Asymptotic estimation theory for a finite-dimensional pure state model

1998 J. Phys. A: Math. Gen. 31 4633

(http://iopscience.iop.org/0305-4470/31/20/006)

J. Phys. A: Math. Gen. **31** (1998) 4633–4655. Printed in the UK PII: S0305-4470(98)87977-8

Asymptotic estimation theory for a finite-dimensional pure state model

Masahito Hayashi†
Department of Mathematics, Kyoto University, Kyoto 606-8502, Japan

Received 29 September 1997

Abstract. The optimization of measurement for n samples of pure states are studied. The error of the optimal measurement for n samples is asymptotically compared with the one of the maximum likelihood estimators from n data given by the optimal measurement for one sample.

1. Introduction

Recently, there has been a rise in the necessity for studies about statistical estimation for the unknown state, related to the corresponding advance in measuring technologies in quantum optics. An investigation including both quantum theory and mathematical statistics is necessary for an essential understanding of quantum theory because it has statistical aspects [1, 2]. Therefore, it is indeed important to optimize the measuring process with respect to the estimation of the unknown state. Such research is known as quantum estimation, and was initiated by Helstrom in the late 1960s, originating in the optimization of the detecting process in optical communications [1]. In the classical statistical estimation, one searches for the most suitable estimator for which probability measure describes the objective probabilistic phenomenon. In quantum estimation, one searches for the most suitable measurement for which density operator describes the objective quantum state.

Contained among important results are three estimation problems. The first is of the complex amplitude of coherent light in thermal noise and the second is of the expectation parameters of quantum Gaussian state. The former was studied by Yuen and Lax [3] and the latter by Holevo [2]. These studies discovered that heterodyning is the most suitable for the estimation of the complex amplitude of coherent light in thermal noise. The third is a formulation of the covariant measurement with respect to an action of a group. It was studied by Holevo [2, 4]. In the formulation, he established a quantum analogue of Hunt–Stein theorem.

Quantum estimation, was first used in the evaluation of the estimation error of a single sample of the unknown state as it had advanced in connection with the optimization of the measuring process in optical communications. Thus, early studies were lacking in asymptotic aspects, i.e. there was little research with respect to reducing the estimation error by quantum correlations between samples.

Recently, studies concerning the estimation of the unknown state have attracted many physicists [5–8]. Some of them were drawn by the variation of the measuring precision with respect to the number of samples of the unknown state [9, 10].

† E-mail address: masahito@kusm.kyoto-u.ac.jp

Nagaoka [11] studied, for the first time, asymptotic aspects of quantum estimation. He paid particular attention to the quantum correlations between samples of the unknown state, and studied the relation between the asymptotic estimation and the local detection of a one-parameter family of quantum states.

In the early 1990s, Fujiwara and Nagaoka [12–14] studied the estimation problem for a multiparameter family consisting of pure states. They pioneered studies into the estimation problem of the complex amplitude of noiseless coherent light. However, there had been some studies with respect to that of coherent light in thermal noise. The research found that heterodyning is the most suitable for the estimation of the complex amplitude of noiseless coherent light. In 1996, Matsumoto [15] established a more general formulation of the estimation for a multiparameter family consisting of pure states. Moreover in 1991, Nagaoka [16] treated the estimation problem for two-parameter families of mixed states in a spin-$\frac{1}{2}$ system, and in 1997 Hayashi [17, 18] treated it for three-parameter families of mixed states in a spin-$\frac{1}{2}$ system. However, there are no asymptotic aspects in these works concerning multiparameter families. There is more need of this type of investigation into one- and multiparameter families.

Can quantum estimation reduce the estimation error by using the quantum correlations between samples, under the preparation of sufficient samples of the unknown state? To answer this question, in this paper, we treat a family, consisting of pure states on a Hilbert space $\mathcal{H}$† under the preparation of n samples of the unknown state, with the estimation problem. In section 2, we use, as a tool, the composite system consisting of n samples as a single system. The quantum i.i.d. condition is introduced as the quantum counterpart of the independent and identical distributions condition (3). In section 3, we review Holevo's result concerning covariant measurements which will be used in the following sections. In section 4, we apply Holevo's result to the optimization of measurements on the composite system, which results in obtaining the most suitable measurement (theorem 3). We asymptotically calculate the estimation error by the optimal measurement in the sense of both the error mean square and large deviation (see (9)–(11) and (13)). The first term of the right-hand side of (10) is consistent with the value conjectured from the results in Fujiwara and Nagaoka [14] and Matsumoto [15]. However, the optimal measurement may be too difficult for modern technology to realize when using more than one sample.

In section 5, we use this estimation problem under the following guidelines. The samples are divided into pairs consisting of a maximum of m samples. By measuring each pair with the optimal measurement of section 4, we create some data. The estimated value is given by manipulating these data. The restricted condition is m-semiclassical (see (14)). We compare an m-semiclassical measurement with the optimal measurement of section 4 with respect to the estimation error under the preparation of a sufficient amount of samples. When we use the maximum likelihood estimator to manipulate the data, the error mean square of both asymptotically coincide in the first order (see (10) and (19)). However, when the radius of allowable errors is finite, the error of large deviations in the latter type is smaller than that in the former type (see (11) and (20)). Both coincide in the case of the maximum likelihood estimator under the limit where the radius goes to infinitesimal (13), (21). Can we asymptotically realize a small estimation error as the optimal measurements in section 4 has? It is, physically, sufficient to construct the optimal measurement for one sample. In section 5, we show how to construct it (see (25)).

Most of the proofs of this paper are given in the appendices. In view of multiparameter families of mixed states in spin-$\frac{1}{2}$ system, Hayashi [19] discussed the same problem using

† Where $\mathcal{H}$ denotes a finite-dimensional Hilbert space which corresponds to the physical system of interest.

Cramér–Rao-type bound.

2. Pure state n-i.i.d. model

In this section, we use the mathematical formulation of the estimation for pure states. Let k be the dimension of the Hilbert space $\mathcal{H}$, and $\mathcal{P}(\mathcal{H})$ be the set of pure states on $\mathcal{H}$.

In quantum physics, the most general description of a quantum measurement is probability given by the mathematical concept of a *positive operator-valued measure* (POVM) [1, 2] on the system of state space. Generally, if Ω is measurable space, a measurement M satisfies the following

$$M(B) = M(B)^* \qquad M(B) \geqslant 0 \qquad M(\emptyset) = 0 \qquad M(\Omega) = \mathrm{Id} \qquad \text{on } \mathcal{H} \forall B \subset \Omega$$
$$M(\cup_i B_i) = \sum_i M(B_i) \qquad \text{for } B_i \cap B_j = \emptyset (i \neq j) \qquad \{B_i\} \text{ is countable subsets of } \Omega.$$

In this paper, $\mathcal{M}(\Omega, \mathcal{H})$ denotes the set of POVMs on $\mathcal{H}$ whose measurable set is Ω. A measurement $M \in \mathcal{M}(\Omega, \mathcal{H})$ is said to be simple if $M(B)$ is a projection for any Borel $B \subset \Omega$. A measurement M is random if it is described as a convex combination of simple measurements. A random measurement $M = \sum_i a_i M_i$ (M_i is simple and $a_i > 0$) can be realized when every measurement M_i is taken with the probability a_i.

In this paper, we consider measurements whose measurable set is $\mathcal{P}(\mathcal{H})$ since it is known that the unknown state is included in $\mathcal{P}(\mathcal{H})$.

Next, we define two distances characterizing the homogeneous space $\mathcal{P}(\mathcal{H})$.

Definition 1. The Fubini–Study distance d_{fs} (which is the geodesic distance of the Fubini–Study metric) is defined as:

$$\cos d_{fs}(\rho, \hat{\rho}) = \sqrt{\operatorname{tr} \rho\hat{\rho}} \qquad 0 \leqslant d_{fs}(\rho, \hat{\rho}) \leqslant \frac{\pi}{2}. \tag{1}$$

The Bures distance d_b is defined in the usual way:

$$d_b(\rho, \hat{\rho}) := \sqrt{1 - \operatorname{tr} \rho\hat{\rho}}. \tag{2}$$

It was introduced by Bures [20] in a mathematical context.

Let $W(\rho, \hat{\rho})$ be a measure of deviation of the measured value $\hat{\rho}$ from the actual value ρ, then we have the following equivalent conditions.

- $W(\rho, \hat{\rho}) = W(g\rho g^*, g\hat{\rho}g^*) \forall \rho, \forall \hat{\rho} \in \mathcal{P}(\mathcal{H}) \forall g \in SU(k)$.
- There exists a function h on $[0, 1]$ such that $W(\rho, \hat{\rho}) = h \circ d_{fs}(\rho, \hat{\rho})$.

It is natural to assume that a deviation measure $W(\rho, \hat{\rho})$ is monotone increasing with respect to the Fubini–Study distance d_{fs}.

If $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are n Hilbert spaces which correspond to the physical systems, then their composite system is represented by the tensor Hilbert space:

$$\mathcal{H}^{(n)} := \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n = \bigotimes_{i=1}^n \mathcal{H}_i.$$

Thus, a state on the composite system is denoted by a density operator ρ on $\mathcal{H}^{(n)}$. In particular if n element systems $\{\mathcal{H}_i\}$ of the composite system $\mathcal{H}^{(n)}$ are independent of each other, there exists a density ρ_i on $\mathcal{H}_i$ such that

$$\rho^{(n)} = \rho_1 \otimes \cdots \otimes \rho_n = \bigotimes_{i=1}^n \rho_i.$$

The condition:

$$\mathcal{H}_1 = \cdots = \mathcal{H}_n = \mathcal{H}, \rho_1 = \cdots = \rho_n = \rho \tag{3}$$

corresponds to the independent and identically distributed (i.i.d.) condition in the classical case. In this paper, we use this estimation problem under condition (3), called the quantum i.i.d. condition. This condition means that identical n samples are independently prepared. The model $\{\rho^{(n)} = \underbrace{\rho \otimes \cdots \otimes \rho}_{n} | \rho \in \mathcal{P}(\mathcal{H})\}$ is called n-i.i.d. model. As ρ is a pure state, $\mathcal{H}^{(n)}$ and $\rho^{(n)}$ are simplified as follows. Letting $\rho = |\phi\rangle\langle\phi| \in \mathcal{P}(\mathcal{H})$, we have

$$\rho^{(n)} = |\phi^{(n)}\rangle\langle\phi^{(n)}| \qquad \phi^{(n)} := \overbrace{\phi \otimes \cdots \otimes \phi}^{n}.$$

Because all of the vectors $\phi^{(n)}$ are included in n-times symmetric tensor space, for any measurement $M \in \mathcal{M}(\Omega, \mathcal{H}^{(n)})$ on the n-times tensor space $\mathcal{H}^{(n)}$, the measurement $\tilde{M}(\mathrm{d}\omega) := P_{\mathcal{H}_s^{(n)}} M(\mathrm{d}\omega) P_{\mathcal{H}_s^{(n)}} \in \mathcal{M}(\Omega, \mathcal{H}_s^{(n)})$ on the n-times symmetric tensor space $\mathcal{H}_s^{(n)}$ satisfies that:

$$\mathrm{tr}\, M(\mathrm{d}\omega)\rho^{(n)} = \mathrm{tr}\, \tilde{M}(\mathrm{d}\omega)\rho^{(n)} \qquad \forall \rho \in \mathcal{H}$$

where $\mathcal{H}_s^{(n)}$ denotes the n-times symmetric tensor space on $\mathcal{H}$. Therefore, all possible measurements can be regarded as elements of $\mathcal{M}(\mathcal{P}(\mathcal{H}), \mathcal{H}_s^{(n)})$. The mean error of the measurement $\Pi \in \mathcal{M}(\mathcal{P}(\mathcal{H}), \mathcal{H}_s^{(n)})$ with respect to a deviation measure $W(\rho, \hat{\rho})$, provided that the actual state is ρ, is equal to

$$\mathcal{D}_\rho^{W,(n)}(\Pi) := \int_{\mathcal{P}(\mathcal{H})} W(\rho, \hat{\rho})\, \mathrm{tr}(\Pi(\mathrm{d}\hat{\rho})\rho^{(n)}).$$

In minimax approach the maximum possible error with respect to a deviation measure $W(\rho, \hat{\rho})$

$$\mathcal{D}^{W,(n)}(\Pi) := \max_{\rho \in \mathcal{P}(\mathcal{H})} \mathcal{D}_\rho^{W,(n)}(\Pi)$$

is minimized.

3. Quantum Hunt–Stein theorem

In this section, the quantum Hunt–Stein theorem, established by Holevo [2, 4], is summarized. Let G be a compact transitive Lie group of all transformations on a compact parametric set Θ, and $\{V_g\}$ a continuous unitary irreducible representation of G in a finite-dimensional Hilbert space $\mathcal{H}' := \mathbb{C}^{k'}$, and μ a σ-finite invariant measure on group G such that $\mu(G) = 1$. In this section, we consider the following measurement condition.

Definition 2. A measurement $\Pi \in \mathcal{M}(\Theta, \mathcal{H}')$ is covariant with respect to $\{V_g\}$ if

$$V_g^* \Pi(B) V_g = \Pi(B_{g^{-1}})$$

for any $g \in G$ and any Borel $B \subset \Theta$, where

$$B_g := \{g\theta | \theta \in B\}.$$

$\mathcal{M}(\Theta, V)$ denotes the set of covariant measurements with respect to $\{V_g\}$.

Covariant measurements are characterized by the following theorem.

Theorem 1. The map V^θ from the set $\mathcal{S}(\mathcal{H}')$ of densities on $\mathcal{H}'$ to $\mathcal{M}(\Theta, V)$ is surjective for any $\theta \in \Theta$, where a POVM $V^\theta(P)$ is defined as follows

$$V^\theta(P)(B) := k' \int_{\{g\theta \in B\}} V_g P V_g^* \mu(\mathrm{d}g) \qquad \forall B \in \mathcal{B}(\Theta)$$

for any $P \in \mathcal{S}(\mathcal{H}')$.

In this section, we use the following condition for a family of states.

Definition 3. The family is called covariant under the representation $\{V_g\}$ of group G acting on Θ, if

$$S_{g\theta} = V_g S_\theta V_g^* \qquad \forall \theta \in \Theta, \forall g \in G.$$

Assuming that the object is prepared in one of the states $\{S_\theta | \theta \in \Theta\}$ but the actual value of θ is unknown, then the difficulty is estimating this value as close as possible to a measurement on the object. We shall solve this problem by means of quantum statistical decision theory.

Let $W(\theta, \hat{\theta})$ be a measure of deviation of the measured value $\hat{\theta}$ from the actual value θ. It is natural to assume that $W(\theta, \hat{\theta})$ is invariant:

$$W(\theta, \hat{\theta}) = W(g\theta, g\hat{\theta}) \qquad \forall \theta, \forall \hat{\theta} \in \Theta, \forall g \in G. \tag{4}$$

The mean error of the measurement $\Pi \in \mathcal{M}(\Theta, \mathcal{H}')$ with respect to a deviation measure $W(\theta, \hat{\theta})$, provided that the actual state is S_θ, is equal to

$$\mathcal{D}_\theta^{W,S}(\Pi) := \int_\Theta W(\theta, \hat{\theta})\, \mathrm{tr}(\Pi(\mathrm{d}\hat{\theta}) S_\theta).$$

Following the classical statistical decision theory, we can form two functionals of $\mathcal{D}_\theta^W$ giving a total measure of precision of the measurement Π.

In Bayes' approach we take the mean of $\mathcal{D}_\theta^W$ with respect to a given prior distribution $\pi(\mathrm{d}\theta)$. The measurement minimizing the resulting functional:

$$\mathcal{D}_\pi^{W,S}(\Pi) := \int_\Theta \mathcal{D}_\theta^{W,S}(\Pi) \pi(\mathrm{d}\theta)$$

is called Bayesian. This quantity represents the mean error in the situation where θ is a random parameter with known distribution $\pi(\mathrm{d}\theta)$. In particular, as Θ, G are compact and 'nothing is known' about θ, it is natural to take for $\pi(\mathrm{d}\theta)$ the 'uniform' distribution, i.e. normalized invariant measure $\nu(\mathrm{d}\theta)$ defined as follows

$$\nu(B) := \mu(\{g\theta \in B\}).$$

It is independent of the choice of $\theta \in \Theta$.

In minimax approach the maximum possible error with respect to a deviation measure $W(\theta, \hat{\theta})$

$$\mathcal{D}^{W,S}(\Pi) := \max_{\theta \in \Theta} \mathcal{D}_\theta^{W,S}(\Pi)$$

is minimized. The minimizing measurement is called minimax.

Because G is compact, we shall show that in the covariant case the minima of Bayes and minimax criteria coincide and are achieved on a covariant measurement. We obtain the following quantum Hunt–Stein theorem [2, 4]. It is easy to prove the theorem.

Theorem 2. For a covariant measurement $\Pi \in \mathcal{M}(\Theta, V)$, we obtain the following equations:

$$\mathcal{D}_\theta^{W,S}(\Pi) = \mathcal{D}_\nu^{W,S}(\Pi) = \mathcal{D}^{W,S}(\Pi).$$

For $\Pi \in \mathcal{M}(\Theta, \mathcal{H}')$, denote

$$\Pi_g(B) := V_g \Pi(B_g) V_g^* \qquad \text{for } B \in \mathcal{B}(\Theta).$$

Introducing the 'averaged' measurement

$$\bar{\Pi}(B) := \int_G \Pi_{g^{-1}}(B)\mu(\mathrm{d}g)$$

we have

$$\mathcal{D}_\nu^{W,S}(\bar{\Pi}) = \int_G \mathcal{D}_\nu^{W,S}(\Pi_{g^{-1}})\mu(\mathrm{d}g) = \mathcal{D}_\nu^{W,S}(\Pi).$$

Thus,

$$\mathcal{D}^{W,S}(\Pi) \geqslant \mathcal{D}_\nu^{W,S}(\Pi) = \mathcal{D}_\nu^{W,S}(\bar{\Pi}).$$

In this case, minimax approach and Bayes' approach with respect to $\nu(\mathrm{d}\theta)$ are equivalent. Therefore we minimize the following

$$\mathcal{D}_\theta^{W,S} \circ V^\theta(P) = k' \int_G W(\theta, g\theta) \operatorname{tr} S_\theta V_g P V_g^* \mu(\mathrm{d}g) = \operatorname{tr} \hat{W}(\theta) P$$

where

$$\begin{aligned}\hat{W}(\theta) :&= k' \int_G W(\theta, g\theta) V_g^* S_\theta V_g \mu(\mathrm{d}g) \\ &= k' \int_\Theta W(\theta, \hat{\theta}) S_{\hat{\theta}} \nu(\mathrm{d}\hat{\theta}).\end{aligned}$$

Thus, it is sufficient to consider the following minimization:

$$\min_{P \in \mathcal{S}(\mathcal{H})} \operatorname{tr} \hat{W}(\theta) P = \min_{P \in \mathcal{P}(\mathcal{H}')} \operatorname{tr} \hat{W}(\theta) P.$$

4. Optimal measurement in pure state n-i.i.d. model

In this section we apply the theory of section 3 to the problem of section 2.

We let

$$\Theta := \mathcal{P}(\mathcal{H}) \qquad \mathcal{H}' := \mathcal{H}_s^{(n)} \qquad G := \mathrm{SU}(k) \qquad S_\rho := \rho^{(n)}.$$

Then, the invariant measure ν on $\mathcal{P}(\mathcal{H})$ is equivalent to the measure defined by the volume bundle induced by the Fubini–Study metric. We let the action $\{V_g\}$ of $G = \mathrm{SU}(k)$ to $\mathcal{H}_s^{(n)}$ be the tensor representation of the natural representation. In this case, we have $k' = \binom{n+k-1}{k-1}$.

Theorem 3. If a deviation measure $W(\rho, \hat{\rho})$ is monotone increasing with respect to the Fubini–Study distance d_{fs}, we find

$$\min_{P_0 \in \mathcal{P}(\mathcal{H}_s^{(n)})} \operatorname{tr} \hat{W}(\rho) P_0 = \operatorname{tr} \hat{W}(\rho) \rho^{(n)}.$$

For a proof see appendix A. Thus, $V^\rho(\rho^{(n)})$ is the optimal measurement with respect to a deviation measure $W(\rho, \hat{\rho})$. The optimal measurement is independent of the choice of ρ and W since $V^{\rho_0}(\rho_0^{(n)}) = V^\rho(\rho^{(n)})$. This optimal measurement is denoted by Π_n and is described as follows

$$\Pi_n(\mathrm{d}\hat{\rho}) := \binom{n+k-1}{k-1} \hat{\rho}^{(n)} \nu(\mathrm{d}\hat{\rho}).$$

Under the following chart (6), the optimal measurements are denoted as:

$$\Pi_n(\mathrm{d}\theta) = \binom{n+k-1}{k-1} |\phi(\theta)^{(n)}\rangle\langle\phi(\theta)^{(n)}|\nu(\mathrm{d}\theta) \tag{5}$$

for $\theta \in \{\theta \in \mathbb{R}^{2k-2} | \theta_i \in [0, 2\pi) 1 \leqslant j \leqslant k-1, \theta_j \in [0, \pi/2]\}$, where

$$\phi(\theta) := \begin{pmatrix} \cos\theta_1 \\ \mathrm{e}^{\mathrm{i}\theta_k} \sin\theta_1 \cos\theta_2 \\ \mathrm{e}^{\mathrm{i}\theta_{k+1}} \sin\theta_1 \sin\theta_2 \cos\theta_3 \\ \vdots \\ \mathrm{e}^{\mathrm{i}\theta_{2k-3}} \sin\theta_1 \sin\theta_2 \sin\theta_3 \ldots \sin\theta_{k-2} \cos\theta_{k-1} \\ \mathrm{e}^{\mathrm{i}\theta_{2k-2}} \sin\theta_1 \sin\theta_2 \sin\theta_3 \ldots \sin\theta_{k-2} \sin\theta_{k-1} \end{pmatrix}. \tag{6}$$

The invariant measure $\nu(\mathrm{d}\theta)$ described above is from [21, p 31]

$$\nu(\mathrm{d}\theta) = \frac{(k-1)!}{\pi^{k-1}} \sin^{2k-3}\theta_1 \sin^{2k-5}\theta_2 \ldots \sin\theta_{k-1} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{k-1}\, \mathrm{d}\theta_1\, \mathrm{d}\theta_2 \ldots \mathrm{d}\theta_{2k-2}. \tag{7}$$

Lemma 1. If the deviation measure W is characterized as $W(\rho, \hat{\rho}) = h \circ d_{fs}(\rho, \hat{\rho})$, we can describe the maximum possible error of the optical measurement Π_n as:

$$\mathcal{D}^{W,(n)}(\Pi_n) = 2(k-1)\binom{n+k-1}{k-1} \int_0^{\frac{\pi}{2}} h(\theta) \cos^{2n+1}\theta \sin^{2k-3}\theta \mathrm{d}\theta.$$

For a proof, see appendix B.

Next, we asymptotically calculate the error of the optimal measurements Π_n in the third order.

Theorem 4. When the deviation measure W is described as $W = d_b^{\gamma}$, we can asymptotically calculate the maximum possible error of the optimal measurement as:

$$\lim_{n\to\infty} \mathcal{D}^{d_b^{\gamma},(n)}(\Pi_n) n^{\frac{\gamma}{2}} = \frac{\Gamma(k-1+\gamma/2)}{\Gamma(k-1)}. \tag{8}$$

Particularly, in the case of $\gamma = 2$, we have

$$\mathcal{D}^{d_b^2,(n)}(\Pi_n) n = \frac{(k-1)n}{n+k} = (k-1)\sum_{i=1}^{\infty} \left(-\frac{k}{n}\right)^i \to k-1 \qquad \text{as } n \to \infty. \tag{9}$$

When the deviation measure is defined by the square of the Fubini–Study distance, we can asymptotically calculate the maximum possible error of the optimal measurement as:

$$\mathcal{D}^{d_{fs}^2,(n)}(\Pi_n) n \cong (k-1) - \frac{2}{3}k(k-1)\frac{1}{n} + k(k-1)\frac{23k-7}{45}\frac{1}{n^2} \qquad \text{as } n \to \infty. \tag{10}$$

The error of the sequence $\{\Pi_n\}_{n=1}^{\infty}$ of the optimal measurements can be calculated in the sense of large deviation as:

$$\begin{aligned} &\frac{1}{n}\log(\mathrm{Pr}_{\Pi_n}^{\rho^{(n)}}\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | d_{fs}(\rho, \hat{\rho}) \geqslant \epsilon\}) \\ &\quad \cong \log\cos^2\epsilon + (k-2)\frac{\log n}{n} + (-\log(k-2)! + 2(k-2)\log(\sin\epsilon) \\ &\qquad -2\log(\cos\epsilon))\frac{1}{n} + \left(\frac{k^2-k-2}{2} + (k-2)\cot^2\epsilon\right)\frac{1}{n^2} \qquad \text{as } n \to \infty \end{aligned} \tag{11}$$

where $\mathrm{Pr}_M^S B$ denotes the probability of B with respect to the probability measure $\mathrm{tr}(M(\mathrm{d}\omega)S)$ for a Borel $B \subset \Omega$, a measurement $M \in \mathcal{M}(\Omega, \mathcal{H}')$ and a state $S \in \mathcal{S}(\mathcal{H}')$.

For a proof, see appendix C. The first term on the right-hand side of (11) coincides with the logarithm of the fidelity [22].

In this paper, ϵ in equations (11) is called the admissible radius.

Since

$$\lim_{\epsilon\to 0}\frac{\log\cos^2\epsilon}{\epsilon^2}=-1 \tag{12}$$

we obtain the following large deviation approximation

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{\epsilon^2 n}\log(\mathrm{Pr}^{\rho^{(n)}}_{\Pi_n}\{\hat\rho\in\mathcal{P}(\mathcal{H})|d_{fs}(\rho,\hat\rho)\geqslant\epsilon\})=-1. \tag{13}$$

5. Semiclassical measurement

In this section, we consider measurements which allow the quantum correlation between finite samples only. A measurement M on $\mathcal{H}^{(nm)}$ is called *m-semiclassical* if there exists an estimator T on the probability space $\underbrace{\mathcal{P}(\mathcal{H})\times\cdots\times\mathcal{P}(\mathcal{H})}_{n}$ whose domain is $\mathcal{P}(\mathcal{H})$ such that

$$M(B)=\int_{T^{-1}(B)}\underbrace{\Pi_m(\mathrm{d}\rho_1)\otimes\cdots\otimes\Pi_m(\mathrm{d}\rho_n)}_{n}\qquad\forall B\subset\mathcal{P}(\mathcal{H}). \tag{14}$$

We compare the error between m-semiclassical measurements and the optimal measurement Π_{nm} for nm samples of the unknown state as equations (10), (11) and (13).

In doing this comparison, we bear in mind asymptotic estimation theory in classical statics. In classical statics, it is assumed that the sequence of estimators satisfies the consistency.

Definition 4. A sequence $\{T^{(n)}\}_{n=1}^{\infty}$ of estimators on a probability space Ω is said to be consistent with respect to a family $\{p_\theta|\theta\in\Theta\}$ of probability distributions on Ω, if it satisfies the condition (15), where every $T^{(n)}$ is a probability variable on the probability space $\underbrace{\Omega\times\cdots\times\Omega}_{n}$ whose domain is Θ

$$p_\theta^{(n)}\{d_J(T^{(n)},\theta)>\epsilon\}\to 0\qquad\text{as } n\to\infty\qquad\forall\theta\in\Theta,\forall\epsilon>0 \tag{15}$$

where d_J denotes the geodesic distance defined by the Fisher information metric and $p_\theta^{(n)}$ denotes the probability measure $\underbrace{p_\theta\times\cdots\times p_\theta}_{n}$ on the probability space $\underbrace{\Omega\times\cdots\times\Omega}_{n}$.

It is well known that the following theorem establishes under the preceding consistency [23, 24].

Theorem 5. If a sequence $\{T^{(n)}\}_{n=1}^{\infty}$ of estimators is a consistent estimator with respect to a family $\{p_\theta|\theta\in\Theta\}$ of probability distributions on a probability space Ω which satisfies some regularity, we then have the following inequalities

$$\lim_{n\to\infty} n\underbrace{\int\int\dots\int}_{n}\mathrm{d}_J^2(T^{(n)}(x_1,x_2,\dots,x_n),\theta)\,p_\theta^{(n)}(\mathrm{d}x_1,\dots,\mathrm{d}x_n)\geqslant\dim\Theta \tag{16}$$

$$\lim_{n\to\infty}\frac{1}{n}\log(p_\theta^{(n)}\{D(p_{T^{(n)}}\|p_\theta)\geqslant\epsilon\})\geqslant-\epsilon \tag{17}$$

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{\epsilon^2 n}\log(p_\theta^{(n)}\{d_J(T^{(n)},\theta)\geqslant\epsilon\})\geqslant-\tfrac{1}{2} \tag{18}$$

where $D(p\|q)$ denotes the information divergence of a probability distribution q with respect to another probability distribution p defined by:

$$D(p\|q) := \int_{\Omega} (\log p(\omega) - \log q(\omega)) p(\omega) \mathrm{d}\omega.$$

The lower bounds of (16) and (18) can be attained by the maximum likelihood estimator. The lower bound of (17) can be attained by the maximum likelihood estimator when the family $\{p_\theta | \theta \in \Theta\}$ is exponential, but generally cannot be attained.

For the comparison, we apply theorem 5 to the family of distributions $\{\operatorname{tr} \Pi_m(\mathrm{d}\hat{\rho})\rho^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\}$ given by the measurement Π_m and the family of states $\{\rho^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\}$. We consider the sequence of measurements $\{T_{(n,m)}\}_{n=1}^{\infty}$ which corresponds to the consistent estimator $\{T^{(n)}\}_{n=1}^{\infty}$ on the family of distributions $\{\operatorname{tr} \Pi_m(\mathrm{d}\hat{\rho})\rho^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\}$, where $T_{(n,m)}$ is the measurement on $\mathcal{H}^{(nm)}$ defined by the estimator $T^{(n)}$ and n data given by the measurement $\underbrace{\Pi_m \otimes \cdots \otimes \Pi_m}_{n}$ and the state $\rho^{(nm)}$. From the symmetry of $\mathcal{P}(\mathcal{H})$ and Π_m, the information divergence of a probability measure $\operatorname{tr} \Pi_m(\mathrm{d}\hat{\rho})\rho_1^{(m)}$ with respect to another a probability measure $\operatorname{tr} \Pi_m(\mathrm{d}\hat{\rho})\rho_2^{(m)}$ is determined by the Fubini–Study distance ϵ between ρ_1 and ρ_2. Thus, the divergence may be denoted by $D_m(\epsilon)$. From lemma 2, the geodesic distance d_{Π_m} with respect to Fisher information metric in the family of distributions $\{\operatorname{tr} \Pi_m(\mathrm{d}\hat{\rho})\rho^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\}$ is given by:

$$d_{\Pi_m} = \sqrt{2m} d_{fs}.$$

Since $\dim \mathcal{P}(\mathcal{H}) = 2(k-1)$, we have the following inequalities:

$$\lim_{n\to\infty} nm \mathcal{D}^{d_{fs}^2,(nm)}(T_{(n,m)}) = \lim_{n\to\infty} \max_{\rho\in\mathcal{P}(\mathcal{H})} nm \int_{\mathcal{P}(\mathcal{H})} d_{fs}^2(\rho,\hat{\rho}) \operatorname{tr}(T_{(n,m)}(\mathrm{d}\hat{\rho})\rho^{(nm)}) \geqslant k-1 \tag{19}$$

$$\lim_{n\to\infty} \frac{1}{nm} \log \mathrm{Pr}_{T_{(n,m)}}^{\rho^{(nm)}} \{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | d_{fs}(\rho,\hat{\rho}) \geqslant \epsilon\} \geqslant -\frac{D_m(\epsilon)}{m} \tag{20}$$

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{\epsilon^2 nm} \log \mathrm{Pr}_{T_{(n,m)}}^{\rho^{(nm)}} \{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | d_{fs}(\rho,\hat{\rho}) \geqslant \epsilon\} \geqslant -1. \tag{21}$$

The lower bound of (19) is consistent with the first term on the right-hand side of (10) and the lower bound of (21) is consistent with the right-hand side of (13). When the sequence of measurements $\{T^{(n)}\}_{n=1}^{\infty}$ corresponds to the maximum likelihood estimator, the lower bounds of (19) and (21) can be attained. We have the following lemma concerning the comparison of the lower bound $-\frac{D_m(\epsilon)}{m}$ of (20) and the first term $2\log\cos\epsilon$ on the right-hand side of (11).

Lemma 2. We can calculate the divergence $D_m(\epsilon)$ and the distance d_{Π_m} as:

$$\frac{D_m(\epsilon)}{m} = \sum_{i=1}^{m} \frac{\sin^{2i}\epsilon}{i} \to -\log(1-\sin^2\epsilon) = -\log\cos^2\epsilon \qquad \text{as } m \to \infty \tag{22}$$

$$d_{\Pi_m} = \sqrt{2m} d_{fs}. \tag{23}$$

Therefore, $\frac{D_m(\epsilon)}{m}$ is monotone increasing with respect to m.

For a proof, see appendix D. Equation (22) infers that

$$0 < \frac{-m\log\cos^2\epsilon - D_m(\epsilon)}{m\epsilon^{2m}} \to 0 \qquad \text{as } \epsilon \to 0 \tag{24}$$

which means that the first term of (11) cannot be completely attained by a semiclassical measurement. However, the first term of (10) and the left-hand side of (13) can be asymptotically attained by a 1-semiclassical measurement, i.e. they can be asymptotically attained by measurements without using quantum correlations between samples. Thus, in order to attain them asymptotically, it is sufficient to physically realize the optimal measurement Π_1 on a single sample. Indeed, Π_1 is a random measurement as follows. To denote Π_1 as a random measurement, we will define the simple measurement $E_g (g \in \mathrm{SU}(k))$ whose measurable space $\mathcal{P}(\mathcal{H})$. For an element $g \in \mathrm{SU}(k)$, the vectors $\phi_1(g), \ldots, \phi_k(g)$ in $\mathcal{H}$ are defined as:

$$(\phi_1(g) \ldots \phi_k(g)) = g.$$

The measurement E_g is defined as:

$$E_g(|\phi_i(g)\rangle\langle\phi_i(g)|) = |\phi_i(g)\rangle\langle\phi_i(g)|.$$

Therefore, the optimal measurement Π_1 for a single sample can be described as the following random measurement:

$$\Pi_1 = \int_{\mathrm{SU}(k)} E_g \mu(\mathrm{d}g) \tag{25}$$

where μ is the invariant measure on $\mathrm{SU}(k)$ with $\mu(\mathrm{SU}(k)) = 1$. Therefore, in order to realize the optimal measurement Π_1, it is sufficient to realize the simple measurement E_g for any $g \in \mathrm{SU}(k)$.

6. Conclusion

We have compared two cases. One regards the system consisting of enough samples as the single system, the other regards it as separate systems. Under this comparison, the error mean squares of both cases asymptotically coincide in the first order with respect to the Fubini–Study distance (see (10) and (19)). However, we leave the question of whether they asymptotically coincide in the second order with respect to the Fubini–Study distance to a future study. On the other hand, in view of the evaluation of large deviation, if the allowable radius is finite, neither coincide (see (11) and (20)). However, if the allowable radius goes to infinitesimal, both coincide (see (13) and (21)).

These results depend on the effect of a pure state. Therefore, it is an open question as to whether the error mean squares of both cases asymptotically coincide in the first order in another family. In the case of large deviations, the same question is also open in the limit where the radius of allowable error goes to infinitesimal.

Appendix A. Proof of theorem 3

In this appendix, assume that $\rho = |\phi(0)\rangle\langle\phi(0)|$. Because $\mathcal{H}_s^{(n)}$ is irreducible with respect to the action of $\mathrm{SU}(k)$,

$$\begin{aligned}\mathcal{H}_s^{(n)} &= \left\{ \sum_i a_i V_{g_i} \phi(0)^{(n)} | a_i \in \mathbb{C}, g_i \in \mathrm{SU}(k) \right\} \\ &= \left\{ \sum_i \phi_i^{(n)} | \phi_i \in \mathcal{H} \right\}.\end{aligned} \tag{26}$$

We assume that $W(\rho, \hat{\rho}) = h(\operatorname{tr} \rho\hat{\rho})$. As h is monotone decreasing, there exists a measure h' on $[0, 1]$ such that $h(x) = h'([x, 1])$.

The function h_β on $[0,1]$ and the deviation measure W_β are defined as follows

$$h_\beta(x) := \begin{cases} 1 & \text{for } x \leqslant \beta \\ 0 & \text{for } x > \beta \end{cases}$$

$$W_\beta(\rho, \hat{\rho}) := h_\beta(\operatorname{tr} \rho\hat{\rho}).$$

From lemma 3, for any measurement Π we have

$$\mathcal{D}_\rho^{W,(n)}(\Pi) = \int_{[0,1]} \mathcal{D}_\rho^{W_\beta,(n)}(\Pi) h'(\mathrm{d}\beta).$$

From (26), it is sufficient to show the following for $\{\phi_i\} \subset \mathcal{H}$ in the case of $W = W_\beta$.

$$\frac{\operatorname{tr} \hat{W}_\beta(\rho) | \sum_i \phi_i^{(n)}\rangle\langle \sum_i \phi_i^{(n)}|}{\langle \sum_i \phi_i^{(n)} | \sum_i \phi_i^{(n)}\rangle} \geqslant \operatorname{tr} \hat{W}_\beta(\rho) |\phi(0)^{(n)}\rangle\langle\phi(0)^{(n)}|. \tag{27}$$

From lemma 4 it is sufficient for (27) to prove the following

$$\left\langle \sum_i \phi_i^{(n)} \middle| \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle \cdot \langle \phi(0)^{(n)} | \operatorname{Id} - \hat{W}_\beta(\rho) | \phi(0)^{(n)}\rangle$$
$$\geqslant \langle \phi(0)^{(n)} \middle| \hat{W}_\beta(\rho) \middle| \phi(0)^{(n)}\rangle \cdot \left\langle \sum_i \phi_i^{(n)} \middle| \operatorname{Id} - \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle. \tag{28}$$

Remark that $|\langle \phi(\theta)|\phi(0)\rangle|^2 = \cos^2\theta_1$. From lemma 5, we obtain

$$\left\langle \sum_i \phi_i^{(n)} \middle| \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle = \frac{k' \cdot (k-1)!}{\pi^{(k-1)}} \int_\alpha^{\frac{\pi}{2}} f_1(\theta_1) \cos\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1$$

$$\left\langle \sum_i \phi_i^{(n)} \middle| \operatorname{Id} - \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle = \frac{k' \cdot (k-1)!}{\pi^{(k-1)}} \int_0^{\alpha} f_1(\theta_1) \cos\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1$$

$$\langle \phi(0)^{(n)} | \hat{W}_\beta(\rho) | \phi(0)^{(n)}\rangle = C \int_\alpha^{\frac{\pi}{2}} \cos^{2n+1}\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1$$

$$\langle \phi(0)^{(n)} | \operatorname{Id} - \hat{W}_\beta(\rho) | \phi(0)^{(n)}\rangle = C \int_0^{\alpha} \cos^{2n+1}\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1$$

where

$$\beta := \cos^2\alpha$$

$$f_1(\theta_1) := \underbrace{\int_0^{\frac{\pi}{2}} \cdots \int_0^{\frac{\pi}{2}}}_{k-2} f_2(\theta_1, \ldots, \theta_{k-1}) \lambda(\mathrm{d}\theta_2 \ldots \mathrm{d}\theta_{k-1})$$

$$f_2(\theta_1, \ldots, \theta_{k-1}) := \underbrace{\int_0^{2\pi} \cdots \int_0^{2\pi}}_{k-1} \sum_{i,j} \langle \phi_i | \phi(\theta)\rangle^n \langle \phi(\theta) | \phi_j\rangle^n \,\mathrm{d}\theta_k \ldots \mathrm{d}\theta_{2k-2}$$

$$C := \frac{k' \cdot (k-1)!}{\pi^{(k-1)}} \underbrace{\int_0^{2\pi} \cdots \int_0^{2\pi}}_{k-1} \underbrace{\int_0^{\frac{\pi}{2}} \cdots \int_0^{\frac{\pi}{2}}}_{k-2} \lambda(\mathrm{d}\theta_2\,\mathrm{d}\theta_3 \ldots \mathrm{d}\theta_{k-1}) \mathrm{d}\theta_k \ldots \mathrm{d}\theta_{2k-2}$$

$$\lambda(\mathrm{d}\theta_2, \ldots, \mathrm{d}\theta_{k-1}) := \sin^{2k-5}\theta_2 \ldots \sin\theta_{k-1} \cos\theta_2 \ldots \cos\theta_{k-1} \,\mathrm{d}\theta_2 \ldots \mathrm{d}\theta_{k-1}.$$

Therefore, it is sufficient for equation (28) to show that for $\pi/2 \geqslant \theta_1 > \theta_1' \geqslant 0$

$$f_1(\theta_1) \sin^{2k-3}\theta_1 \cos^{2n+1}\theta_1' \sin^{2k-3}\theta_1' \geqslant f_1(\theta_1') \sin^{2k-3}\theta_1' \cos^{2n+1}\theta_1 \sin^{2k-3}\theta_1.$$

It suffices to verify that for $\theta_i \in [0, \frac{\pi}{2}]$, $2 \leqslant i \leqslant k-1$, $\pi/2 \geqslant \theta_1 > \theta_1' \geqslant 0$

$$\frac{f_2(\theta_1, \theta_2, \ldots, \theta_{n-1})}{\cos^{2n}\theta_1} \geqslant \frac{f_2(\theta_1', \theta_2, \ldots, \theta_{n-1})}{\cos^{2n}\theta_1'}.$$

Thus, it is sufficient to prove that the following is monotone decreasing about θ_1 for any $\theta_2, \ldots, \theta_{k-1}$:

$$\frac{1}{\cos^{2n}\theta_1} \underbrace{\int_0^{2\pi} \cdots \int_0^{2\pi}}_{k-1} \sum_{i,j} \langle\phi_i|\phi(\theta)\rangle^n \langle\phi(\theta)|\phi_j\rangle^n \,\mathrm{d}\theta_k \ldots \mathrm{d}\theta_{2k-2}. \tag{29}$$

Letting

$$\phi_i := \begin{pmatrix} \mathrm{e}^{\mathrm{i}\psi_i^1}\phi_i^1 \\ \mathrm{e}^{\mathrm{i}\psi_i^2}\phi_i^2 \\ \vdots \\ \mathrm{e}^{\mathrm{i}\psi_i^k}\phi_i^k \end{pmatrix}$$

we find

$$\frac{\langle\phi_i|\phi(\theta)\rangle^n}{\cos^n\theta_1} = \Bigg(\mathrm{e}^{\mathrm{i}\psi_i^1}\phi_i^1 + \sum_{j=2}^{k-1} \mathrm{e}^{\mathrm{i}(\theta_{k-2+j}-\psi_i^j)} \tan\theta_1 \sin\theta_2 \ldots \sin\theta_j \cos\theta_{j+1}\phi_i^j$$
$$+\mathrm{e}^{\mathrm{i}(\theta_{2k-2}-\psi_i^{k-1})} \tan\theta_1 \sin\theta_2 \cdots \sin\theta_{k-1}\phi_i^k\Bigg)^n.$$

Letting $x := \tan\theta_1$, lemma 6 infers that (29) is monotone decreasing about θ_1. The proof is complete.

Lemma 3. If the deviation measure $W(\rho, \hat\rho) = h'([\mathrm{tr}\,\rho\hat\rho, 1])$, then

$$\mathcal{D}_\rho^{W,(n)}(\Pi) = \int_{[0,1]} \mathcal{D}_\rho^{W_\beta,(n)}(\Pi) h'(\mathrm{d}\beta). \tag{30}$$

Proof. For the probability measure π on $\mathcal{P}(\mathcal{H})$, we have

$$\begin{aligned}\int_{\mathcal{P}(\mathcal{H})} W(\rho, \hat\rho)\pi(\mathrm{d}\hat\rho) &= \int_{\mathcal{P}(\mathcal{H})} h(\mathrm{tr}\,\rho\hat\rho)\pi(\mathrm{d}\hat\rho) \\ &= \int_{\mathcal{P}(\mathcal{H})}\int_{[0,1]} h_\beta(\mathrm{tr}\,\rho\hat\rho)h'(\mathrm{d}\beta)\pi(\mathrm{d}\hat\rho) \\ &= \int_{[0,1]}\left(\int_{\mathcal{P}(\mathcal{H})} h_\beta(\mathrm{tr}\,\rho\hat\rho)\pi(\mathrm{d}\hat\rho)\right)h'(\mathrm{d}\beta) \\ &= \int_{[0,1]}\left(\int_{\mathcal{P}(\mathcal{H})} W_\beta(\rho, \hat\rho)\pi(\mathrm{d}\hat\rho)\right)h'(\mathrm{d}\beta).\end{aligned}$$

Substituting $\pi(\mathrm{d}\hat\rho)$ for $\mathrm{tr}(\Pi(\mathrm{d}\hat\rho)\rho^{(n)})$, we obtain (30). □

Lemma 4. Let $\mathcal{H}$ be any finite-dimensional Hilbert space. For any elements $\phi, \psi \in \mathcal{H}$ and any self-adjoint operator A on $\mathcal{H}$, the following are equivalent

- $\frac{\langle\phi|A|\phi\rangle}{\langle\phi|\phi\rangle} \geqslant \frac{\langle\psi|A|\psi\rangle}{\langle\psi|\psi\rangle}$
- $\langle\phi|A|\phi\rangle\langle\psi|\,\mathrm{Id}-A|\psi\rangle \geqslant \langle\psi|A|\psi\rangle\langle\phi|\,\mathrm{Id}-A|\phi\rangle$.

Lemma 5. We have

$$\left\langle \sum_i \phi_i^{(n)} \middle| \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle = \frac{k' \cdot (k-1)!}{\pi^{(k-1)}} \int_\alpha^{\frac{\pi}{2}} f_1(\theta_1) \cos\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1$$

$$\left\langle \sum_i \phi_i^{(n)} \middle| \mathrm{Id} - \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle = \frac{k' \cdot (k-1)!}{\pi^{(k-1)}} \int_0^{\alpha} f_1(\theta_1) \cos\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1$$

$$\langle \phi(0)^{(n)} | \hat{W}_\beta(\rho) | \phi(0)^{(n)} \rangle = C \int_\alpha^{\frac{\pi}{2}} \cos^{2n+1}\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1$$

$$\langle \phi(0)^{(n)} | \mathrm{Id} - \hat{W}_\beta(\rho) | \phi(0)^{(n)} \rangle = C \int_0^{\alpha} \cos^{2n+1}\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1.$$

Proof. $\hat{W}_\beta(\rho)$ is denoted as follows

$$\begin{aligned} \hat{W}_\beta(\rho) &= k' \int_{\mathcal{P}(\mathcal{H})} W_\beta(\rho, \hat{\rho}) \hat{\rho}^{(n)} \nu(\mathrm{d}\hat{\rho}) \\ &= k' \int_{\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | \operatorname{tr} \hat{\rho}\rho \leqslant \beta\}} \hat{\rho}^{(n)} \nu(\mathrm{d}\hat{\rho}). \end{aligned}$$

We obtain

$$\begin{aligned} \left\langle \sum_i \phi_i^{(n)} \middle| \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle &= \left\langle \sum_i \phi_i^{(n)} \middle| k' \int_{\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | \operatorname{tr} \hat{\rho}\rho \leqslant \beta\}} \hat{\rho}^{(n)} \nu(\mathrm{d}\hat{\rho}) \middle| \sum_i \phi_i^{(n)} \right\rangle \\ &= \sum_{i,j} k' \int_{\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | \operatorname{tr} \hat{\rho}\rho \leqslant \beta\}} \langle \phi_i^{(n)} | \hat{\rho}^{(n)} | \phi_j^{(n)} \rangle \nu(\mathrm{d}\hat{\rho}) \\ &= \sum_{i,j} k' \int_{\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | \operatorname{tr} \hat{\rho}\rho \leqslant \beta\}} \langle \phi_i | \hat{\rho} | \phi_j \rangle^n \nu(\mathrm{d}\hat{\rho}) \\ &= \frac{k' \cdot (k-1)!}{\pi^{(k-1)}} \int_\alpha^{\frac{\pi}{2}} f_1(\theta_1) \cos\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1. \end{aligned}$$

Similarly,

$$\begin{aligned} \left\langle \sum_i \phi_i^{(n)} \middle| \mathrm{Id} - \hat{W}_\beta(\rho) \middle| \sum_i \phi_i^{(n)} \right\rangle &= \sum_{i,j} k' \int_{\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | \operatorname{tr} \hat{\rho}\rho > \beta\}} \langle \phi_i | \hat{\rho} | \phi_j \rangle^n \nu(\mathrm{d}\hat{\rho}) \\ &= \frac{k' \cdot (k-1)!}{\pi^{(k-1)}} \int_0^{\alpha} f_1(\theta_1) \cos\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1 \end{aligned}$$

$$\begin{aligned} \langle \phi(0)^{(n)} | \hat{W}_\beta(\rho) | \phi(0)^{(n)} \rangle &= k' \int_{\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | \operatorname{tr} \hat{\rho}\rho \leqslant \beta\}} \langle \phi(0) | \hat{\rho} | \phi(0) \rangle^n \nu(\mathrm{d}\hat{\rho}) \\ &= C \int_\alpha^{\frac{\pi}{2}} \cos^{2n+1}\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1 \end{aligned}$$

$$\begin{aligned} \langle \phi(0)^{(n)} | \mathrm{Id} - \hat{W}_\beta(\rho) | \phi(0)^{(n)} \rangle &= k' \int_{\{\hat{\rho} \in \mathcal{P}(\mathcal{H}) | \operatorname{tr} \hat{\rho}\rho > \beta\}} \langle \phi(0) | \hat{\rho} | \phi(0) \rangle^n \nu(\mathrm{d}\hat{\rho}) \\ &= C \int_0^{\alpha} \cos^{2n+1}\theta_1 \sin^{2k-3}\theta_1 \, \mathrm{d}\theta_1. \end{aligned}$$

□

Lemma 6. The following function $f(x)$ is monotone decreasing on $[0, \infty)$:

$$f(x) := \sum_{a=1}^{m} \sum_{b=1}^{m} \underbrace{\int_0^{2\pi} \dots \int_0^{2\pi}}_{k} \left(c_a^0 \mathrm{e}^{\mathrm{i} d_a^0} + x \sum_{j=1}^{k} \mathrm{e}^{\mathrm{i}(\theta_j + d_a^j)} c_a^j \right)^n \times \left(c_b^0 \mathrm{e}^{\mathrm{i} d_b^0} + x \sum_{j=1}^{k} \mathrm{e}^{-\mathrm{i}(\theta_j + d_b^j)} c_b^j \right)^n \mathrm{d}\theta_1 \dots \mathrm{d}\theta_k$$

where c_n^j, d_n^j are any real numbers.

Proof. The set K_n^m is defined as follows

$$K_n^m := \left\{ I = (I_1, \cdots, I_m) \in (\mathbb{N}^{+,0})^m \,\middle|\, \sum_{j=1}^{m} I_j = n. \right\}.$$

The number $C(I)$ is defined for $I \in K_n^m$ as sufficing the following condition:

$$\left(\sum_{j=1}^{m} x_j \right)^n = \sum_{I \in K_n^m} C(I) x_1^{I_0} \dots x_m^{I_m}.$$

Therefore,

$$\left(c_a^0 + x \sum_{j=1}^{k} \mathrm{e}^{\mathrm{i}(\theta_j + d_a^j)} c_a^j \right)^n = \sum_{I \in K_n^{k+1}} C(I) \mathrm{e}^{\mathrm{i} d_a^0} (c_a^0)^{I_0} \mathrm{e}^{\mathrm{i} I_1 (\theta_1 + d_a^1)} (c_a^1)^{I_1} \dots \mathrm{e}^{\mathrm{i} I_d (\theta_k + d_a^k)} (c_a^k)^{I_k} x^{n - I_0}.$$

Thus,

$$\begin{aligned} f(x) = \sum_{a=1}^{m} \sum_{b=1}^{m} (2\pi)^k \sum_{I} C(I) \mathrm{e}^{\mathrm{i} I_0 (d_a^0 - d_b^0)} (c_a^0 c_b^0)^{I_0} \mathrm{e}^{\mathrm{i} I_1 (d_a^1 - d_b^1)} \\ \times (c_a^1 c_b^1)^{I_1} \dots \mathrm{e}^{\mathrm{i} I_k (d_a^k - d_b^k)} (c_a^k c_b^k)^{I_k} x^{2n - 2I_0} \\ = (2\pi)^k \sum_{I} C(I) \sum_{a=1}^{m} \sum_{b=1}^{m} \mathrm{e}^{\mathrm{i}(\sum_{j=0}^{k} I_i d_a^i - \sum_{j=0}^{k} I_i d_b^i)} \\ \times (c_a^0)^{I_0} \dots (c_a^k)^{I_k} (c_b^0)^{I_0} \dots (c_b^k)^{I_k} x^{2n - 2I_0} \\ = (2\pi)^k \sum_{I} C(I) D(I) x^{2n - 2I_0} \end{aligned}$$

where

$$D(I) := \sum_{a=1}^{m} \sum_{b=1}^{m} \mathrm{e}^{\mathrm{i}(\sum_{j=0}^{k} I_i d_a^i - \sum_{j=0}^{k} I_i d_b^i)} (c_a^0)^{I_0} \dots (c_a^k)^{I_k} (c_b^0)^{I_0} \dots (c_b^k)^{I_k}.$$

It is sufficient to show $D(I) \geqslant 0$. Letting

$$v_a := (c_a^0)^{I_0} \dots (c_a^k)^{I_k}$$

$$y_a := \sum_{j=0}^{k} I_i d_a^i$$

$$w_{a,b} := \cos(y_a - y_b)$$

we have

$$D(I) = \sum_{a=1}^{m} \sum_{b=1}^{m} v_a w_{a,b} v_b.$$

Then

$$w_{a,b} = \cos(y_a - y_b) = \cos y_a \cos y_b + \sin y_a \sin y_b.$$

As $\{\cos y_a \cos y_b\}$ and $\{\sin y_a \sin y_b\}$ are nonnegative, $\{w_{a,b}\}$ is nonnegative matrix. Therefore, we obtain $D(I) \geqslant 0$. □

Appendix B. Proof of lemma 1

$$\begin{aligned}
\mathcal{D}^{W,(n)}(\Pi_n) &= \int_{\mathcal{P}(\mathcal{H})} h(d_{fs}(\rho,\hat{\rho})\,\mathrm{tr}(\Pi_n(\mathrm{d}\hat{\rho})\rho^{(n)}) \\
&= \int_{\mathcal{P}(\mathcal{H})} h(\theta_1)\binom{n+k-1}{k-1}|\langle\phi(\theta)^{(n)}|\phi(0)^{(n)}\rangle|^2\nu(\mathrm{d}\theta) \\
&= \int_0^{\frac{\pi}{2}} h(\theta_1)\binom{n+k-1}{k-1}\frac{(k-1)!}{\pi^{k-1}}\cos^{2n+1}\theta_1\sin^{2k-3}\theta_1\,\mathrm{d}\theta_1 \\
&\quad\times \underbrace{\int_0^{2\pi}\cdots\int_0^{2\pi}}_{k-1}\underbrace{\int_0^{\frac{\pi}{2}}\cdots\int_0^{\frac{\pi}{2}}}_{k-2}\sin^{2k-5}\theta_2\cdots\sin\theta_{k-1} \\
&\quad\times\cos\theta_2\ldots\cos\theta_{k-1}\,\mathrm{d}\theta_2\ldots\mathrm{d}\theta_{2k-2} \\
&= \int_0^{\frac{\pi}{2}} h(\theta_1)\binom{n+k-1}{k-1}\frac{(k-1)!}{\pi^{k-1}}\cos^{2n+1}\theta_1\sin^{2k-3}\theta_1\,\mathrm{d}\theta_1 \\
&\quad\times\underbrace{\int_0^1 x^{2k-5}\,\mathrm{d}x\ldots\int_0^1 x\,\mathrm{d}x}_{k-2}\cdot(2\pi)^{k-1} \\
&= \int_0^{\frac{\pi}{2}} h(\theta)\binom{n+k-1}{k-1}\frac{(k-1)!}{\pi^{k-1}}\cos^{2n+1}\theta\sin^{2k-3}\theta\,\mathrm{d}\theta\frac{(2\pi)^{k-1}}{2^{k-2}(k-2)!} \\
&= 2(k-1)\binom{n+k-1}{k-1}\int_0^{\frac{\pi}{2}} h(\theta)\cos^{2n+1}\theta\sin^{2k-3}\theta\,\mathrm{d}\theta.
\end{aligned}$$

The proof is complete.

Appendix C. Proof of theorem 4

Definition 1 and lemma 1 infer that:

$$\mathcal{D}^{d_b^{\gamma},(n)}(\Pi_n) = 2(k-1)\binom{n+k-1}{k-1}\int_0^{\frac{\pi}{2}}\cos^{2n+1}\theta\sin^{2k-3+\gamma}\theta\,\mathrm{d}\theta. \tag{31}$$

Since

$$\int_0^{\frac{\pi}{2}}\cos^x\theta\sin^y\theta\,\mathrm{d}\theta = \frac{\Gamma(\frac{x+1}{2})\Gamma(\frac{y+1}{2})}{2\Gamma(\frac{x+y}{2}+1)} \qquad \forall x, y \in \mathbb{R}$$

we have

$$\begin{aligned}
\mathcal{D}^{d_b^{\gamma},(n)}(\Pi_n) &= 2(k-1)\binom{n+k-1}{k-1}\frac{\Gamma(n+1)\Gamma(k-1+\gamma/2)}{\Gamma(n+k+\gamma/2)} \\
&= \frac{\Gamma(n+1)\Gamma(k-1+\gamma/2)\Gamma(n+k)}{\Gamma(n+k+\gamma/2)\Gamma(n+1)\Gamma(k-1)}
\end{aligned}$$

$$= \frac{\Gamma(n+k)}{\Gamma(n+k+\gamma/2)} \frac{\Gamma(k-1+\gamma/2)}{\Gamma(k-1)}. \tag{32}$$

Therefore we obtain (8) from the following formula of Γ function:

$$\lim_{n\to\infty} \frac{\Gamma(n+x)}{\Gamma(n)n^x} = 1.$$

Letting $\gamma := 2$, we obtain

$$\mathcal{D}^{d_b^2,(n)}(\Pi_n) = \frac{\Gamma(n+k)}{\Gamma(n+k+1)} \frac{\Gamma(k-1+1)}{\Gamma(k-1)} = \frac{k-1}{n+k}.$$

Thus, we obtain (9).

Next, we will prove (10). θ^2 can be expanded as:

$$\theta^2 = \sum_{i=0}^{\infty} \frac{(2i-2)!!}{(2i-1)!!} \frac{\sin^{2i}\theta}{i}$$

where we put $(2n)!! = 2n(2n-2)\ldots 4\cdot 2$, $(2n-1)!! = (2n-1)(2n-3)\ldots 3\cdot 1$, $0!! = (-1)!! = 1$. From (32) we have

$$\begin{aligned}\mathcal{D}^{d_{fs}^2,(n)}(\Pi_n) &= \sum_{i=0}^{\infty} \frac{(2i-2)!!}{(2i-1)!!i} \mathcal{D}^{d_b^{2i},(n)}(\Pi_n)\\ &= \sum_{i=0}^{\infty} \frac{(2i-2)!!}{(2i-1)!!i} \prod_{j=0}^{i-1} \frac{k-1+j}{n+k+j}\\ &\cong (k-1)\frac{1}{n} - \frac{2}{3}k(k-1)\frac{1}{n^2} + k(k-1)\frac{23k-7}{45}\frac{1}{n^3}.\end{aligned}$$

Thus we obtain (10). Lemma 1 infers that

$$\begin{aligned}&\log \mathrm{Pr}_{\Pi_n}^{\rho^{(n)}}\{\hat\rho \in \mathcal{P}(\mathcal{H}) | d_{fs}(\rho,\hat\rho) \geqslant \epsilon\}\\ &\quad = \log\left((k-1)\binom{n+k-1}{k-1} \int_0^{\cos^2\epsilon} x(1-x)^{k-2}\,\mathrm{d}x\right).\end{aligned} \tag{33}$$

Therefore, it is sufficient for (13) to show that

$$\log\binom{n+k-1}{k-1} \cong (k-1)\log n - \log(k-1)! + \frac{1}{n}\frac{(k-1)k}{2} \tag{34}$$

$$\begin{aligned}\log\left(\int_0^{\cos^2\epsilon} x(1-x)^{k-2}\,\mathrm{d}x\right) &\cong 2n\log\cos\epsilon - \log n + 2(k-2)\log\sin\epsilon - 2\log\cos\epsilon\\ &\quad -\frac{1}{n}(1+(k-2)\cot^2\epsilon).\end{aligned} \tag{35}$$

The left-hand side of (34) is calculated as:

$$\begin{aligned}\log\binom{n+k-1}{k-1} &= \sum_{i=0}^{k-1} \log\frac{n+i}{i}\\ &= (k-1)\log n - \log(k-1)! + \sum_{i=1}^{k-1}\log\left(1+\frac{i}{n}\right)\\ &\cong (k-1)\log n - \log(k-1)! + \sum_{i=1}^{k-1}\frac{i}{n}\\ &= (k-1)\log n - \log(k-1)! + \frac{1}{n}\frac{(k-1)k}{2}.\end{aligned}$$

Therefore, we have (34). The left-hand side of (35) is calculated as:

$$\log\left(\int_0^{\cos^2\epsilon} x(1-x)^{k-2}\,\mathrm{d}x - 2n\log\cos\epsilon\right) = \log\left(\int_0^{\cos^2\epsilon}\left(\frac{x}{\cos^2\epsilon}\right)^n(1-x)^{k-2}\,\mathrm{d}x\right)$$
$$= \log\left(\int_0^1 x^n(1-\cos^2 x)^{k-2}\frac{1}{\cos^2\epsilon}\,\mathrm{d}x\right)$$
$$= -2\log\cos\epsilon + \log\left(\sum_{i=1}^{k-2}\binom{k-2}{i}(-\cos^2\epsilon)^i\frac{1}{n+i+1}\right)$$
$$= -2\log\cos\epsilon - \log n + \log\left(\sum_{i=1}^{k-2}\binom{k-2}{i}(-\cos^2\epsilon)^i\frac{1}{1+\frac{i+1}{n}}\right)$$
$$\cong -2\log\cos\epsilon - \log n + \log\left(\sum_{i=1}^{k-2}\binom{k-2}{i}(-\cos^2\epsilon)^i\left(1-\frac{i+1}{n}\right)\right)$$
$$= -2\log\cos\epsilon - \log n + \log((1-\cos^2\epsilon)^{k-2}$$
$$-\frac{1}{n}(1-\cos^2\epsilon)^{k-3}(1-(k-1)\cos^2\epsilon))$$
$$= -2\log\cos\epsilon - \log n + \log(1-\cos^2\epsilon)^{k-2}$$
$$+\log\left(1-\frac{1}{n}\frac{1-(k-1)\cos^2\epsilon}{1-\cos^2\epsilon}\right)$$
$$\cong -2\log\cos\epsilon - \log n + (k-2)\log(1-\cos^2\epsilon) - \frac{1}{n}\frac{1-(k-1)\cos^2\epsilon}{1-\cos^2\epsilon}.$$

We obtain (35).

Appendix D. Proof of lemma 2

From the symmetry of $\mathcal{P}(\mathcal{H})$ and Π_m, we may assume that $\rho_1 = |\phi_0\rangle\langle\phi_0|$, $\rho_2 = |\phi_\epsilon\rangle\langle\phi_\epsilon|$. First, we consider the case of $k = 2$. For the following calculation, we prepare the following equations:

$$|\langle\phi_\epsilon|\phi(\theta)\rangle|^{2n} = (\cos^2\epsilon\cos^2\theta_1 + \sin^2\theta_1\sin^2\epsilon + 2\cos\epsilon\sin\epsilon\cos\theta_1\sin\theta_1\cos\theta_2)^n \tag{36}$$
$$\int_0^{2\pi}\log(1+2a\cos\theta+a^2)\mathrm{d}\theta = 4\pi\psi(|a|)\log|a| \tag{37}$$

where the function ψ is defined as:

$$\psi(x) = \begin{cases} 1 & x \geqslant 1 \\ 0 & x < 0. \end{cases}$$

Paying attention to (5) and (7), we have

$$\frac{-D_{\Pi_m}(\rho_1^{(m)}\|\rho_2^{(m)}) - m\log\cos^2\epsilon}{m}$$
$$= -\frac{1}{m}\left((m+1)\int_{\mathcal{P}(\mathcal{H})}\log\frac{|\langle\phi_0|\phi(\theta)\rangle|^{2m}}{|\langle\phi_\epsilon|\phi(\theta)\rangle|^{2m}}|\langle\phi_0|\phi(\theta)\rangle|^{2m}\nu(\mathrm{d}\theta) + m\log\cos\epsilon\right)$$
$$= -\frac{2(m+1)}{\pi}$$

$$\times \int_0^{2\pi}\int_0^{\frac{\pi}{2}} \log\left(\frac{\cos^2\theta_1\cos^2\epsilon}{(\cos\theta_1\cos\epsilon+\sin\theta_1\cos\theta_2\sin\epsilon)^2+(\sin\theta_1\sin\theta_2\sin\epsilon)^2}\right)$$
$$\times \cos^{2m+1}\theta_1\sin\theta_1\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2$$
$$= \frac{(m+1)}{\pi}\int_0^{\frac{\pi}{2}}\int_0^{2\pi}\log(1+2\tan\theta_1\tan\epsilon\cos\theta_2+(\tan\theta_1\tan\epsilon)^2)\mathrm{d}\theta_2$$
$$\times\cos^{2m+1}\theta_1\sin\theta_1\,\mathrm{d}\theta_1$$
$$= \frac{(m+1)}{\pi}\int_0^{\frac{\pi}{2}}4\pi\log(\tan\theta_1\tan\epsilon)\psi(\tan\theta_1\tan\epsilon)\cos^{2m+1}\theta_1\sin\theta_1\,\mathrm{d}\theta_1$$
$$= 2(m+1)\int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}}\log(\tan^2\theta_1\tan^2\epsilon)\cos^{2m+1}\theta_1\sin\theta_1\,\mathrm{d}\theta_1$$
$$= (m+1)\int_0^{\sin^2\epsilon}\log\left(\frac{1-x}{x}\tan^2\epsilon\right)x^m\,\mathrm{d}x. \tag{38}$$

Substitute $a=\tan^2\epsilon$ into (51) of lemma 8, then

$$(m+1)\int_0^{\sin^2\epsilon}x^m\log\left(\frac{1-x}{x}\tan^2\epsilon\right)\mathrm{d}x = -\log\cos^2\epsilon-\sum_{i=1}^m\frac{\sin^{2i}\epsilon}{i}. \tag{39}$$

From (38) and (39), we have

$$\frac{D_{\Pi_m}(\rho_1^{(m)}\|\rho_2^{(m)})}{m} = \sum_{i=1}^m\frac{\sin^{2i}\epsilon}{i}. \tag{40}$$

Therefore, we can prove (22).

Next, we consider the case of $k\geqslant 3$. In this case, we have:

$$|\langle\phi_\epsilon|\phi(\theta)\rangle|^{2n} = (\cos^2\epsilon\cos^2\theta_1+\sin^2\theta_1\cos^2\theta_2\sin^2\epsilon$$
$$+2\cos\epsilon\sin\epsilon\cos\theta_1\sin\theta_1\cos\theta_2\cos\theta_k)^n. \tag{41}$$

Paying attention to (5), (37) and lemma 7, we can calculate

$$\frac{-D_{\Pi_m}(\rho_1^{(m)}\|\rho_2^{(m)})-m\log\cos^2\epsilon}{m} = -\frac{1}{m}\left(\binom{m+k-1}{k-1}\int_{\mathcal{P}(\mathcal{H})}\log\left(\frac{|\langle\phi_0|\phi(\theta)\rangle|^{2m}}{|\langle\phi_\epsilon|\phi(\theta)\rangle|^{2m}}\right)\right.$$
$$\left.\times|\langle\phi_0|\phi(\theta)\rangle|^{2m}\nu(\mathrm{d}\theta)+2m\log\cos\epsilon\right)$$
$$= -\frac{2(k-1)(k-2)}{\pi}\binom{m+k-1}{k-1}\int_0^{2\pi}\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}\log(\cos^2\theta_1\cos^2\epsilon)$$
$$\times[(\cos^2\epsilon\cos^2\theta_1+\sin^2\theta_1\cos^2\theta_2\sin^2\epsilon$$
$$+2\cos\epsilon\sin\epsilon\cos\theta_1\sin\theta_1\cos\theta_2\cos\theta_k)]^{-1}$$
$$\times\cos^{2m+1}\theta_1\sin^{2k-3}\theta_1\cos\theta_2\sin^{2k-5}\theta_2\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2\,\mathrm{d}\theta_k$$
$$= \frac{2(k-1)(k-2)}{\pi}\binom{m+k-1}{k-1}$$
$$\times\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}\int_0^{2\pi}\log(1+2\tan\theta_1\cos\theta_2\tan\epsilon\cos\theta_k+(\tan\theta_1\cos\theta_1\tan\epsilon)^2)\,\mathrm{d}\theta_k$$
$$\times\cos^{2m+1}\theta_1\sin^{2k-3}\theta_1\cos\theta_2\sin^{2k-5}\theta_2\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2$$
$$= \frac{2(k-1)(k-2)}{\pi}\binom{m+k-1}{k-1}\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}4\pi\psi(\tan\theta_1\cos\theta_2\tan\epsilon)$$
$$\times\log(\tan\theta_1\cos\theta_2\tan\epsilon)\cos^{2m+1}\theta_1\sin^{2k-3}\theta_1\cos\theta_2\sin^{2k-5}\theta_2\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2. \tag{42}$$

Substituting $a := \tan^2\epsilon$, $s := \sin^2\theta_2$, $y := \cos^2\theta_1$, then the condition $\tan\theta_1\cos\theta_2\tan\epsilon \geqslant 1$ turns into the following conditions:

$$1 - \frac{y}{a(1-y)} \geqslant x \geqslant 0 \qquad \frac{a}{1+a} \geqslant y \geqslant 0.$$

Using (42) and (50), we have

$$\begin{aligned}
\frac{-D_{\Pi_m}(\rho_1^{(m)}\|\rho_2^{(m)}) - m\log\cos^2\epsilon}{m} &= (k-1)(k-2)\binom{m+k-1}{k-1} \\
&\quad\times \int_0^{\frac{a}{1+a}} \left(\int_0^{1-\frac{y}{a(1-y)}} x^{k-3}\log\left((1-x)\frac{a(1-y)}{y}\right)\mathrm{d}x\right) y^m(1-y)^{k-2}\,\mathrm{d}y \\
&= (k-1)\binom{m+k-1}{k-1}\int_0^{\frac{a}{1+a}} \left(-\log\left(\frac{y}{a(1-y)}\right) - \sum_{i=1}^{k-2}\frac{1}{i}\left(\frac{a-(1+a)y}{a(1-y)}\right)^i\right) \\
&\quad\times y^m(1-y)^{k-2}\,\mathrm{d}y \\
&= -(k-1)\binom{m+k-1}{k-1}\int_0^{\frac{a}{1+a}}\log\left(\frac{y}{a(1-y)}\right) y^m(1-y)^{k-2}\,\mathrm{d}y \\
&\quad -(k-1)\binom{m+k-1}{k-1} f\left(\frac{a}{1+a}\right)
\end{aligned} \tag{43}$$

where $f(x)$ is defined as:

$$f(x) := \int_0^x \sum_{i=1}^{k-2}\frac{1}{i}\left(1-\frac{y}{x}\right)^i y^m(1-y)^{k-2-i}\,\mathrm{d}y.$$

From lemma 9, the derivative of $f(x)$ can be calculated as:

$$\begin{aligned}
f'(x) &= \int_0^x \frac{y}{x^2}\left(\sum_{i=1}^{k-2}\left(\frac{1-\frac{y}{x}}{1-y}\right)^{i-1}\right) y^m(1-y)^{k-3}\,\mathrm{d}y \\
&= \int_0^x \frac{y}{x^2}\left(\frac{1-\left(\frac{1-\frac{y}{x}}{1-y}\right)^{k-2}}{1-\frac{1-\frac{y}{x}}{1-y}}\right) y^m(1-y)^{k-3}\,\mathrm{d}y \\
&= \frac{1}{x(1-x)}\int_0^x \left(1-\left(\frac{1-\frac{y}{x}}{1-y}\right)^{k-2}\right) y^m(1-y)^{k-2}\,\mathrm{d}y \\
&= \frac{1}{x(1-x)}\int_0^x y^m(1-y)^{k-2}\,\mathrm{d}y - \frac{1}{x(1-x)}\int_0^x\left(1-\frac{y}{x}\right)^{k-2} y^m\,\mathrm{d}y \\
&= \frac{1}{x(1-x)}\sum_{i=0}^{k-2}\binom{k-2}{i}\int_0^x(-1)^i y^{m+i}\,\mathrm{d}y - \frac{x^{m+1}}{x(1-x)}\int_0^1(1-t)^{k-2}t^m\,\mathrm{d}t \\
&= \frac{x^m}{x(1-x)}\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-x)^i}{m+i+1} - \frac{x^{m+1}}{x(1-x)}\binom{m+k-1}{k-2}^{-1}\frac{1}{m+1}.
\end{aligned} \tag{44}$$

By (51) and lemma 9, the first term of (43) is calculated as:

$$\begin{aligned}
-(k-1)&\binom{m+k-1}{k-1}\int_0^{\frac{a}{1+a}} y^m(1-y)^{k-2}\log\left(\frac{y}{a(1-y)}\right)\mathrm{d}y \\
&= -(k-1)\binom{m+k-1}{k-1}\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\int_0^{\frac{a}{1+a}} y^{m+i}\log\left(\frac{y}{a(1-y)}\right)\mathrm{d}y
\end{aligned}$$

$$
\begin{aligned}
&= -(k-1)\binom{m+k-1}{k-1}\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-1)^i}{m+i+1} \\
&\quad\times\left(-\log(1+a)+\sum_{j=1}^{m+i}\frac{1}{j}\left(\frac{a}{1+a}\right)^j\right) \\
&= -(k-1)\binom{m+k-1}{k-1}\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-1)^i}{m+i+1} \\
&\quad\times\left(-\log(1+a)+\sum_{j=1}^{m}\frac{1}{j}\left(\frac{a}{1+a}\right)^j+\sum_{j=m+1}^{m+i}\frac{1}{j}\left(\frac{a}{1+a}\right)^j\right) \\
&= -(k-1)\binom{m+k-1}{k-1}\left(\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-1)^i}{m+i+1}\right) \\
&\quad\times\left(-\log(1+a)+\sum_{j=1}^{m}\frac{1}{j}\left(\frac{a}{1+a}\right)^j\right) \\
&\quad-(k-1)\binom{m+k-1}{k-1}\sum_{i=0}^{k-2}\sum_{j=1}^{i}\binom{k-2}{i}\frac{(-1)^i}{(m+i+1)(j+m)}\left(\frac{a}{1+a}\right)^{j+m} \\
&= -(k-1)\binom{m+k-1}{k-1}\left(\binom{m+k-1}{k-2}^{-1}\frac{1}{m+1}\right) \\
&\quad\times\left(-\log(1+a)+\sum_{j=1}^{m}\frac{1}{j}\left(\frac{a}{1+a}\right)^j\right)+(k-1)\binom{m+k-1}{k-1}g\left(\frac{a}{1+a}\right) \\
&= \log(1+a)-\sum_{j=1}^{m}\frac{1}{j}\left(\frac{a}{1+a}\right)^j-(k-1)\binom{m+k-1}{k-1}g\left(\frac{a}{1+a}\right) \qquad (45)
\end{aligned}
$$

where $g(x)$ is defined as:

$$
g(x) := \sum_{i=0}^{k-2}\sum_{j=1}^{i}\binom{k-2}{i}\frac{(-1)^i}{(m+i+1)(j+m)}x^{j+m}.
$$

By lemma 9, the derivative of $g(x)$ is calculated as:

$$
\begin{aligned}
g'(x) &= \sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-1)^i}{(m+i+1)}\sum_{j=1}^{i}x^{j+m-1} = \sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-1)^i}{(m+i+1)}x^m\frac{1-x^i}{1-x} \\
&= \frac{x^m}{1-x}\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-1)^i}{(m+i+1)}-\frac{x^m}{1-x}\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-x)^i}{(m+i+1)} \\
&= \frac{x^m}{1-x}\binom{m+k-1}{k-2}^{-1}\frac{1}{m+1}-\frac{x^m}{1-x}\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-x)^i}{(m+i+1)}. \qquad (46)
\end{aligned}
$$

From (44) and (46), we have $f'(x) = -g'(x)$. The definitions of $f(x)$ and $g(x)$ mean that $f(0) = g(0) = 0$. Then we obtain $f(x) = -g(x)$. By (43) and (45), we have

$$
\begin{aligned}
\frac{-D_{\Pi_m}(\rho_1^{(m)}\|\rho_2^{(m)})-2m\log\cos\epsilon}{m} &= (k-1)(k-2)\binom{m+k-1}{k-1}\int_0^{\frac{a}{1+a}} \\
&\quad\times\left(\int_0^{1-\frac{y}{a(1-y)}}x^{k-3}\log\left((1-x)\frac{a(1-y)}{y}\right)\mathrm{d}x\right)y^m(1-y)^{k-2}\mathrm{d}y
\end{aligned}
$$

$$
\begin{aligned}
&= \log(1+a) - \sum_{j=1}^{m} \frac{1}{j}\left(\frac{a}{1+a}\right)^j \\
&\quad -(k-1)\binom{m+k-1}{k-1}\left(g\left(\frac{a}{1+a}\right) + f\left(\frac{a}{1+a}\right)\right) \\
&= \log(1+a) - \sum_{j=1}^{m} \frac{1}{j}\left(\frac{a}{1+a}\right)^j \\
&= -\log\cos^2\epsilon - \sum_{j=1}^{m} \frac{1}{j}\sin^{2j}\epsilon.
\end{aligned}
$$

Then we obtain:

$$\frac{D_{\Pi_m}(\rho_1^{(m)}\|\rho_2^{(m)})}{m} = \sum_{j=1}^{m} \frac{\sin^{2i}\epsilon}{i}.$$

We proved (22).

Next we will prove (23). We consider the tangent space $T_\rho\mathcal{P}(\mathcal{H})$ at $\rho := |\phi(0)\rangle\langle\phi(0)|$. If $c(t)$ is a curve on $\mathcal{P}(\mathcal{H})$ such that $c(0) = \rho$, $\dot{c}$ denotes the element of $T_\rho\mathcal{P}(\mathcal{H})$ defined by $c(t)$. The Fubini–Study metric g_{fs} is defined as:

$$g_{fs}(\dot{c},\dot{c}) := \left(\lim_{t\to 0}\frac{d_{fs}(c(0),c(t))}{t}\right)^2.$$

Therefore, it is sufficient to show that

$$J_{\Pi_n}^{\rho} = 2n g_{fs}.$$

Let $c(t) := |\phi_t\rangle\langle\phi_t|, \phi_t := \phi(t,0,\ldots,0)$. (See equation (6).) Because $g_{fs}(\dot{c},\dot{c}) = 1$, it is sufficient to prove that

$$J_{\Pi_n}^{\rho}(\dot{c},\dot{c}) = 2n.$$

We assume that $k \geqslant 3$. From (41), we have

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\log(|\langle\phi_t|\phi(\theta)\rangle|^{2n})|_{t=0}\right)^2 |\langle\phi_0|\phi(\theta)\rangle|^{2n} = 4n^2\cos^{2n-2}\theta_1\sin^2\theta_1\cos^2\theta_2\cos^2\theta_k. \tag{47}$$

By (47) and (49), we have:

$$
\begin{aligned}
&\binom{m+k-1}{k-1}\int_{\mathcal{P}(\mathcal{H})}\left(\frac{\mathrm{d}}{\mathrm{d}t}\log(|\langle\phi_t|\phi(\theta)\rangle|^{2m})|_{t=0}\right)^2 |\langle\phi_0|\phi(\theta)\rangle|^{2m}\nu(\mathrm{d}\theta) \\
&\quad = \frac{2(k-1)(k-2)}{\pi}\binom{m+k-1}{k-1}4m^2\int_0^{\frac{\pi}{2}}\cos^{2n-1}\theta_1\sin^{2k-1}\theta_1\,\mathrm{d}\theta_1 \\
&\qquad \times\int_0^{\frac{\pi}{2}}\cos^3\theta_2\sin^{2k-5}\theta_2\,\mathrm{d}\theta_2\int_0^{2\pi}\cos^2\theta_k\,\mathrm{d}\theta_k \\
&\quad = \frac{2(k-1)(k-2)}{\pi}\binom{m+k-1}{k-1}4m^2\frac{(m-1)!(k-1)!}{2(m+k-1)!}\frac{1!(k-3)!}{2(k-1)!}\pi \\
&\quad = 2m.
\end{aligned} \tag{48}
$$

We obtain (23). In the case of $k = 2$, we can similarly prove (23).

Lemma 7. If $k \geqslant 3$, we then have

$$\int_{\mathcal{P}(\mathcal{H})} f(\theta_1, \theta_2, \theta_k)\nu(\mathrm{d}\theta) = \frac{2(k-1)(k-2)}{\pi} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} f(\theta_1, \theta_2, \theta_k) \times \cos\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1 \cos\theta_2 \sin^{2k-5}\theta_2 \,\mathrm{d}\theta_2 \,\mathrm{d}\theta_k. \tag{49}$$

Proof. From (7) the left-hand side of (49) is calculated as:

$$\begin{aligned}
\int_{\mathcal{P}(\mathcal{H})} f(\theta_1, \theta_2, \theta_k)\nu(\mathrm{d}\theta) &= \frac{(k-1)!}{\pi^{k-1}} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} f(\theta_1, \theta_2, \theta_k) \\
&\quad \times \cos\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1 \cos\theta_2 \sin^{2k-5}\theta_2 \,\mathrm{d}\theta_2 \,\mathrm{d}\theta_k \\
&\quad \times \underbrace{\int_0^{\frac{\pi}{2}} \ldots \int_0^{\frac{\pi}{2}}}_{k-3} \sin^{2k-7}\theta_3 \ldots \sin\theta_{k-1} \cos\theta_3 \cdots \cos\theta_{k-1} \,\mathrm{d}\theta_2 \ldots \,\mathrm{d}\theta_{k-1} \\
&\quad \times \underbrace{\int_0^{2\pi} \ldots \int_0^{2\pi}}_{k-2} \mathrm{d}\theta_{k+1} \ldots \mathrm{d}\theta_{2k-2} \\
&= \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} f(\theta_1, \theta_2, \theta_k) \cos\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1 \cos\theta_2 \sin^{2k-5}\theta_2 \,\mathrm{d}\theta_2 \,\mathrm{d}\theta_k \\
&\quad \times \frac{(k-1)!}{\pi^{k-1}} \underbrace{\int_0^1 x^{2k-7} \,\mathrm{d}x \ldots \int_0^1 x \,\mathrm{d}x}_{k-3} \cdot (2\pi)^{k-2} \\
&= \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} f(\theta_1, \theta_2, \theta_k) \cos\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1 \cos\theta_2 \sin^{2k-5}\theta_2 \,\mathrm{d}\theta_2 \,\mathrm{d}\theta_k \\
&\quad \times \frac{(k-1)!}{\pi^{k-1}} \frac{(2\pi)^{k-2}}{2^{k-3}(k-3)!} \\
&= \frac{2(k-1)(k-2)}{\pi} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} f(\theta_1, \theta_2, \theta_k) \\
&\quad \times \cos\theta_1 \sin^{2k-3}\theta_1 \,\mathrm{d}\theta_1 \cos\theta_2 \sin^{2k-5}\theta_2 \,\mathrm{d}\theta_2 \,\mathrm{d}\theta_k.
\end{aligned}$$

We then obtain (49). □

Lemma 8. The following integral can be calculated as:

$$\int_0^{1-a} x^m \log\left(\frac{1-x}{a}\right) \mathrm{d}x = \frac{1}{m+1}\left(-\log a - \sum_{i=1}^{m+1} \frac{(1-a)^i}{i}\right) \tag{50}$$

$$\int_0^{\frac{a}{1+a}} x^m \log\left(\frac{x}{a(1-x)}\right) \mathrm{d}x = \frac{1}{m+1}\left(-\log(1+a) + \sum_{i=1}^{m} \frac{1}{i}\left(\frac{a}{1+a}\right)^i\right). \tag{51}$$

Proof. Equation (50) is derived by

$$\int_0^{\alpha} x^m \log(1-x)\mathrm{d}x = \frac{1}{m+1}\left((\alpha^{m+1}-1)\log(1-\alpha) - \sum_{i=1}^{m+1} \frac{\alpha^i}{i}\right). \tag{52}$$

Also, equation (51) is derived by (52) and the following:

$$\int_0^{\alpha} x^m \log x \, \mathrm{d}x = \frac{1}{m+1}\left(\alpha^{m+1}\left(\log\alpha - \frac{1}{m+1}\right)\right). \tag{53}$$

□

Lemma 9. We have the following equations:

$$\sum_{i=0}^{n}\binom{n}{i}\frac{(-1)^i}{m+i} = \int_0^1 x^{m-1}(1-x)^n \, \mathrm{d}x = \binom{m+n}{n}^{-1}\frac{1}{m}. \tag{54}$$

It is easily derived.

References

[1] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[2] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[3] Yuen H P and Lax M 1973 *IEEE Trans.* **19** 740
[4] Holevo A S 1977 *Rep. Math. Phys.* **12** 251
[5] Jones K R W 1994 *Phys. Rev.* A **50** 3682
[6] Jones K R W 1991 *J. Phys. A: Math. Gen.* **24** 121
[7] D'Ariano G M 1997 Homodyning as universal detection *Quantum Communication, Computing, and Measurement* ed O Hirota *et al* (New York: Plenum) p 253
D'Ariano G M 1997 *LANL e-print* quant-ph/9701011
[8] D'Ariano G M and Yuen H P 1996 *Phys. Rev. Lett.* **76** 2832
[9] Bužek V, Adam G and Drobný G 1996 *Phys. Rev.* A **54** 804
[10] Massar S and Popescu S 1995 *Phys. Rev. Lett.* **74** 1259
[11] Nagaoka H 1992 On the relation between Kullback divergence and Fisher information—from classical systems to quantum systems *Proc. Society Information Theory and its Applications in Japan* (in Japanese) pp 63–72
Nagaoka H 1994 Two quantum analogues of the large deviation Cramér–Rao inequality *Proc. 1994 IEEE Int. Symp. on Information Theory* p 118
[12] Fujiwara A 1994 *METR* 94–09, 94–10 University of Tokyo
[13] Fujiwara A and Nagaoka H 1995 *Phys. Lett.* A **13** 199
[14] Fujiwara A and Nagaoka H 1996 Coherency in view of quantum estimation theory *Quantum Coherence and Decoherence* ed K Fujikawa and Y A Ono (Amsterdam: Elsevier) p 303
[15] Matsumoto K 1997 A Geometrical approach to quantum estimation theory *Doctoral Thesis* Graduate School of Mathematical Sciences, University of Tokyo
Matsumoto K 1996 *METR* 96–09 University of Tokyo
Matsumoto K 1997 *LANL e-print* quant-ph/9711008
[16] Nagaoka H 1991 *Trans. Jap. Soc. Ind. App. Math.* **4** 305 (in Japanese)
[17] Hayashi M 1997 A linear programming approach to attainable Cramér-rao type bound *Quantum Communication, Computing, and Measurement* ed O Hirota *et al* (New York: Plenum) p 99
[18] Hayashi M 1997 *Kyoto-Math* 97–08 (Kyoto University)
Hayashi M 1997 *LANL e-print* quant-ph/9704044
[19] Hayashi M 1997 *LANL e-print* quant-ph/9710040
[20] Bures D 1969 *Trans. Am. Math. Soc.* **135** 199
[21] Griffiths P and Harris J 1978 *Principle of Algebraic Geometry* (New York: Wiley)
[22] Jozsa R 1994 *J. Mod. Opt.* **41** 2315
[23] Bahadur R, Zabell S and Gupta J 1980 Large deviations, tests, and estimates *Asymptotic Theory of Statistical Tests and Estimation* ed I M Chatcravarti (New York: Academic) p 33
[24] Fu J C 1973 *Ann. Stat.* **1** 745
Fu J C 1982 *Ann. Stat.* **10** 762