

where $\wp_{\mathcal{H}_\perp}$ refers to the projection into the space orthogonal to $|\psi_{\eta,t}\rangle$. Notice that p_η and α_η are orthogonal to other parameters, in the sense that

$$\mathrm{Tr}\,\rho L_{t_j} L_{p_\eta} = \mathrm{Tr}\,\rho L_{\alpha_\eta} L_{p_\eta} = 0$$

and

$$\mathrm{Tr}\,\rho L_{t_j} L_{\alpha_\eta} = \frac{2ip_\eta \sin 2\alpha_\eta}{d}$$

Table 3. Comparison between our results and existing results

Topic	Our result	Existing results
Class of estimators	Local asymptotic covariance	Unbiased, [6]
		Locally unbiased [5], etc.
Local optimal estimator	Achieving Holevo bound	Achieving Holevo bound [19]
	Order estimate	
	Uniform convergence	
Global estimator	Achieving Holevo bound	Achieving separable bound [8,11]
	General cost function	
Nuisance parameter	Achieving Holevo bound	Special cases, e.g. qubit models [27,29]
	Order estimate	
	Uniform convergence	
Tail probability of limiting dist.	Gaussian states	Special pure states model [49,50]
	General D-inv. model	

is purely imaginary. We also have

$$\operatorname{Tr} \rho L_{t_j} L_{t_k} = 4 p_\eta N^2 |\langle \psi | N \rangle_j|^2 \left(\delta_{jk} - |\langle \psi | N \rangle_k|^2 \right)$$
$$= \frac{4 p_\eta N^2 \sin^2 \alpha_\eta}{d} \left(\delta_{jk} - \frac{\sin^2 \alpha_\eta}{d} \right).$$

Therefore, the SLD Fisher information matrix and the D matrix are of the forms

$$J = \begin{pmatrix} J_t & 0 & 0 \\ 0 & J_{\alpha_\eta} & 0 \\ 0 & 0 & J_{p_\eta} \end{pmatrix} \qquad D = \begin{pmatrix} 0 & D_{t,\alpha_\eta} & 0 \\ D_{t,\alpha_\eta}^T & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Substituting the above into the bound (144), we immediately get an attainable bound

$$\operatorname{tr} W V \left[\wp_{t_0, t | \mathsf{m}} \right] \geq \operatorname{tr} W J_t^{-1}, \quad (J_t)_{ij} = \frac{4 p_\eta N^2 \sin^2 \alpha_\eta}{d} \left(\delta_{ij} - \frac{\sin^2 \alpha_\eta}{d} \right) \qquad (186)$$

for any locally asymptotically covariant measurement m. Taking W to be the identity, one will see that for small η the sum of the variances scales as N^2/d^2, while for $\eta \to 1$ it scales as N^2/d, losing the boost in scaling compared to separate measurement of the phases. The bound (186) coincides with the SLD bound and the RLD bound. By Corollary 4, we conclude that the SLD (RLD) bound can be attained in the case of joint estimation of multiple phases. In addition, we stress that the ultimate precision does not depend on whether or not the noisy parameter η is known aprior: If η is unknown, one can obtain the same precision as when η is known by estimating η without disturbing the parameters of interest.

11. Conclusion

In this work, we completely solved the attainability problem of precision bounds for quantum state estimation under the local asymptotic covariance condition. We provided an explicit construction for the optimal measurement which attains the bounds globally. The key building block of the optimal measurement is the quantum local asymptotic

normality, derived in [16,17] for a particular type of parametrization and generalized here to arbitrary parameterizations. Besides the bound of MSE, we also derived a bound for the tail probability of estimation. Our work provides a general tool of constructing benchmarks and optimal measurements in multiparameter state estimation. In Table 3, we compare our result with existing results.

Here, we should remark the relation with the results by Yamagata et al. [19], which showed a similar statement for this kind of achievability in a local model scenario by a kind of local quantum asymptotic normality. In Theorem 7, we have shown the compact uniformity with the order estimation in our convergence, but they did not such properties. In the evaluation of global estimator, these properties for the convergence is essential. The difference between our evaluation and their evaluation comes from the key tools. The key tool of our derivation is Q-LAN (Proposition 2) by [16,17], which gives the state conversion, i.e., the TP-CP maps converting the states family with precise evaluation of the trace norm. However, their method is based on the algebraic central limit theorem [38,52], which gives only the behavior of the expectation of the function of operators R_i. This idea of applying this method to the achievability of the Holevo bound was first mentioned in [18]. Yamagata et al. [19] derived the detailed discussion in this direction.

Indeed, the algebraic version of Q-LAN by [38,52] can be directly applied to the vector X of Hermitian matrices to achieve the Holevo bound while use of the state conversion of Q-LAN requires some complicated procedure to handle the the vector X of Hermitian matrices, which is the disadvantage of our approach. However, since the algebraic version of Q-LAN does not give a state conversion directly, it is quite difficult to give the compact uniformity and the order estimate of the convergence. In this paper, to overcome the disadvantage of our approach, we have derived several advanced properties for Gaussian states in Sects. 3.2 and 3.3 by using symplectic structure. Using these properties, we could smoothly handle complicated procedure to fill the gap between the full quit model and arbitrary submodel.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A. Proof of Lemma 2

In this appendix, we show Lemma 2. For this aim, we discuss the existence of PDF. First, we show the following lemma.

Lemma 19. *Let P be a probability measure on $\mathbb{R}$. Define the location shift family $\{P_t\}$ as $P_t(B) := P(B + t)$. For an arbitrary disjoint decomposition $\mathcal{A} := \{A_i\}$ of $\mathbb{R}$, we assume that the probability distribution family $\{P_{\mathcal{A},t}\}$ has finite Fisher information $J_{\mathcal{A},t}$, where $P_{\mathcal{A},t}(i) := P_t(A_i)$. We also assume that $J_t := \sup_{\mathcal{A}} J_{\mathcal{A},t} < \infty$. Also, we define $x_+ := \inf\{x' | P(x', \infty) = 0\}$ and $x_- := \sup\{x' | P((-\infty, x']) = 0\}$.*

Then, for $x = x_-, x_+$, the derivative $p(x) := \frac{d}{dx} P((-\infty, x])$ is zero. For $x \in (x_-, x_+)$, the derivative $p(x) := \frac{d}{dx} P((-\infty, x])$ exists. For $x \in [x_-, x_+]$, $p(x)$ is Hölder continuous with order $1/2$, i.e.,

$$\limsup_{\epsilon \to 0} \left| \frac{p(x+\epsilon) - p(x)}{\epsilon^{1/2}} \right| < J_0^{3/2}. \tag{187}$$

Also, $p(x)$ is bounded, i.e., $\sup_x p(x) < \sqrt{J_0}$.

Proof. Assume that $x_+ < \infty$. We choose $A_1 := (-\infty, x_+]$, $A_2 := (x_+, \infty)$, and $\mathcal{A} = \{A_1, A_2\}$. The fidelity between $P_{\mathcal{A},0}$ and $P_{\mathcal{A},t}$ is

$$\sqrt{P((-\infty, x_+])}\sqrt{P((-\infty, x_+ - t])} = P((-\infty, x_+])\sqrt{1 - \frac{P((x_+ - t, x_+])}{P((-\infty, x_+])}}$$

$$\cong P((-\infty, x_+])\left(1 - \frac{1}{2}\frac{P((x_+ - t, x_+])}{P((-\infty, x_+])} - \frac{1}{8}\left(\frac{P((x_+ - t, x_+])}{P((-\infty, x_+])}\right)^2\right)$$

$$= 1 - \frac{1}{2}P((x_+ - t, x_+]) - \frac{1}{8}P((x_+ - t, x_+])^2 \tag{188}$$

Hence, we have

$$\lim_{t \to 0} \frac{8}{2t^2} P((x_+ - t, x_+]) = J_{\{(-\infty, x_+], (x_+, \infty)\}, 0}, \tag{189}$$

which implies the existence of $p(x_+)$ and $p(x_+) = 0$.

Similarly, we can show that there exists $p(x_-)$ and $p(x_-) = 0$ when $x_- > -\infty$.

Next, we choose $x \in (x_-, x_+)$. We choose $A_1 := (-\infty, x]$, $A_2 := (x, \infty)$, and $\mathcal{A} := \{A_1, A_2\}$. The fidelity between $P_{\mathcal{A},0}$ and $P_{\mathcal{A},t}$ is

$$\sqrt{P((-\infty, x])}\sqrt{P((-\infty, x + t])} + \sqrt{P((x, \infty))}\sqrt{P((x + t, \infty))}$$

$$= P((-\infty, x])\sqrt{1 + \frac{P((x, x + t])}{P((-\infty, x])}} + P((x, \infty))\sqrt{1 - \frac{P((x, x + t])}{P((x, \infty))}}$$

$$\cong P((-\infty, x])\left(1 + \frac{1}{2}\frac{P((x, x + t])}{P((-\infty, x])} - \frac{1}{8}\left(\frac{P((x, x + t])}{P((-\infty, x])}\right)^2\right)$$

$$+ P((x, \infty))\left(1 - \frac{1}{2}\frac{P((x, x + t])}{P((x, \infty))} - \frac{1}{8}\left(\frac{P((x, x + t])}{P((x, \infty))}\right)^2\right)$$

$$= P((-\infty, x]) + \frac{1}{2}P((x, x + t]) - \frac{1}{8}\frac{P((x, x + t])^2}{P((-\infty, x])}$$

$$+ P((x, \infty)) - \frac{1}{2}P((x, x + t]) - \frac{1}{8}\frac{P((x, x + t])^2}{P((x, \infty))}$$

$$= 1 - \frac{1}{8}\frac{P((x, x + t])^2}{P((-\infty, x])} - \frac{1}{8}\frac{P((x, x + t])^2}{P((x, \infty))}. \tag{190}$$

Hence, we have

$$\lim_{t \to 0} \frac{P((x, x+t])^2}{t^2 P((-\infty, x])} + \frac{P((x, x+t])^2}{t^2 P((x, \infty))} = J_{\{(-\infty, x], (x, \infty)\}, 0}, \tag{191}$$

which implies the existence of $p(x)$. Thus,

$$|p(x)| = \sqrt{\frac{J_{\{(-\infty, x], (x, \infty)\}, 0}}{\frac{1}{P((-\infty, x])} + \frac{1}{P((x, \infty))}}} \leq \sqrt{\frac{J_0}{\frac{1}{P((-\infty, x])} + \frac{1}{P((x, \infty))}}}$$

$$\leq \sqrt{\frac{J_0}{\min\left(\frac{1}{P((-\infty, x])}, \frac{1}{P((x, \infty))}\right)}} = \sqrt{J_0 \max(P((-\infty, x]), P((x, \infty)))} \leq \sqrt{J_0}. \tag{192}$$

Hence, $p(x)$ is bounded.

For $x \in [x_-, x_+]$, we consider the sets $\mathcal{A}_{x,d} := \{(x, x+d], (x, x+d]^c\}$. Then, we have

$$J_{\mathcal{A}_{x,d}, 0} = \frac{(\frac{d}{dt} P((x+t, x+d+t]))^2}{P((x, x+d])} + \frac{(\frac{d}{dt} P((x+t, x+d+t]^c))^2}{P((x, x+d]^c)}$$

$$= \frac{p((x, x+d])^2}{P((x, x+d])} + \frac{p((x, x+d]^c)^2}{P((x, x+d]^c)}. \tag{193}$$

Hence, we have

$$\frac{(p(x+d) - p(x))^2}{P((x, x+d])} < J_0. \tag{194}$$

Hence, when $d \to 0$

$$\frac{(p(x+d) - p(x))^2}{p(x)d} < J_0, \tag{195}$$

i.e.,

$$\frac{(p(x+d) - p(x))^2}{d} < p(x) J_0 \leq J_0^{3/2}, \tag{196}$$

which implies that $p(x)$ is Hölder continuous with order $1/2$.

Using the previous Lemma, we are in position to prove Lemma 2.

Proof of Lemma 2. Let $\mathcal{A} := \{A_i\}$ be an arbitrary disjoint finite decomposition of $\mathbb{R}$. Let $\mathcal{G}_{\mathcal{A}}$ be the coarse-graining map from a distribution on $\mathbb{R}$ to a distribution on the meshes A_i. Then, the Fisher information $J_{\mathcal{A}, n, t}$ of $\{\mathcal{G}_{\mathcal{A}}(\wp^n_{t_0, t | M_n})\}_t$ is not greater than the Fisher information J_t of $\{\rho^n_{t_0, t}\}_{t \in \Delta_n}$. Hence, the Fisher information $J_{\mathcal{A}, t}$ of $\{\mathcal{G}_{\mathcal{A}}(\wp_{t_0, t | \mathrm{m}})\}_t$ satisfies

$$J_{\mathcal{A}, t} = \lim_{n \to \infty} J_{\mathcal{A}, n, t} \leq J_{t_0}. \tag{197}$$

Therefore, we can apply Lemma 19 to $\wp_{t_0, t | \mathrm{m}}$. Lemma 19 guarantees the existence of the PDF of the limiting distribution $\wp_{t_0, t | \mathrm{m}}$.   □

B. Lemmas Used for Asymptotic Evaluations

In this appendix, we prepare two lemmas for asymptotic evaluations of information quantity of probability distributions.

Lemma 20. *Assume that two sequences of probability distributions $\{(P_n, Q_n)\}$ on $\mathbb{R}$ converges to a pair of probability distributions (P, Q) on $\mathbb{R}$, respectively. Then, the inequality*

$$F(P||Q) \geq \limsup_{n \to \infty} F(P_n||Q_n) \tag{198}$$

holds.

Proof. Let p and q be the Radon–Nikodým derivative of P and Q with respect to $P + Q$. Given $N > 0$, let $\mathcal{G}_N$ be the coarse-grained map from a distribution on $\mathbb{R}$ to a distribution on the Borel subsets $B \subset (-N, N]$ and the subset $(-N, N]^c$. Given $N > 0$ and $N' > 0$, let $\mathcal{G}_{N',N}$ be the coarse-grained map from a distribution on $\mathbb{R}$ to a distribution on meshes $(i/N', (i+1)/N']$ with $-NN' \leq i \leq NN' - 1$ and its complement $(-N, N]^c$.

We define $\bar{\mathcal{G}}_{N',N}(P)(x) := \frac{\mathcal{G}_{N',N}(P)(i)}{(P+Q)((i/N',(i+1)/N'])}$ for $x \in (i/N', (i+1)/N'] \subset (-N, N]$. Then, we have $\bar{\mathcal{G}}_{N',N}(P)(x) \leq \frac{\mathcal{G}_{N',N}(P)(i)}{P((i/N',(i+1)/N'])} = 1/2$ and $\bar{\mathcal{G}}_{N',N}(Q)(x) \leq 1/2$. Since $\sqrt{\bar{\mathcal{G}}_{N',N}(P)(x)}\sqrt{\bar{\mathcal{G}}_{N',N}(Q)(x)} \leq \sqrt{1/2}\sqrt{1/2} = 1/2$ for $x \in (-N, N]$, Lebesgue convergence theorem guarantees that

$$\lim_{N' \to \infty} \int_{-N}^{N} \sqrt{\bar{\mathcal{G}}_{N',N}(P)(x)}\sqrt{\bar{\mathcal{G}}_{N',N}(Q)(x)}(P + Q)(dx)$$
$$= \int_{-N}^{N} \sqrt{p(x)}\sqrt{q(x)}(P + Q)(dx). \tag{199}$$

Since

$$F\left(\mathcal{G}_{N',N}(P)||\mathcal{G}_{N',N}(Q)\right)$$
$$= \int_{-N}^{N} \sqrt{\bar{\mathcal{G}}_{N',N}(P)(x)}\sqrt{\bar{\mathcal{G}}_{N',N}(Q)(x)}(P + Q)(dx)$$
$$+ \sqrt{P((-N, N]^c)}\sqrt{Q((-N, N]^c)}, \tag{200}$$

Eq. (199) implies that

$$\lim_{N' \to \infty} F\left(\mathcal{G}_{N',N}(P)||\mathcal{G}_{N',N}(Q)\right) = F\left(\mathcal{G}_N(P)||\mathcal{G}_N(Q)\right). \tag{201}$$

Also, we have

$$\lim_{N \to \infty} F\left(\mathcal{G}_N(P)||\mathcal{G}_N(Q)\right) = F\left(P||Q\right). \tag{202}$$

Then, information processing inequality for the fidelity yields that

$$F\left(P_n||Q_n\right) \leq F\left(\mathcal{G}_{N',N}(P_n)||\mathcal{G}_{N',N}(Q_n)\right).$$

Since the number of meshes is finite, we have

$$\limsup_{n \to \infty} F\left(\mathcal{G}_{N',N}(P_n)||\mathcal{G}_{N',N}(Q_n)\right) = F\left(\mathcal{G}_{N',N}(P)||\mathcal{G}_{N',N}(Q)\right)$$

for every $N', N > 0$. Hence, we have

$$\limsup_{n \to \infty} F(P_n \| Q_n) \leq F(\mathcal{G}_{N',N}(P) \| \mathcal{G}_{N',N}(Q)). \tag{203}$$

Hence, using (201), (202) and (203), we have

$$\limsup_{n \to \infty} F(P_n \| Q_n) \leq F(P \| Q).$$

Lemma 21. *Let Θ be an open subset of $\mathbb{R}^{k'}$. Assume that a sequence of probability distributions $\{P_{t,n}\}_{t \in \Theta}$ on $\mathbb{R}^k$ converges to a family of probability distributions $\{P_t\}_{t \in \Theta}$ on $\mathbb{R}^k$. We denote their ϵ-difference Fisher information matrices by $J_{t,\epsilon}^n$ and $J_{t,\epsilon}$, respectively. For an vector t and $\epsilon > 0$, we also assume that there exists a Hermitian matrix J_ϵ such that $J_{t,\epsilon}^n \leq J_\epsilon$. Then, $P_{t+\epsilon e_j}$ is absolutely continuous with respect to P_t for $j = 1, \ldots, k'$, and the inequality*

$$\liminf_{n \to \infty} \langle a | J_{t,\epsilon}^n - J_{t,\epsilon} | a \rangle \geq 0 \tag{204}$$

holds for any complex vector $a \in \mathbb{C}^{k'}$.

Proof of Lemma 21. Since $J_{t,\epsilon}^n$ and $J_{t,\epsilon}$ are real matrices, it is sufficient to show (204) for a real vector a. In this proof, we fix the vector t.

Step (1): We show that $P_{t+\epsilon e_j}$ is absolutely continuous with respect to P_t for $j = 1, \ldots, k'$ by contradiction. Assume that there exists an integer j such that $P_{t+\epsilon e_j}$ is not absolutely continuous with respect to P_t. There exists a Borel set $B \subset \mathbb{R}^k$ such that $P_{t+\epsilon e_j}(B) > 0$ and $P_t(B) = 0$. Let $\mathcal{G}$ be the coarse-grained map from a distribution P on $\mathbb{R}^k$ to a binary distribution $(P(B), P(B^c))$ on two events $\{B, B^c\}$. Let $J_{t,B,\epsilon}$ and $J_{t,B,\epsilon}^n$ be the ϵ-difference Fisher information matrices of $\{\mathcal{G}(P_t)\}$ and $\{\mathcal{G}(P_{t,n})\}$, respectively. Information processing inequality implies that $J_{t,B,\epsilon}^n \leq J_{t,\epsilon}^n \leq J_\epsilon$. Also, $J_{t,B,\epsilon}^n \to J_{t,B,\epsilon}$ as $n \to \infty$. Hence, $J_{t,B,\epsilon} \leq J_\epsilon$. However, the j-th diagonal element of $J_{t,B,\epsilon}$ is infinity. It contradicts the assumption of contradiction.

Step (2): Let $p_{t+\epsilon e_j}$ be the Radon–Nikodým derivative of $P_{t+\epsilon e_j}$ with respect to P_t. We show that

$$\lim_{R \to \infty} P_t(\{x | p_{t+\epsilon e_j}(x) > R\}) = 0 \tag{205}$$

for $N, \epsilon > 0$, and any integer $j = 1, \ldots, k'$ by contradiction. We denote the LHS of (205) by C_j and assume there exists an integer j such that $C_j > 0$.

We set $R = \sqrt{J_{\epsilon; j, j} / C_j} + 2$. Setting B to be $\{x | p_{t+\epsilon e_j}(x) > R\}$, we repeat the same discussion as Step (1). Then, we obtain the contradiction as follows.

$$J_{\epsilon; j, j} \geq J_{t,B,\epsilon; j, j} \geq \int_{B_j} (p_{t+\epsilon e_j}(x) - 1)^2 P_t(dx) \geq \int_{B_j} (R - 1)^2 P_t(dx)$$

$$= (R - 1)^2 / C_j = \left(\sqrt{J_{\epsilon; j, j} / C_j} + 1 \right)^2 / C_j > J_{\epsilon; j, j}. \tag{206}$$

Step (3): We show (204) for a real vector a. We define the subsets

$$\mathcal{C}_R := \{x | \exists j, \, p_{t+\epsilon e_j}(x) > R\} \tag{207}$$

$$\mathcal{C}_{N,R} := ((-N, N]^k)^c \cup \mathcal{C}_R. \tag{208}$$

Given $R > 0$, let $\mathcal{G}_R$ be the coarse-grained map from a distribution on $\mathbb{R}^k$ to a distribution on the family of measurable sets $\{B \subset \mathbb{R}^k \setminus \mathcal{C}_R\} \cup \{\mathcal{C}_R\}$, where B is any Borel set in $\mathbb{R}^k \setminus \mathcal{C}_R$. Given $N > 0$ and $R > 0$, let $\mathcal{G}_{N,R}$ be the coarse-grained map from a distribution on $\mathbb{R}^k$ to a distribution on the family of measurable sets $\{B \subset \mathbb{R}^k \setminus \mathcal{C}_{N,R}\} \cup \{\mathcal{C}_{N,R}\}$, where B is any Borel subset in $\mathbb{R}^k \setminus \mathcal{C}_{N,R}$. Given $N > 0$, $R > 0$, and $N' > 0$, let $\mathcal{G}_{N',N,R}$ be the coarse-grained map from a distribution on $\mathbb{R}^k$ to a distribution on meshes $(\prod_{j=1}^k (i_j/N', (i_j+1)/N']) \setminus \mathcal{C}_{N,R}$ with $-NN' \le i_j \le NN' - 1$ and the complement $\mathcal{C}_{N,R}$.

We define

$$\bar{\mathcal{G}}_{N',N,R}(p)(x) := \frac{\mathcal{G}_{N',N,R}(p)(i_1, \ldots, i_k)}{P_t((\prod_{j=1}^k (i_j/N', (i_j+1)/N']) \setminus \mathcal{C}_{N,R})}$$

for $x \in (\prod_{j=1}^k (i_j/N', (i_j+1)/N']) \setminus \mathcal{C}_{N,R}$. Let $J_{t,\epsilon,N',N,R}$ be the ϵ-difference Fisher information matrix for the distribution family $\{\mathcal{G}_{N',N,R}(p_t)\}_t$. Let $J_{t,\epsilon,N,R}$ be the ϵ-difference Fisher information matrix for the distribution family $\{\mathcal{G}_{N,R}(p_t)\}_t$. Let $J_{t,\epsilon,N}$ be the ϵ-difference Fisher information matrix for the distribution family $\{\mathcal{G}_R(p_t)\}_t$.

Since $\bar{\mathcal{G}}_{N',N,R}(p_{\epsilon e_j})(x)\bar{\mathcal{G}}_{N',N,R}(p_{\epsilon e_{j'}})(x) \le R^2$ for $x \in \mathcal{C}_{N,R}^c$ and

$$\lim_{N' \to \infty} \bar{\mathcal{G}}_{N',N,R}(p_{t+\epsilon e_j})(x)\bar{\mathcal{G}}_{N',N,R}(p_{t+\epsilon e_{j'}})(x)$$
$$= \mathcal{G}_{N,R}(p_{t+\epsilon e_j})(x)\mathcal{G}_{N,R}(p_{t+\epsilon e_{j'}})(x), \tag{209}$$

Lebesgue convergence theorem guarantees that

$$\lim_{N' \to \infty} (J_{t,\epsilon,N',N,R})_{j,j'} = (J_{t,\epsilon,N,R})_{j,j'} \tag{210}$$

in the same way as (201). Since $p_t(((-N, N]^k)^c) \to 0$ as $N \to \infty$, we have

$$\lim_{N \to \infty} J_{t,\epsilon,N,R} = J_{t,\epsilon,R}. \tag{211}$$

Using (205), we have

$$\lim_{R \to \infty} J_{t,\epsilon,R} = J_{t,\epsilon}. \tag{212}$$

Let $J_{t,\epsilon,N',R,N}^n$ be the ϵ-difference Fisher information matrix for the distribution family $\{\mathcal{G}_{N',R,N}(p_{t,n})\}_t$. Then, information processing inequality (93) for the ϵ-difference Fisher information matrix yields that

$$\langle a | J_{t,\epsilon}^n - J_{t,\epsilon,N',R,N}^n | a \rangle \ge 0. \tag{213}$$

Since the number of meshes is finite, we have

$$\lim_{n \to \infty} J_{t,\epsilon,N',R,N}^n = J_{t,\epsilon,N',R,N}. \tag{214}$$

The combination of (210), (211), (212) (213), and (214) implies

$$\liminf_{n \to \infty} \langle a | J_{t,\epsilon}^n - J_{t,\epsilon,N',R,N} | a \rangle \ge 0. \tag{215}$$

Hence, using (210), (211), (212), and (215), we have $\liminf_{n \to \infty} \langle a | J_{t,\epsilon}^n - J_{t,\epsilon} | a \rangle \ge 0$.

C. Proof of Lemma 3

Before starting our proof of Lemma 3, we prepare the following lemmas.

Lemma 22. *Consider a canonical quantum Gaussian states family* $\{\Phi[\boldsymbol{\theta}, \boldsymbol{\beta}]\}$. *When a symplectic matrix S satisfies*

$$SE_q\left(e^{-\boldsymbol{\beta}}\right)S^T = E_q\left(e^{-\boldsymbol{\beta}}\right), \tag{216}$$

where E_q is the matrix defined in Eq. (56), *there exists a unitary operator U_S such that*

$$\Phi[\boldsymbol{\theta}, \boldsymbol{\beta}] = U_S\Phi[S\boldsymbol{\theta}, \boldsymbol{\beta}]U_S^\dagger.$$

Proof. Consider any coordinate $\boldsymbol{\theta}' = (\boldsymbol{\theta}'^C, \boldsymbol{\theta}'^Q)$, where $\boldsymbol{\theta}'^Q$ obtained by a reversible linear transformation S on the Q-LAN coordinate $\boldsymbol{\theta}^Q$, i.e. $\boldsymbol{\theta}'^Q = S\boldsymbol{\theta}^Q$.

Define $\hat{q}_j = \frac{1}{\sqrt{2}}(\hat{a}_j + \hat{a}_j^\dagger)$, $\hat{p}_j = \frac{1}{i\sqrt{2}}(\hat{a}_j - \hat{a}_j^\dagger)$, and $\boldsymbol{x} = (\hat{q}_1, \hat{p}_1, \ldots, \hat{q}_q, \hat{p}_q)^T$. We have

$$
\begin{aligned}
\Phi[\boldsymbol{\theta}_V^Q, \boldsymbol{\beta}] &= \Phi[S^{-1}\boldsymbol{\theta}'^Q, \boldsymbol{\beta}_V] \\
&= Z_\beta \exp[i\boldsymbol{x}^T S^{-1}\boldsymbol{\theta}'^Q]\exp\left[-\frac{\boldsymbol{x}^T E_q(\boldsymbol{\beta}_V)\boldsymbol{x}}{2}\right]\exp\Phi[-i\boldsymbol{x}^T S^{-1}\boldsymbol{\theta}'^Q] \\
&= Z_\beta \exp[i\boldsymbol{y}^T\boldsymbol{\theta}'^Q]\exp\left[-\frac{\boldsymbol{y}^T SE_q(\boldsymbol{\beta}_V)S^T\boldsymbol{y}}{2}\right]\exp\Phi[-i\boldsymbol{y}^T\boldsymbol{\theta}'^Q] \quad (217)
\end{aligned}
$$

where $\boldsymbol{y} := (S^{-1})^T\boldsymbol{x}$ and $Z_\beta > 0$ is a normalizing constant. Now, by the definition of $E_q(\boldsymbol{x})$ in Eq. (56) and $SE_q(e^{-\boldsymbol{\beta}_V})'S^T = E_q\left(e^{-(\boldsymbol{\beta}_V)'}\right)$, S must be of the block diagonal form $S = \bigoplus_i O_{\mathsf{s}_i}$. Here $\{\mathsf{s}_i\}$ is a partition of $\{1, \ldots, 2q\}$ and $j, k \in \mathsf{s}_i$ if and only if $\beta_j' = \beta_k'$, and O_{s_i} is an orthogonal matrix acting on any component $j \in \mathsf{s}_i$. Since $\boldsymbol{\beta}_V'$, $\boldsymbol{\beta}_V$ and $\ln\boldsymbol{\beta}_V$ are in one-to-one correspondence, we $SE_q(e^{-\boldsymbol{\beta}_V})S^T = E_q(e^{-\boldsymbol{\beta}_V})$. Substituting it into Eq. (217), we have

$$\Phi[\boldsymbol{\theta}_V^Q, \boldsymbol{\beta}_V] = Z_\beta \exp[i\boldsymbol{y}^T\boldsymbol{\theta}'^Q]\exp\left[-\frac{\boldsymbol{y}^T E_q(\boldsymbol{\beta}_V)\boldsymbol{y}}{2}\right]\exp\Phi[-i\boldsymbol{y}^T\boldsymbol{\theta}'^Q, \boldsymbol{\beta}_V].$$

That is, $(S^{-1})^T$ can be regarded as a transformation of $\boldsymbol{x}$. Finally, S is symplectic since $SDS^T = D$, and there exists a unitary U_S such that [50]

$$\Phi[\boldsymbol{\theta}_V^Q, \boldsymbol{\beta}_V] = U_S\Phi[\boldsymbol{\theta}'^Q, \boldsymbol{\beta}_V]U_S^\dagger. \tag{218}$$

Therefore we have $\Phi[\boldsymbol{\theta}_V^Q, \boldsymbol{\beta}_V] \cong \Phi[\boldsymbol{\theta}'^Q, \boldsymbol{\beta}_V]$ as desired. $\square$

Proof of Lemma 3. Using the imaginary part $\mathsf{Im}(\Gamma)$, we distinguish the classical and the quantum parts. Specifically, the kernel and support of $\mathsf{Im}(\Gamma)$ are

$$\mathrm{Ker}\,\mathsf{Im}(\Gamma) := \left\{\boldsymbol{x} \in \mathbb{R}^k : \mathsf{Im}(\Gamma)\boldsymbol{x} = 0\right\} \tag{219}$$

and

$$\text{Supp Im}(\Gamma) := (\text{Ker Im}(\Gamma))^{\perp}, \tag{220}$$

respectively. We introduce the classical parameters $\boldsymbol{\theta}^C$ and the quantum parameters $\boldsymbol{\theta}^Q$ in Ker Im(Γ) and Supp Im(Γ), respectively. That is, the classical parameter $\boldsymbol{\theta}^C$ and the quantum parameter $\boldsymbol{\theta}^Q$ are given by an invertible linear transformation T' such that $\boldsymbol{\theta}' := (\boldsymbol{\theta}^C, \boldsymbol{\theta}^Q) = T'\boldsymbol{t}$ satisfies

$$\text{Ker Im}(\Gamma) = \{T'^{-1}(\boldsymbol{\theta}^C, 0) | \boldsymbol{\theta}^C \in \mathbb{R}^{d^C}\} \tag{221}$$

$$\text{Supp Im}(\Gamma) = \{T'^{-1}(0, \boldsymbol{\theta}^Q) | \boldsymbol{\theta}^C \in \mathbb{R}^{2d^Q}\}. \tag{222}$$

Since the above separation is unique up to the linear conversion and any classical Gaussian states can be converted to each other via scale conversion, the remaining problem is to show the desired statement for the quantum part.

Next, we focus on the quantum part $((T'^{-1})^T \Gamma T'^{-1})^Q$ of the Hermitian matrix $(T'^{-1})^T \Gamma T'^{-1}$ It is now convenient to define the matrix

$$A := |\text{Im}(((T'^{-1})^T \Gamma T'^{-1})^Q)|^{1/2}. \tag{223}$$

The role of A is to normalize the D-matrix. Indeed, since $\text{Im}(((T'^{-1})^T \Gamma T'^{-1})^Q)$ is skew symmetric, $A^{-1}\text{Im}(((T'^{-1})^T \Gamma T'^{-1})^Q)A^{-1}$ is similar to Ω_{d^Q}, namely that there exists an orthogonal matrix S_0 so that $S_0 A^{-1}\text{Im}(((T'^{-1})^T \Gamma T'^{-1})^Q)A^{-1}S_0^T = \Omega_{d^Q}$. Moreover, since $S_0 A^{-1}\text{Re}(((T'^{-1})^T \Gamma T'^{-1})^Q)A^{-1}S_0^T$ is a real symmetric matrix, there exists a symplectic matrix S and a vector $\boldsymbol{\beta}$ such that [53]

$$S^T S_0 A^{-1}\text{Im}(((T'^{-1})^T \Gamma T'^{-1})^Q)A^{-1}S_0^T S = E_{d^Q}(e^{-\boldsymbol{\beta}}). \tag{224}$$

Meanwhile, we have $S S_0 A^{-1}\text{Im}(((T'^{-1})^T \Gamma T'^{-1})^Q)A^{-1}S_0 S^T = \Omega_{d^Q}$ since S is symplectic. Overall, when T is given as $(I \oplus (S S_0 A^{-1}))T'$, the desired requirement is satisfied.

The uniqueness of $\boldsymbol{\beta}$ is guaranteed by the uniqueness of symplectic eigenvalues. Hence, when two linear conversions T and $\tilde{T}$ satisfies the condition of the statement, $T\Gamma T^T = \tilde{T}\Gamma \tilde{T}^T$. Thus, Lemma 22 guarantees that the canonical Gaussian states $G(T^{-1}\boldsymbol{\alpha}, T\Gamma T^T)$ and $G(\tilde{T}^{-1}\boldsymbol{\alpha}, \tilde{T}\Gamma \tilde{T}^T)$ are unitarily equivalent. $\quad\square$

D. Proof of Lemma 5

$(3) \Rightarrow (1)$: When a Gaussian states family is given in the RHS of (64), it is clearly D-invariant.

$(1) \Leftrightarrow (2)$: Assume that the Gaussian states $G[\boldsymbol{\alpha}, \Gamma]$ is generated by the operators $\boldsymbol{R} = (R_1, \ldots, R_d)$. Due to Lemma 4, the SLDs of $\{G[T(\boldsymbol{t}), \Gamma]\}$ are $L_j := \sum_{k,k'} T_{k,j} (A^{-1})_{k,k'} R_{k'}$. Lemma 4 guarantees that

$$\mathcal{D}(L_j) = \sum_{k,k'} T_{k,j}(A^{-1})_{k,k'}\mathcal{D}(R_{k'}) = \sum_{k,k'} T_{k,j}(A^{-1})_{k,k'} \sum_{j'} 2B_{k',j'} R_{j'}$$

$$= -2\sum_k R_k (BA^{-1}T)_{k,j}.$$

Hence, the D-invariance is equivalent to the condition (2).

(2) $\Rightarrow$ (3): First, we separate the system into the classical and the quantum parts. In the Gaussian states family $\{G[\boldsymbol{\alpha}, \Gamma]\}$ this separation can be done by considering the Kernel of $Im(\Gamma)$ as in the proof of Lemma 3. In the Gaussian states family $G[T(t), \Gamma]$ this separation can be done by considering the Kernel of the D-matrix D_0 in the same way. Since the relation (64) for the classical part is easily done, we show the relation (64) when only the quantum part exists.

Under the above assumption, we define the $k \times (d - k)$ matrix T' such that $F := (T \oplus T')$ is invertible and $T'^T A^{-1} T = 0$. Then, Lemma 3 guarantees that $G[F(t, t'), \Gamma]$ is unitarily equivalent to $G[(t, t'), F^{-1}\Gamma(F^T)^{-1}]$. Since $T'^T A^{-1} T = 0$, we have

$$
\begin{aligned}
F^{-1}\Gamma(F^T)^{-1} &= F^{-1}A(F^T)^{-1}F^T A^{-1}\Gamma A^{-1}F F^{-1}A(F^T)^{-1} \\
&= (F^T T'^T F)^{-1}F^T A^{-1}\Gamma A^{-1}F(F^T T'^T F)^{-1} = \Gamma_T \oplus \Gamma_0, \quad (225)
\end{aligned}
$$

where

$$
\Gamma_0 := ((T'^T A^{-1} T')^{-1} + i(T'^T A^{-1} T')^{-1}(T'^T B T')(T'^T A^{-1} T')^{-1}).
$$

Hence, $G[(t, t'), F^{-1}\Gamma(F^T)^{-1}] = G[t, \Gamma_T] \otimes G[t', \Gamma_0]$. Putting $t' = 0$, we obtain the condition (3).

E. Proof of Lemma 6

Since Lemma 3 shows that general Gaussian states can be reduced to the canonical Gaussian states, we discuss only the canonical Gaussian states.

Step 1 We show the statement when we have only the quantum part and $X = R$. For a given state ρ, we define the POVM M_ρ by

$$
M_\rho(\mathsf{B}) := \int_{\mathsf{B}} T_{\boldsymbol{\alpha}} \rho T_{\boldsymbol{\alpha}}^\dagger d\boldsymbol{\alpha}. \quad (226)
$$

When ρ is a squeezed state with $\mathrm{Tr}\,\rho Q_j = \mathrm{Tr}\,\rho_j P = 0$, the output distribution $\wp_{\boldsymbol{\alpha}|M[\rho]}$ of $M[\rho]$ is the $2d^Q$-dimensional normal distribution of average $\boldsymbol{\alpha}$ and the following covariance matrix [5];

$$
E_{d^Q}(\boldsymbol{\beta}) + V_\rho, \quad \text{with } V_\rho := \begin{pmatrix} (\mathrm{Tr}\,Q_i Q_j\rho)_{i,j} & (\mathrm{Tr}\,Q_i P_j\rho)_{i,j} \\ (\mathrm{Tr}\,P_i Q_j\rho)_{i,j} & (\mathrm{Tr}\,P_i P_j\rho)_{i,j} \end{pmatrix}. \quad (227)
$$

In the single-mode case, without loss of generality, we can assume that W is a diagonal matrix $\begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}$ because this diagonalization can be done by applying the orthogonal transformation between Q and P. Then,

$$
\frac{1}{2}\sqrt{W}^{-1}|\sqrt{W}\Omega_{d^Q}\sqrt{W}|\sqrt{W}^{-1} = \begin{pmatrix} \frac{\sqrt{w_1}}{2\sqrt{w_1}} & 0 \\ 0 & \frac{\sqrt{w_1}}{2\sqrt{w_2}} \end{pmatrix}. \quad (228)
$$

We define the squeezed state $\rho[w]$ by $V_\rho = \begin{pmatrix} \frac{\sqrt{w}}{2} & 0 \\ 0 & \frac{1}{2\sqrt{w}} \end{pmatrix}$. Then the squeezed state $\rho[\frac{w_2}{w_1}]$ satisfies the condition $V_{\rho[\frac{w_2}{w_1}]} = \frac{1}{2}\sqrt{W}^{-1}|\sqrt{W}\Omega_{d^Q}\sqrt{W}|\sqrt{W}^{-1}$. Hence, the POVM $M[\rho[\frac{w_2}{w_1}]]$ satisfies the requirement.

In the multiple-mode case, we choose a symplectic matrix S such that SWS^T is a diagonal matrix with diagonal element $w_1, w_2, \ldots, w_{2dQ}$. The matrix

$$\frac{1}{2}S\sqrt{W}^{-1}|\sqrt{W}\Omega_{dQ}\sqrt{W}|\sqrt{W}^{-1}S^T$$

is the diagonal matrix with diagonal element $\frac{\sqrt{w_1}}{2\sqrt{w_2}}, \frac{\sqrt{w_1}}{2\sqrt{w_2}}, \ldots, \frac{\sqrt{w_{2dQ}}}{2\sqrt{w_{2dQ-1}}}, \frac{\sqrt{w_{2dQ-1}}}{2\sqrt{w_{2dQ}}}$. Employing symplectic representation $U(S)$ given in [50, Sect. 7.8], we apply the unitary operator $U(S)$. Then, the state $\rho[\frac{w_2}{w_1}] \otimes \cdots \otimes \rho[\frac{w_{2dQ}}{w_{2dQ-1}}]$ satisfies the condition $V_{U(S)\rho[\frac{w_2}{w_1}]\otimes\cdots\otimes\rho[\frac{w_{2dQ}}{w_{2dQ-1}}]U(S)^\dagger} = \frac{1}{2}\sqrt{W}^{-1}|\sqrt{W}\Omega_{dQ}\sqrt{W}|\sqrt{W}^{-1}$. Therefore, Hence, the POVM $M[U(S)\rho[\frac{w_2}{w_1}] \otimes \cdots \otimes \rho[\frac{w_{2dQ}}{w_{2dQ-1}}]U(S)^\dagger]$ satisfies the requirement.

Step 2 As the next step, we show the statement when we have only quantum part and the vector X is given in the following way with an integer $\tilde{k}$;

$$X_1 = Q_1, \ X_2 = P_1, \ \ldots, \ X_{2\tilde{k}-1} = Q_{\tilde{k}}, \ X_{2\tilde{k}} = P_{\tilde{k}} \tag{229}$$

$$X_{2\tilde{k}+1} = Q_{\tilde{k}+1}, \ X_{2\tilde{k}+2} = Q_{\tilde{k}+2}, \ \ldots, \ X_k = Q_{k-\tilde{k}}. \tag{230}$$

For the latter $k - 2\tilde{k}$ parameters, we just apply the measurements for the observables $Q_{\tilde{k}+1}, \ldots, Q_{k-\tilde{k}}$. We denote the part of the initial $2\tilde{k}$ parameters for W by W_Q. We substituting W_Q into W in the construction of the POVM $M[U(S)\rho[\frac{w_2}{w_1}] \otimes \cdots \otimes \rho[\frac{w_{2dQ}}{w_{2dQ-1}}]U(S)^\dagger]$. Then, as the output distribution, it realizes the normal distribution with average t and covariance matrix $\mathsf{Re}((Z_t(X)) + \sqrt{W}^{-1}|\sqrt{W}\mathsf{Im}(Z_t(X))\sqrt{W}|\sqrt{W}^{-1}$.

Step 3 Next, we consider the case when we have only quantum part and X does not have the form (229), (230). Let $2\tilde{k}$ be the rank of $\mathsf{Im}(Z_t(X))$. We choose an invertible matrix T' such that non-zero entries of the matrix $T'\mathsf{Im}(Z_t(X))T'^T$ are limited in the first $2\tilde{k}$ components and are given as $\Omega_{\tilde{k}}$. When we apply the matrix T' to the outcome of M, we obtain another POVM, which is denoted by M'. M' can be considered as a measurement for $X' := T'(X)$ instead of X. Hence, we have the relation $V_\alpha(M') = T'V_\alpha(M)T'^T$. Thus, it is sufficient to discuss the case with X' and the weight matrix $T'^T WT'$.

Then, we define $Y_j := (T'X)_j$ for $j = 1, \ldots, 2\tilde{k}$ and $Y_{2\tilde{k}+2j'-1} := (T'X)_{2\tilde{k}+j'}$ for $j' = 1, \ldots, k - 2\tilde{k}$. Next, we choose $k' - k$ operators $Y_{2\tilde{k}+2j}$ for $j = 1, \ldots, k - 2\tilde{k}$ and $Y_{2k-2\tilde{k}+j'}$ for $j' = 1, \ldots, k'$ as linear combinations of Q_j and P_j with $j = 1, \ldots, k'/2$ such that $\mathsf{Im}(Z_t(Y)) = \Omega_{k'/2}$. Hence, there exists a symplectic matrix S such that $Y = SQ$, where the vector Q is defined as $(Q)_{2j-1} = Q_j$ and $(Q)_{2j} = P_j$. Employing symplectic representation $U(S)$ given in [50, Sect. 7.8], we apply the unitary operator $U(S)$ to the Hilbert space so that $U(S)(T'X)_j$ satisfies the conditions (229) and (230). Hence, our problem is reduced to Step 2.

Step 4 We consider the case when our system is composed of the classical and the quantum parts. X_i is given as a linear combination of the operators Q_j and P_j, and the classical random variables $Z_{j'}$. Then, we divide $X = (X_i)$ to the sum of the quantum part $X^Q = (X_i^Q)$ and the classical part $X^C = (X_i^C)$. Then, we have $Z_\alpha(X) = Z_t(X^Q) + Z_\alpha(X^C)$, which implies that $\operatorname{tr} W\mathsf{Re}(Z_\alpha(X)) + \operatorname{tr}|\sqrt{W}\mathsf{Im}(Z_\alpha(X))\sqrt{W}| = \operatorname{tr} W\mathsf{Re}(Z_\alpha(X^Q)) + \operatorname{tr}|\sqrt{W}\mathsf{Im}(Z_\alpha(X^Q))\sqrt{W}|\operatorname{tr}|\sqrt{W}\mathsf{Im}(Z_\alpha(X^C))\sqrt{W}|$. Hence, when the outcome is given by the sum of X^C and the outcome of Step 3 with X^Q, the desired properties are satisfied.

F. Proof of Eq. (70)

By additivity of quantum Fisher information, the quantum Fisher information of $\{\rho_{\boldsymbol{\theta},n}\}$ is equal to the quantum Fisher information of $\{\rho_{\boldsymbol{\theta}_0+\boldsymbol{\theta}}\}$. For the latter, solving the equations

$$\frac{\partial \rho_{\boldsymbol{\theta}_0+\boldsymbol{\theta}}}{\partial \theta_{j,k}^{\mathrm{R/I}}} = \frac{1}{2}\left(L_{\theta_{j,k}^{\mathrm{R/I}}}\rho_{\boldsymbol{\theta}_0+\boldsymbol{\theta}} + \rho_{\boldsymbol{\theta}_0+\boldsymbol{\theta}} L_{\theta_{j,k}^{\mathrm{R/I}}}\right), \tag{231}$$

we get

$$L_{\theta_{j,k}^{\mathrm{R}}} = \frac{2\sqrt{\theta_{0,j}-\theta_{0,k}}}{(\theta_{0,j}+\theta_{0,k})}\cdot T_{j,k}^{\mathrm{I}} \qquad L_{\theta_{j,k}^{\mathrm{I}}} = -\frac{2\sqrt{\theta_{0,j}-\theta_{0,k}}}{(\theta_{0,j}+\theta_{0,k})}\cdot T_{j,k}^{\mathrm{R}}. \tag{232}$$

Using Eq. (232), we can evaluate the SLD quantum Fisher information

$$(J_{\boldsymbol{\theta}_0})_{j,k}^{\mathrm{R}} = (J_{\boldsymbol{\theta}_0})_{j,k}^{\mathrm{I}} = \frac{4(1-\beta_{j,k})}{1+\beta_{j,k}} \qquad (J_{\boldsymbol{\theta}_0})_{j,k}^{(\mathrm{R,I})} = (J_{\boldsymbol{\theta}_0})_{j,k}^{(\mathrm{I,R})} = 0 \tag{233}$$

and the D-matrix

$$(D_{\boldsymbol{\theta}_0})_{j,k}^{\mathrm{R}} = (D_{\boldsymbol{\theta}_0})_{j,k}^{\mathrm{I}} = 0 \qquad (D_{\boldsymbol{\theta}_0})_{j,k}^{(\mathrm{R,I})} = -(D_{\boldsymbol{\theta}_0})_{j,k}^{(\mathrm{I,R})} = -\frac{8(1-\beta_{j,k})^2}{(1+\beta_{j,k})^2}, \tag{234}$$

having used the definition $\beta_{j,k} := \theta_{0,k}/\theta_{0,j}$. It can be immediately verified that $(\tilde{J}_{\boldsymbol{\theta}_0})^{-1} = (J_{\boldsymbol{\theta}_0})^{-1} + \frac{i}{2}(J_{\boldsymbol{\theta}_0})^{-1}D_{\boldsymbol{\theta}_0}(J_{\boldsymbol{\theta}_0})^{-1}$ matches Eq. (70).

The case with the displaced thermal state $\rho_{\theta_{j,k}^{\mathrm{R}}+i\theta_{j,k}^{\mathrm{I}},\beta_{j,k}}$ follows from Lemma 4.

G. Proof of Lemma 12

Denote by $Q(\boldsymbol{y}) := \frac{1}{\pi^{k/2}}\langle \boldsymbol{y}|F|\boldsymbol{y}\rangle$ the Q-function of F [54]. Expanding displaced thermal states into a convex combination of coherent states, Eq. (104) can be rewritten as

$$f(\boldsymbol{\alpha}) = \int d\boldsymbol{y}\left(\prod_j \sqrt{\frac{1-\gamma_j}{\gamma_j}}\, e^{-(1-\gamma_j)(y_j-\alpha_j)^2/\gamma_j}\right) Q(\boldsymbol{y}). \tag{235}$$

Taking the Fourier transform $\mathcal{F}_{\boldsymbol{y}\to\boldsymbol{\xi}}(g) := \int d\boldsymbol{y}\, e^{i\boldsymbol{y}\cdot\boldsymbol{\xi}}g$ on both sides, we get

$$\mathcal{F}_{\boldsymbol{\alpha}\to\boldsymbol{\xi}}(f(\boldsymbol{\alpha}))$$

$$= \int\int d\boldsymbol{\alpha}d\boldsymbol{y}\left(\prod_j \sqrt{\frac{1-\gamma_j}{\gamma_j}}\, e^{i\xi_j\alpha_j-(1-\gamma_j)(y_j-\alpha_j)^2/\gamma_j}\right) Q(\boldsymbol{y})$$

$$= \left(\prod_j \int d\alpha_j\sqrt{\frac{1-\gamma_j}{\gamma_j}}\, e^{-\frac{1-\gamma_j}{\gamma_j}\left(\alpha_j-y_j-\frac{i\xi_j\gamma_j}{2(1-\gamma_j)}\right)^2}\right)\int d\boldsymbol{y}\, e^{i\xi_j y_j-\sum_j \frac{\gamma_j\xi_j^2}{4(1-\gamma_j)}} Q(\boldsymbol{y})$$

$$= \sqrt{\pi^k}e^{-\frac{1}{4}\sum_j \frac{\gamma_j\xi_j^2}{1-\gamma_j}}\mathcal{F}_{\boldsymbol{y}\to\boldsymbol{\xi}}(Q(\boldsymbol{y})). \tag{236}$$

In addition, we know that the P-function $P(\mathbf{y})$ [55] of F can be evaluated via the Q-function as (see, for instance, [56])

$$P(\mathbf{y}) = \mathcal{F}_{\boldsymbol{\xi} \to \mathbf{y}}^{-1} \left(e^{\frac{1}{4} \sum_j \xi_j^2} \mathcal{F}_{\boldsymbol{\alpha} \to \boldsymbol{\xi}} \left(Q(\boldsymbol{\alpha}) \right) \right). \tag{237}$$

The combination of (236) and (237) yields

$$P(\mathbf{y}) = \mathcal{F}_{\boldsymbol{\xi} \to \mathbf{y}}^{-1} \left(\sqrt{\pi^k} e^{\frac{1}{4} \sum_j \frac{\xi_j^2}{1-\gamma_j}} \mathcal{F}_{\boldsymbol{\alpha} \to \boldsymbol{\xi}} \left(f(\boldsymbol{\alpha}) \right) \right).$$

By definition of the P-function $P(\mathbf{y})$, F satisfies

$$F = \int d\mathbf{y} \, P(\mathbf{y}) |\mathbf{y}\rangle\langle\mathbf{y}| = \int d\mathbf{y} \, \mathcal{F}_{\boldsymbol{\xi} \to \mathbf{y}}^{-1} \left(\sqrt{\pi^k} e^{\frac{1}{4} \sum_j \frac{\xi_j^2}{1-\gamma_j}} \mathcal{F}_{\boldsymbol{\alpha} \to \boldsymbol{\xi}} \left(f(\boldsymbol{\alpha}) \right) \right) |\mathbf{y}\rangle\langle\mathbf{y}|. \tag{238}$$

Conversely, we assume that F is given by (105). Then, we choose the function $Q(\boldsymbol{\alpha})$ to satisfy

$$\mathcal{F}_{\boldsymbol{\alpha} \to \boldsymbol{\xi}} \left(f(\boldsymbol{\alpha}) \right) = \sqrt{\pi^k} e^{-\frac{1}{4} \sum_j \frac{\gamma_j \xi_j^2}{1-\gamma_j}} \mathcal{F}_{\mathbf{y} \to \boldsymbol{\xi}} \left(Q(\mathbf{y}) \right). \tag{239}$$

Since $F = \int d\mathbf{y} \, \mathcal{F}_{\boldsymbol{\xi} \to \mathbf{y}}^{-1} \left(e^{\frac{1}{4} \sum_j \xi_j^2} \mathcal{F}_{\boldsymbol{\alpha} \to \boldsymbol{\xi}} \left(Q(\boldsymbol{\alpha}) \right) \right) |\mathbf{y}\rangle\langle\mathbf{y}|$, we have $Q(\mathbf{y}) = \frac{1}{\pi^{k/2}} \langle\mathbf{y}|F|\mathbf{y}\rangle$.

Expanding displaced thermal states into a convex combination of coherent states, we have

$$\mathrm{Tr} \, F G[\boldsymbol{\alpha}, \boldsymbol{\gamma}] = \int d\mathbf{y} \left(\prod_j \sqrt{\frac{1-\gamma_j}{\gamma_j}} \, e^{-(1-\gamma_j)(y_j - \alpha_j)^2/\gamma_j} \right) Q(\mathbf{y}). \tag{240}$$

Applying the inverse of $\mathcal{F}_{\boldsymbol{\alpha} \to \boldsymbol{\xi}}$ to (239), we obtain (235). The combination of (235) and (240) implies (104).

H. Proof of Lemma 16

First we focus on the quantum part and we show that, when a POVM $\{M_0^G\}_{B \subset \mathbb{R}^{2s}}$ is covariant,

$$\wp_{\boldsymbol{\alpha}^Q | M_0^G} \left(\mathsf{T}_{W^Q, c}(\boldsymbol{\alpha}_0^Q) \right) \geq N \left[\mathbf{0}, E_s \left(e^{-\boldsymbol{\beta}} + e/2 \right) \right] \left(\mathsf{T}_{W^Q, c}(\mathbf{0}) \right) \tag{241}$$

where $\boldsymbol{\alpha}^Q = (\boldsymbol{\alpha}^R, \boldsymbol{\alpha}^I)$ and W^Q is a diagonal matrix with eigenvalues $w_j > 0$.

Since M^Q is covariant, we have $\wp_{\boldsymbol{\alpha}^Q | M_0^G} \left(\mathsf{T}_{W^Q, c}(\boldsymbol{\alpha}_0^Q) \right) = \wp_{\mathbf{0} | M_0^G} \left(\mathsf{T}_{W^Q, c}(\mathbf{0}) \right)$. There exists a state τ such that

$$M^Q(\mathsf{B}) = \frac{1}{\pi^s} \int_{\mathsf{B}} d\boldsymbol{\alpha}^Q \, T_{\boldsymbol{\alpha}^Q}^Q \, \tau (T_{\boldsymbol{\alpha}^Q}^Q)^\dagger,$$

where $T_{\alpha \varrho}^{Q}$ is defined as in Eq. (52). Let $\hat{N}_j$ be the number operator on the j-th system. Since $\hat{N}_j \Phi[\mathbf{0}, \boldsymbol{\beta}] = \Phi[\mathbf{0}, \boldsymbol{\beta}]\hat{N}_j$, for the set $\mathsf{T}_{W\varrho,c}(\mathbf{0})$, we have

$$\mathrm{Tr}\, \Phi[\mathbf{0}, \boldsymbol{\beta}]M(\mathsf{T}_{W\varrho,c}(\mathbf{0}))$$

$$= \mathrm{Tr}\, \Phi[\mathbf{0}, \boldsymbol{\beta}]e^{it\hat{N}_j}M(\mathsf{T}_{W\varrho,c}(\mathbf{0}))e^{-it\hat{N}_j}$$

$$= \mathrm{Tr}\, \Phi[\mathbf{0}, \boldsymbol{\beta}]e^{it\hat{N}_j}\frac{1}{\pi^s}\int_{\mathsf{T}_{W\varrho,c}(\mathbf{0})} D_{\alpha \varrho}\tau D_{\alpha \varrho}^{\dagger}d\alpha^Q e^{-it\hat{N}_j}$$

$$= \mathrm{Tr}\, \Phi[\mathbf{0}, \boldsymbol{\beta}]\frac{1}{\pi^s}\int_{\mathsf{T}_{W\varrho,c}(\mathbf{0})} D_{\alpha \varrho}e^{it\hat{N}_j}\tau e^{-it\hat{N}_j} D_{\alpha \varrho}^{\dagger}d\alpha^Q.$$

Since this relation holds with any $\hat{N}_j$. Hence, τ can be replaced by

$$\frac{1}{(2\pi)^s}\int_{[0,2\pi]^s} e^{i\sum_{j=1}^{s} t_j \hat{N}_j}\tau e^{-i\sum_{j=1}^{s} t_j \hat{N}_j}dt_1 \ldots dt_s\,,$$

For a quantum which commutes with $\hat{N}_j$ with any j.

Now, we consider the case with $\tau = |k_1, \ldots, k_s\rangle\langle k_1, \ldots, k_s|$. Then,

$$\mathrm{Tr}\, \Phi[\mathbf{0}, \boldsymbol{\beta}]M(\mathsf{T}_{W\varrho,c}(\mathbf{0}))$$

$$= \frac{1}{\pi^s}\int_{\mathsf{T}_{W\varrho,c}(\mathbf{0})} \langle k_1, \ldots, k_s|D_{\alpha \varrho}^{\dagger}\Phi[\mathbf{0}, \boldsymbol{\beta}]D_{\alpha \varrho}|k_1, \ldots, k_s\rangle d\alpha^Q$$

$$= \frac{1}{\pi^s}\int_{\mathsf{T}_{W\varrho,c}(\mathbf{0})} \langle k_1, \ldots, k_s|\Phi[\boldsymbol{\alpha}^Q, \boldsymbol{\beta}]|k_1, \ldots, k_s\rangle d\alpha^Q$$

$$= \frac{1}{\pi^s}\int_{\mathsf{T}_{W\varrho,c}(\mathbf{0})} \int |\langle k_1, \ldots, k_s|\alpha_1', \ldots, \alpha_s'\rangle|^2 \frac{1}{\pi^s N_1 \ldots N_s}e^{-\frac{\sum_{j=1}^{s}|\alpha_j'-\alpha_j|^2}{N_j}} d\alpha' d\alpha^Q$$

$$= \frac{1}{\pi^s}\int_{\mathsf{T}_{W\varrho,c}(\mathbf{0})} \int e^{-\sum_{j=1}^{s}|\alpha_j'|^2}\frac{|\alpha_1'|^{2k_1}\cdots|\alpha_s'|^{2k_s}}{k_1!\cdots k_s!} \frac{1}{\pi^s N_1 \ldots N_s}e^{-\frac{\sum_{j=1}^{s}|\alpha_j'-\alpha_j|^2}{N_j}} d\alpha' d\alpha^Q$$

$$= \frac{1}{\pi^s}\int e^{-\sum_{j=1}^{s}|\alpha_j'|^2}\frac{|\alpha_1'|^{2k_1}\cdots|\alpha_s'|^{2k_s}}{k_1!\cdots k_s!} f(|\alpha_1'|^2, \ldots, |\alpha_s'|^2)d\alpha'$$

$$= \frac{1}{\pi^s}\int_{[0,\infty)^s} e^{-\sum_{j=1}^{s} r_j}\frac{r_1^{k_1}\cdots r_s^{k_s}}{k_1!\cdots k_s!} f(r_1, \ldots, r_s)d_1 \ldots dr_s$$

$$= \int_{[0,\infty)^s} f(r_1, \ldots, r_s)p_{k_1}(r_1)\cdots p_{k_s}(r_s)dr_1 \ldots dr_s, \tag{242}$$

where $\alpha_j := \alpha_j^R + i\alpha_j^I$, $f(r_1, \ldots, r_s) := \int_{\mathsf{T}_{W\varrho,c}(\mathbf{0})} \frac{1}{\pi^s N_1 \ldots N_s}e^{-\frac{\sum_{j=1}^{s}|\alpha_j'-\alpha_j|^2}{N_j}} d\alpha^Q$ with $r_j = |\alpha_j'|^2$ and $p_k(r)$ is the PDF defined as $e^{-r}\frac{r^k}{\pi k!}$. We find that $f(r_1, \ldots, r_s)$ is monotone increasing for $r_1, \ldots, r_s$. Also, $p_k(r) \geq p_0(r)$ for $r \geq (k!)^{1/k}$, and $p_k(r) <$

$p_0(r)$ for $r < (k!)^{1/k}$. These facts show that

$$\int_{[0,\infty)^s} f(r_1, \ldots, r_s) p_0(r_1) \cdots p_0(r_s) dr_1 \ldots dr_s$$

$$\leq \int_{[0,\infty)^s} f(r_1, \ldots, r_s) p_{k_1}(r_1) \cdots p_{k_s}(r_s) dr_1 \ldots dr_s. \tag{243}$$

When τ is the vacuum state, the equality in (241) hold. So, combining (242) and (243), we obtain (241).

In the classical case, the covariant measurement is unique. So, we have the extension as in Lemma 16.

I. Proof of Lemma 17

(i)$\Rightarrow$ (ii): Since $S^{-1}A_2(S^T)^{-1} = (S^T A_2^{-1} S)^{-1}$, $S^T A_1 S$, and D commute with each other, we have

$$DS^{-1}A_2 A_1 S = DS^{-1}A_2(S^T)^{-1}S^T A_1 S$$

$$= S^T A_1 S S^{-1} A_2(S^T)^{-1} D = S^T A_1 A_2(S^T)^{-1} D. \tag{244}$$

Since $S^T DS = D$, we have $S^T D = DS^{-1}$ and $DS = (S^T)^{-1}D$. Thus,

$$S^T DA_2 A_1 S = S^T A_1 A_2 DS, \tag{245}$$

which implies (ii).

(ii)$\Rightarrow$ (i): Let S be a symplectic matrix to symplectically diagonalize A_1. Combining (244) and (245), we have

$$DS^{-1}A_2(S^T)^{-1}S^T A_1 S = S^T A_1 S S^{-1} A_2(S^T)^{-1} D.$$

Since $D^2 = -1$, we have

$$S^{-1}A_2(S^T)^{-1}S^T A_1 SD = DS^T A_1 S S^{-1} A_2(S^T)^{-1} = S^T A_1 SDS^{-1}A_2(S^T)^{-1}.$$

Hence, $S^{-1}A_2(S^T)^{-1}$ commute with $S^T A_1 SD$. There exists an orthogonal matrix S' such that SS' is a symplectic matrix, and $(SS')^T A_1(SS')$ and $(SS')^{-1}A_2((SS')^T)^{-1}$ are diagonal matrices. Considering the inverse of A_2^{-1}, we obtain (i).

J. Derivation of Eq. (185)

First, SLD operators for x, y, and the nuisance parameter z can be calculated by solving the equation $\frac{\partial \rho_t}{\partial t_j} = \frac{1}{2}\left(\rho_t L_j + L_j \rho_t\right)$:

$$L_x = -\frac{x'}{1 - |\boldsymbol{n}|^2}I + \left(\boldsymbol{a}' + \frac{x'\boldsymbol{n}}{1 - |\boldsymbol{n}|^2}\right) \cdot \boldsymbol{\sigma}$$

$$L_y = -\frac{y'}{1 - |\boldsymbol{n}|^2}I + \left(\boldsymbol{b}' + \frac{y'\boldsymbol{n}}{1 - |\boldsymbol{n}|^2}\right) \cdot \boldsymbol{\sigma}$$

$$L_z = -\frac{z}{1 - |\boldsymbol{n}|^2}I + \left(\boldsymbol{c} + \frac{z\boldsymbol{n}}{1 - |\boldsymbol{n}|^2}\right) \cdot \boldsymbol{\sigma},$$

where $s := \boldsymbol{a} \cdot \boldsymbol{b}$, $\boldsymbol{a}' = \frac{\boldsymbol{a} - s\boldsymbol{b}}{1-s^2}$, $\boldsymbol{b}' = \frac{\boldsymbol{b} - s\boldsymbol{a}}{1-s^2}$, $x' = \frac{x - ys}{1-s^2}$, and $y' = \frac{y - xs}{1-s^2}$. By definition, the SLD Fisher information can be evaluated as $(J_{\boldsymbol{t}})_{ij} := \text{Tr}\, \rho_{\boldsymbol{t}}(L_i L_j + L_j L_i)/2$ and the D-matrix can be evaluated as $(D_{\boldsymbol{t}})_{ij} := i\, \text{Tr}\, \rho_{\boldsymbol{t}}[L_i, L_j]$. Explicitly, we get

$$
J = \begin{pmatrix}
\frac{1}{1-s^2} + \frac{x'^2}{1-|\boldsymbol{n}|^2} & -\frac{s}{1-s^2} + \frac{x'y'}{1-|\boldsymbol{n}|^2} & \frac{x'z}{1-|\boldsymbol{n}|^2} \\[2mm]
-\frac{s}{1-s^2} + \frac{x'y'}{1-|\boldsymbol{n}|^2} & \frac{1}{1-s^2} + \frac{y'^2}{1-|\boldsymbol{n}|^2} & \frac{y'z}{1-|\boldsymbol{n}|^2} \\[2mm]
\frac{x'z}{1-|\boldsymbol{n}|^2} & \frac{y'z}{1-|\boldsymbol{n}|^2} & 1 + \frac{z^2}{1-|\boldsymbol{n}|^2}
\end{pmatrix},
$$

$$
J^{-1} = \begin{pmatrix}
\frac{1-|\boldsymbol{n}|^2+y'^2(1-s^2)+z^2}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & -\frac{x'y'(1-s^2)+(1-|\boldsymbol{n}|^2+z^2)s}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & -\frac{(x'+sy')z}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} \\[3mm]
-\frac{x'y'(1-s^2)+(1-|\boldsymbol{n}|^2+z^2)s}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & \frac{1-|\boldsymbol{n}|^2+x'^2(1-s^2)+z^2}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & -\frac{(sx'+y')z}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} \\[3mm]
-\frac{(x'+sy')z}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & -\frac{(sx'+y')z}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2} & \frac{1-|\boldsymbol{n}|^2+x'^2+2sx'y'+y'^2}{1-|\boldsymbol{n}|^2+x'^2+2x'y's+y'^2+z^2}
\end{pmatrix},
$$

and

$$
D = \frac{1}{\sqrt{1-s^2}} \begin{pmatrix}
0 & -2z & 2y \\
2z & 0 & -2x \\
-2y & 2x & 0
\end{pmatrix}.
$$

Finally, substituting the above into Eq. (144), we get Eq. (185).

References

1. D'Ariano, G.M., Presti, P.L.: Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation. Phys. Rev. Lett. **86**(19), 4195 (2001)
2. Knill, E., Leibfried, D., Reichle, R., Britton, J., Blakestad, R.B., Jost, J.D., Langer, C., Ozeri, R., Seidelin, S., Wineland, D.J.: Randomized benchmarking of quantum gates. Phys. Rev. A **77**(1), 012307 (2008)
3. Heisenberg, W.: Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. Zeitschrift fur Physik **43**, 172–198 (1927)
4. Schnabel, R., Mavalvala, N., McClelland, D.E., Lam, P.K.: Quantum metrology for gravitational wave astronomy. Nat. Commun. **1**, 121 (2010)
5. Holevo, A.S.: Probabilistic and Statistical Aspects of Quantum Theory. Springer, Berlin (2011). (**originally published in Russian, 1980**)
6. Helstrom, C.W.: Quantum Detection and Estimation Theory. Academic Press, Cambridge (1976)
7. Massar, S., Popescu, S.: Optimal extraction of information from finite quantum ensembles. Phys. Rev. Lett. **74**(8), 1259 (1995)
8. Gill, R.D., Massar, S.: State estimation for large ensembles. Phys. Rev. A **61**(4), 042312 (2000)
9. Fujiwara, A.: Estimation of SU(2) operation and dense coding: an information geometric approach. Phys. Rev. A **65**(1), 012316 (2001)
10. Matsumoto, K.: A new approach to the Cramér–Rao-type bound of the pure-state model. J. Phys. A: Math. Gen. **35**(13), 3111 (2002)
11. Hayashi, M., Matsumoto, K.: Statistical model with measurement degree of freedom and quantum physics. In: Hayashi, M. (ed.) Asymptotic Theory of Quantum Statistical Inference: Selected Papers, pp. 162–169. World Scientific (2005). (**originally published in Japanese in Surikaiseki Kenkyusho Kokyuroku No. 1055, 1998**)
12. Gill, R.D., Guţă, M.: On asymptotic quantum statistical inference. In: Banerjee, M., Bunea, F., Huang, J., Koltchinskii, V., Maathuis, M. H. (eds.) From Probability to Statistics and Back: High-Dimensional Models and Processes—A Festschrift in Honor of Jon A. Wellner, pp. 105–127. Institute of Mathematical Statistics (2013)

13. Szczykulska, M., Baumgratz, T., Datta, A.: Multi-parameter quantum metrology. Adv. Phys.: X **1**(4), 621–639 (2016)
14. Hayashi, M., Matsumoto, K.: Asymptotic performance of optimal state estimation in quantum two level system. (2004). arXiv:quant-ph/0411073
15. Hayashi, M., Matsumoto, K.: Asymptotic performance of optimal state estimation in qubit system. J. Math. Phys. **49**(10), 102101 (2008)
16. Kahn, J., Guţă, M.: Local asymptotic normality for finite dimensional quantum systems. Commun. Math. Phys. **289**(2), 597–652 (2009)
17. Kahn, J.: Quantum local asymptotic normality and other questions of quantum statistics. Theses, Université Paris XI (2009)
18. Hayashi, M.: Quantum estimation and the quantum central limit theorem. Am. Math. Soc. Transl. Ser. **2**(227), 95–123 (2009). (**originally published in Japanese in Surikaiseki Kenkyusho Kokyuroku No. 1055, 1998**)
19. Yamagata, K., Fujiwara, A., Gill, R.D.: Quantum local asymptotic normality based on a new quantum likelihood ratio. Ann. Stat. **41**(4), 2197–2217 (2013)
20. Yue, J.-D., Zhang, Y.-R., Fan, H.: Quantum-enhanced metrology for multiple phase estimation with noise. Sci. Rep. **4**, 5933 (2014)
21. Knott, P.A., Proctor, T.J., Hayes, A.J., Ralph, J.F., Kok, P., Dunningham, J.A.: Local versus global strategies in multiparameter estimation. Phys. Rev. A **94**(6), 062312 (2016)
22. Zhang, L., Chan, K.W.C.: Quantum multiparameter estimation with generalized balanced multimode noon-like states. Phys. Rev. A **95**(3), 032321 (2017)
23. Liu, J., Yuan, H.: Control-enhanced multiparameter quantum estimation. Phys. Rev. A **96**(4), 042114 (2017)
24. Proctor, T.J., Knott, P.A., Dunningham, J.A.: Multiparameter estimation in networked quantum sensors. Phys. Rev. Lett. **120**(8), 080501 (2018)
25. Tsuda, Y., Matsumoto, K.: Quantum estimation for non-differentiable models. J. Phys. A: Math. Gen. **38**(7), 1593 (2005)
26. Haah, J., Harrow, A.W., Ji, Z., Xiaodi, W., Nengkun, Y.: Sample-optimal tomography of quantum states. IEEE Trans. Inf. Theory **63**(9), 5628–5641 (2017)
27. Suzuki, J.: Parameter estimation of qubit states with unknown phase parameter. Int. J. Quantum Inf. **13**(01), 1450044 (2015)
28. Kok, P., Dunningham, J., Ralph, J.F.: Role of entanglement in calibrating optical quantum gyroscopes. Phys. Rev. A **95**(1), 012326 (2017)
29. Yang, Y., Chiribella, G., Hayashi, M.: Quantum stopwatch: how to store time in a quantum memory. Proc. R. Soc. Lond. A: Math. Phys. Eng. Sci. **474**(2213), 20170773 (2018)
30. Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. **35**, 99–109 (1943)
31. Hellinger, E.: Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. Journal für die reine und angewandte Mathematik **136**, 210–271 (1909)
32. Fuchs, C.A.: Distinguishability and accessible information in quantum theory (1996). arXiv:quant-ph/9601020
33. van der Vaart, A.W.: Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (1998)
34. Hayashi, M.: Quantum Information Theory: Mathematical Foundation. Springer, Berlin (2017)
35. Hayashi, M.: Comparison between the Cramér–Rao and the mini-max approaches in quantum channel estimation. Commun. Math. Phys. **304**(3), 689–709 (2011)
36. Le Cam, L.: Asymptotic Methods in Statistical Decision Theory. Springer, New York (1986)
37. Suzuki, J.: Explicit formula for the Holevo bound for two-parameter qubit-state estimation problem. J. Math. Phys. **57**(4), 042201 (2016)
38. Petz, D.: An invitation to the algebra of canonical commutation relations, Leuven notes in mathematical and theoretical physics, A2. Leuven University Press, Leuven, Belgium (1990)
39. Guţă, M., Kahn, J.: Local asymptotic normality for qubit states. Phys. Rev. A **73**(5), 052108 (2006)
40. Guţă, M., Janssens, B., Kahn, J.: Optimal estimation of qubit states with continuous time measurements. Commun. Math. Phys. **277**(1), 127–160 (2008)
41. Guţă, M., Bowles, P., Adesso, G.: Quantum-teleportation benchmarks for independent and identically distributed spin states and displaced thermal states. Phys. Rev. A **82**(4), 042310 (2010)
42. Yang, Y., Bai, G., Chiribella, G., Hayashi, M.: Compression for quantum population coding. IEEE Trans. Inf. Theory **64**(7), 4766–4783 (2018)
43. Chapman, D.G., Robbins, H.: Minimum variance estimation without regularity assumptions. Ann. Math. Stat. **22**, 581–586 (1951)
44. Hammersley, J.M.: On estimating restricted parameters. J. R. Stat. Soc. Ser. B (Methodological) **12**(2), 192–240 (1950)

45. Hayashi, M.: Fourier analytic approach to quantum estimation of group action. Commun. Math. Phys. **347**(1), 3–82 (2016)
46. Cox, D.R., Barndorff-Nielsen, O.E.: Inference and Asymptotics, vol. 52. CRC Press, Boca Raton (1994)
47. Lehmann, E.L., Casella, G.: Theory of Point Estimation. Springer, Berlin (2006)
48. Watanabe, Y., Sagawa, T., Ueda, M.: Uncertainty relation revisited from quantum estimation theory. Phys. Rev. A **84**(4), 042121 (2011)
49. Hayashi, M.: Asymptotic estimation theory for a finite-dimensional pure state model. J. Phys. A: Math. Gen. **31**(20), 4633 (1998)
50. Hayashi, M.: A Group Theoretic Approach to Quantum Information. Springer, Berlin (2017)
51. Humphreys, P.C., Barbieri, M., Datta, A., Walmsley, I.A.: Quantum enhanced multiple phase estimation. Phys. Rev. Lett. **111**(7), 070403 (2013)
52. Giri, N., von Waldenfels, W.: An algebraic version of the central limit theorem. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **42**(2), 129–134 (1978)
53. Williamson, J.: On the algebraic problem concerning the normal forms of linear dynamical systems. Am. J. Math. **58**(1), 141–163 (1936)
54. Husimi, K.: Some formal properties of the density matrix. Proc. Physico-Math. Soc. Jpn. 3rd Ser. **22**(4), 264–314 (1940)
55. Glauber, R.J.: Coherent and incoherent states of the radiation field. Phys. Rev. **131**(6), 2766 (1963)
56. Schleich, W.P.: Quantum Optics in Phase Space. Wiley, Hoboken (2011)