Optimal sequence of quantum measurements in the sense of Stein's lemma in quantum hypothesis testing

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 10759–10773 PII: S0305-4470(02)54390-0

Optimal sequence of quantum measurements in the sense of Stein's lemma in quantum hypothesis testing

Masahito Hayashi

Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN 2-1, Hirosawa, Wako, Saitama 351-0198, Japan

E-mail: masahito@brain.riken.go.jp

Received 3 October 2002
Published 4 December 2002
Online at stacks.iop.org/JPhysA/35/10759

Abstract

We derive a necessary and sufficient condition for a sequence of quantum measurements to achieve the optimal performance in quantum hypothesis testing. We discuss what quantum measurement we should perform in order to attain the optimal exponent of the second error probability under the condition that the first error probability goes to 0. As an asymptotically optimal measurement, we propose a projection measurement characterized by the irreducible representation theory of the special linear group $SL(\mathcal{H})$. Especially, in the spin-$1/2$ system, it is realized by the simultaneous measurement of the total momentum and a momentum of a specified direction. As a by-product, we obtain another proof of quantum Stein's lemma. In addition, an asymptotically optimal measurement is constructed in the quantum Gaussian case, and it is physically meaningful.

PACS numbers: 03.67.−a, 02.50.Tt

1. Introduction

Estimating the true quantum state based on the decision for one of two alternative hypotheses is called quantum hypothesis testing, which is one of the most fundamental problems among quantum information theory for the following reasons.

The difficulty derived from non-commutativity of matrices (operators) appears as a simple form in this problem. This problem can be applied to other related topics in quantum information, for example, quantum channel coding [1–4], distillable entanglement [5], quantum estimation [6], quantum universal variable-length source coding [7] and quantum coin tossing [8].

When the null hypothesis is the tensor product of a certain quantum state ρ and the alternative hypothesis is that of another quantum state σ, we sometimes focus on the asymptotic behaviour of the first error probability (we reject the null hypothesis though it is correct) and the second error probability (we accept the null hypothesis though it is incorrect). Hiai and Petz [9], and Ogawa and Nagaoka [10] discussed the optimal second error exponent under the assumption that the first error probability is less than a certain constant $\epsilon > 0$. Combining their results, we obtain that the optimal second error exponent is independent of $\epsilon > 0$, and coincides with the quantum relative entropy. Hiai and Petz [9] proved the direct part, i.e., the attainability of the quantum relative entropy, and Ogawa and Nagaoka [10] proved the converse part, i.e., the impossibility of surpassing the quantum relative entropy. The converse part was simplified by Nagaoka [2]. In addition, the quantum relative entropy coincides with the optimal second error exponent under the condition that the first error probability asymptotically goes to 0. Moreover, Ogawa and Hayashi [11] discussed the second error exponent under the constant constraint for the first error exponent.

We divide this testing process into two parts: one is the quantum part, i.e., the quantum measurement process. When ρ and σ are non-commutative, the choice of this quantum measurement is difficult and essential. The other is the classical part, i.e., the classical data processing. In this paper, we focus on the former process i.e., we study what kind of measurement is suitable in order to achieve the optimal second error exponent. As mentioned in section 4, it is sufficient for this kind of hypothesis testing to discuss our quantum measurement of a certain class. We derive a necessary and sufficient condition for a quantum measurement to attain the optimal second error exponent among this class. This condition depends on the alternative hypothesis σ, and is almost independent of the null one ρ. As a by-product, we obtain another proof of quantum Stein's lemma.

In our setting, the unknown state is a tensor product state, but our measurement is not necessarily a tensor product. Therefore, in order to treat the classical part we need to discuss our data processing after our measurement as a classical hypothesis testing with two general sources. In classical information theory, by using the information-spectrum method, Han [12, 13] studied hypothesis testing based on such a general setting. We apply it to our proof of the main result, and such an application to quantum hypothesis testing was initiated by Nagaoka [14, 15]. However, this paper is organized so that the reader can understand the statement of the main result without any knowledge of the information-spectrum method. A quantum version of this method was discussed by Nagaoka and Hayashi [16], but it is not treated in this paper because it is not directly related to this issue. This work was motivated by Nagaoka's [14, 15] earlier works. Unfortunately, these papers of Nagaoka were written in Japanese, but Nagaoka and Hayashi [16] contains a part of the results by Nagaoka [14, 15].

This paper is organized as follows. In section 2, we formulate quantum hypothesis testing with tensor product states as an asymptotic problem. In order to discuss our quantum asymptotic setting, we prepare some non-asymptotic characterizations in section 3. After these preparations, we state the main results, i.e., we characterize a quantum measurement to attain the optimal second error exponent in section 4. We treat quantum Gaussian states as a special example of the infinite-dimensional case in section 5, while we assume that the dimension of the Hilbert space of interest is finite in section 4. In the quantum Gaussian case, we give an asymptotically optimal measurement whose physical interpretation is clear. In order to prove our main theorem, we have to discuss a general sequence of classical information sources. Thus, in section 6, by using the information-spectrum method, we prepare a lemma which is applicable to such a general sequence, and then apply it to our issue. We prove the main theorem in section 7 with the help of this lemma. We use some fundamental inequalities in section 7, and these inequalities are given in section 8. Moreover we assume a fundamental fact in section 4 which is proved from a representation viewpoint in section 9.

2. Asymptotic formulation of quantum hypothesis testing

Let $\mathcal{H}$ be the Hilbert space of interest, and $\mathcal{S}(\mathcal{H})$ be the set of density matrices on $\mathcal{H}$. When we perform a measurement corresponding to positive operator valued measure (POVM) $M = \{M_i\}$ to a system in the state ρ, the data obeys the probability $P_\rho^M = \left\{P_\rho^M(i) = \operatorname{Tr} M_i \rho\right\}$. In particular, the POVM $M = \{M_i\}$ is called a projection valued measure (PVM) if each M_i is a projection. In the hypothesis testing, the testing is described by a two-valued POVM $\{M_a, M_r\}$, where M_a corresponds to acceptance and M_r corresponds to rejection. Similarly, an operator A satisfying $0 \leqslant A \leqslant I$ is called a *test*, and is identified with the POVM $\{M_a, M_r\} = \{A, I - A\}$.

Now, we study the quantum hypothesis testing problem for the null hypothesis $H_0 : \rho^{\otimes n} \in \mathcal{S}(\mathcal{H}^{\otimes n})$ versus the alternative hypothesis $H_1 : \sigma^{\otimes n} \in \mathcal{S}(\mathcal{H}^{\otimes n})$, where $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ are the nth-tensor powers of arbitrarily given density operators ρ and σ in $\mathcal{S}(\mathcal{H})$. In the following, an operator A^n on $\mathcal{H}^{\otimes n}$ satisfying $0 \leqslant A^n \leqslant I$ or a sequence $\{A^n\}$ of such operators, is called a *test*. For a test A^n the probabilities of the first and the second are, respectively, defined by

$$\alpha_n(A^n) = \operatorname{Tr} \rho^{\otimes n}(I - A^n) \qquad \text{and} \qquad \beta_n(A^n) = \operatorname{Tr} \sigma^{\otimes n} A^n.$$

We can understand that $\alpha_n(A^n)$ is the probability of erroneously rejecting $\rho^{\otimes n}$ though $\rho^{\otimes n}$ is true and $\beta_n(A^n)$ is the error probability of erroneously accepting $\rho^{\otimes n}$ although $\rho^{\otimes n}$ is not true. We discuss the trade-off of the two types of error probabilities for two n-tensor product states.

The following is known as quantum Stein's lemma.

Theorem 1. *For any $1 > \epsilon > 0$, the equation*

$$-\lim_{n\to\infty} \frac{1}{n} \log \beta_n^*(\epsilon) = D(\rho\|\sigma) := \operatorname{Tr} \rho(\log \rho - \log \sigma)$$

holds, where $\beta_n^*(\epsilon)$ is the minimum second error probability under the constraint that the first error probability is less than ϵ, i.e.,

$$\beta_n^*(\epsilon) := \min\{\beta_n(A^n) | 0 \leqslant A^n \leqslant I, \alpha_n(A^n) \leqslant \epsilon\}.$$

The part $\geqslant$ was proved by Hiai and Petz [9]. Its infinite-dimensional case was proved by Petz [17]. The part $\leqslant$ was proved by Ogawa and Nagaoka [10].

In order to discuss the above theorem, we define the following two quantities:

$$B(\rho\|\sigma) := \sup\left\{\lambda \middle| \exists \vec{A}, \lim_{n\to\infty} \alpha_n(A^n) = 0, \limsup_{n\to\infty} \frac{1}{n} \log \beta_n(A^n) \leqslant -\lambda\right\}$$

$$B^\dagger(\rho\|\sigma) := \sup\left\{\lambda \middle| \exists \vec{A}, \limsup_{n\to\infty} \alpha_n(A^n) < 1, \limsup_{n\to\infty} \frac{1}{n} \log \beta_n(A^n) \leqslant -\lambda\right\}$$

i.e., the former is the optimal second error exponent under the constraint that the first error probability goes to 0 while the latter is the optimal second error exponent under the other constraint that the first error probability does not go to 1. Thus, quantum Stein's lemma (theorem 1) is equivalent to the following relations:

$$B(\rho\|\sigma) = B^\dagger(\rho\|\sigma) = D(\rho\|\sigma).$$

In the following, we divide this testing process into two parts i.e., the quantum part and the classical part. The quantum part corresponds to the choice of a sequence $\vec{M} := \{M^n\}$ of POVMs, and the classical part to the decision process from the classical data. Now, we fix a sequence $\vec{M} := \{M^n\}$ of POVMs, and denote the dataset of M^n by Ω_n. Then, this classical part is formulated as the classical hypothesis testing with the null hypothesis $P_{\rho^{\otimes n}}^{M^n}$ and the alternative hypothesis $P_{\sigma^{\otimes n}}^{M^n}$. In this setting, a test is described by a function T^n from Ω_n to

[0, 1] instead of an operator A^n satisfying $0 \leqslant A^n \leqslant I$. In particular, when a test T^n is a test function with the support S^n, this test is equivalent to the test with the acceptance region S^n. Similar to $B(\rho\|\sigma)$ and $B^\dagger(\rho\|\sigma)$, we define

$$B^{\vec{M}}(\rho\|\sigma) := \sup\left\{\lambda \left| \exists \vec{A},\ \lim_{n\to\infty} E^{M^n}_{\rho^{\otimes n}}(1-T^n) = 0, \limsup_{n\to\infty} \frac{1}{n}\log E^{M^n}_{\sigma^{\otimes n}}(T^n) \leqslant -\lambda \right.\right\}$$

$$B^{\dagger,\vec{M}}(\rho\|\sigma) := \sup\left\{\lambda \left| \exists \vec{A}, \limsup_{n\to\infty} E^{M^n}_{\rho^{\otimes n}}(1-T^n) < 1, \limsup_{n\to\infty} \frac{1}{n}\log E^{M^n}_{\sigma^{\otimes n}}(T^n) \leqslant -\lambda \right.\right\}$$

where E^M_ρ denotes the expectation regarding the probability P^M_ρ. We can easily check that

$$B(\rho\|\sigma) = \sup_{\vec{M}:\text{POVMs}} B^{\vec{M}}(\rho\|\sigma) \leqslant B^\dagger(\rho\|\sigma) = \sup_{\vec{M}:\text{POVMs}} B^{\dagger,\vec{M}}(\rho\|\sigma).$$

Therefore, theorem 1 is equivalent to the following relations:

$$\sup_{\vec{M}:\text{POVMs}} B^{\vec{M}}(\rho\|\sigma) = \sup_{\vec{M}:\text{POVMs}} B^{\dagger,\vec{M}}(\rho\|\sigma) = D(\rho\|\sigma). \tag{1}$$

In this paper, we focus on a sequence $\vec{M} := \{M^n\}_{n=1}^\infty$ of POVMs that satisfies the condition

$$B^{\vec{M}}(\rho\|\sigma) = D(\rho\|\sigma) \tag{2}$$

and call such a sequence $\vec{M}$ of POVMs (PVMs) an *optimal sequence of POVMs (PVMs) in the sense of Stein's lemma*. The main issue is a characterization of a sequence $\vec{M} := \{M^n\}_{n=1}^\infty$ of POVMs that satisfies condition (2). As mentioned in section 4, our characterization of such an optimal sequence is independent of the null hypothesis ρ, and depends only on the alternative hypothesis σ. Of course, in section 4, we construct such an optimal sequence. Indeed, if a sequence $\vec{M}$ of POVMs satisfies condition (2), there exists a sequence $\vec{A} := \{A^n\}$ of tests satisfying

$$\lim_{n\to\infty} \alpha_n(A^n) = 0 \qquad -\lim_{n\to\infty} \frac{1}{n}\log\beta_n(A^n) = D(\rho\|\sigma) - \epsilon \tag{3}$$

for any $\epsilon > 0$. In the following, we assume that the dimension of $\mathcal{H}$ is finite (k) and the inverse σ^{-1} of σ exists.

3. Non-asymptotic characterization of PVMs

In order to treat condition (2), we need some characterizations concerning PVMs in the non-asymptotic setting. One may think that these characterizations have no relation with condition (2), but they are essential for our issue.

A state ρ is called *commutative* with a PVM E (=$\{E_i\}$) on $\mathcal{H}$ if $\rho E_i = E_i\rho$ for any index i. The spectral decomposition of any operator X can be regarded as a PVM and it is denoted by $E(X)$. In particular, we have $E(\sigma) = E(\log\sigma)$. The map $\mathcal{E}_E$ with respect to a PVM E is defined as

$$\mathcal{E}_E : \rho \mapsto \sum_i E_i\rho E_i$$

which is a linear map from the set of Hermite operators to itself. Note that the state $\mathcal{E}_E(\rho)$ is commutative with a PVM E. The number $\sup_i \operatorname{rank} E_i$ of a PVM $E = \{E_i\}$ is an important quantity in the following, and is denoted by $w(E)$. Next, we focus on two PVMs E (=$\{E_i\}_{i\in I}$), F (=$\{F_j\}_{j\in J}$). We write $E \leqslant F$ if there exists a subset $(F/E)_i$ of the index set J such that $E_i = \sum_{j\in(F/E)_i} F_j$ for any index $i \in I$. If a PVM $F = \{F_j\}$ is commutative with a PVM $E = \{E_i\}$, then we can define the PVM $F \times E = \{F_jE_i\}$, which satisfies $F \times E \geqslant E$ and $F \times E \geqslant F$, and can be regarded as the simultaneous measurement of E and F.

Lemma 2. *If ρ and σ are commutative with a PVM E, then the equation*

$$\inf\{\beta(A)|\alpha(A)\leqslant\epsilon\}=\inf\{\beta(A)|\exists M:\mathrm{PVM},\ M\geqslant E, M\geqslant E(A),\alpha(A)\leqslant\epsilon, w(M)=1\}$$

holds.

Proof. For any A, the relations $\beta(\mathcal{E}_E(A))=\beta(A)$ and $\alpha(\mathcal{E}_E(A))=\alpha(A)$ hold. Since the PVM $E(\mathcal{E}_E(A))$ commutes with the PVM E, there exists a PVM M such that $M\geqslant E, M\geqslant E(\mathcal{E}_E(A))$ and $w(M)=1$. □

Indeed, if a test A and a PVM M satisfy $M\geqslant E(A)$, the test A is performed by combining the quantum measurement M and suitable data processing. Therefore, when ρ and σ are commutative with a PVM E, we may discuss only PVMs M satisfying $M\geqslant E$ i.e., we can restrict our tests.

4. Main result

In this section, we discuss condition (2) under the assumption that there exist PVMs E^n such that each PVM E^n is commutative with states $\sigma^{\otimes n}$ and $\rho^{\otimes n}$ and $w(E^n)\leqslant(n+1)^{k-1}$. This existence is proved by the representation theory in section 9. Since, it follows from lemma 2 that we may treat only a PVM satisfying $M^n\geqslant E^n$ and $w(M^n)=1$, we obtain

$$\begin{aligned}B(\rho\|\sigma)&=\sup_{\vec{M}:\mathrm{POVMs}}B^{\vec{M}}(\rho\|\sigma)=\sup_{\vec{M}=\{M^n\}:\mathrm{PVMs\ s.t.}\ M^n\geqslant E^n,w(M^n)=1}B^{\vec{M}}(\rho\|\sigma)\\ B^{\dagger}(\rho\|\sigma)&=\sup_{\vec{M}:\mathrm{POVMs}}B^{\dagger,\vec{M}}(\rho\|\sigma)=\sup_{\vec{M}=\{M^n\}:\mathrm{PVMs\ s.t.}\ M^n\geqslant E^n,w(M^n)=1}B^{\dagger,\vec{M}}(\rho\|\sigma)\end{aligned}\tag{4}$$

i.e., we can discuss $B^{\vec{M}}(\rho\|\sigma)$ and $B^{\dagger,\vec{M}}(\rho\|\sigma)$ only of a sequence $\vec{M}$ satisfying the condition $M^n\geqslant E^n, w(M^n)=1$.

Therefore, our main issue is the asymptotic behaviour of the variable $\frac{1}{n}\log\frac{P^{M^n}_{\rho^{\otimes n}}}{P^{M^n}_{\sigma^{\otimes n}}}$ for the probability distribution $P^{M^n}_{\rho^{\otimes n}}$ under the condition $M^n\geqslant E^n, w(M^n)=1$.

Theorem 3. *For any sequence $\vec{M}=\{M^n\}$ of PVMs satisfying $M^n\geqslant E^n, w(M^n)=1$, the relation*

$$B^{\vec{M}}(\rho\|\sigma)=D(\rho\|\sigma)\tag{5}$$

holds, if and only if the variable $-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}}$ converges to $-\mathrm{Tr}\,\rho\log\sigma$ in the probability distribution $P^{M^n}_{\rho^{\otimes n}}$.

For example, if a PVM M^n is commutative with $\sigma^{\otimes n}$ and satisfies $M^n\geqslant E^n$, $w(M^n)=1$, the equations

$$\begin{aligned}\sum_i P^{M^n}_{\rho^{\otimes n}}(i)\left(\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}}(i)-\mathrm{Tr}\,\rho\log\sigma\right)^2&=\mathrm{Tr}\,\mathcal{E}_{M^n}(\rho^{\otimes n})\left(\frac{1}{n}\log\mathcal{E}_{M^n}(\sigma^{\otimes n})-\mathrm{Tr}\,\rho\log\sigma\right)^2\\ &=\mathrm{Tr}\,\rho^{\otimes n}\left(\frac{1}{n}\log\sigma^{\otimes n}-\mathrm{Tr}\,\rho\log\sigma\right)^2=\mathrm{Tr}\,\rho^{\otimes n}\left(\frac{1}{n}(\log\sigma)^{(n)}-\mathrm{Tr}\,\rho\log\sigma\right)^2\\ &=\frac{1}{n}\mathrm{Tr}\,\rho(\log\sigma-\mathrm{Tr}\,\rho\log\sigma)^2\end{aligned}\tag{6}$$

hold. Equation (6) implies that the variable $\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}}$ converges to $\mathrm{Tr}\,\rho\log\sigma$ in probability. Therefore, it is optimal in the sense of Stein's lemma. This PVM coincides with the PVM proposed by Hayashi [18]. In particular, as guaranteed in section 9, in the spin-1/2 system, $E^n \times E(\sigma^{\otimes n})$ can be regarded as a simultaneous measurement of the total momentum and a momentum of the specified direction. As a by-product, we can prove the following theorem.

Theorem 4. *Any sequence $\vec{M}$ of POVMs satisfies*

$$D(\rho\|\sigma) \geqslant B^{\dagger,\vec{M}}(\rho\|\sigma).$$

Since the existence of E^n is proven in section 9, theorems 3 and 4 yield relation (1), which is equivalent to quantum Stein's lemma.

5. Quantum Gaussian states

In this section, we discuss a quantum hypothesis testing whose hypotheses are quantum Gaussian states ρ_θ on an infinite-dimensional space $L^2(\mathbb{R})$:

$$\rho_\theta := \frac{1}{\pi \overline{N}} \int_{\mathbb{C}} |\alpha\rangle\langle\alpha| \, \mathrm{e}^{-\frac{|\alpha-\theta|^2}{\overline{N}}} \, \mathrm{d}^2\alpha \qquad \forall\theta \in \mathbb{C}$$

where we define the boson coherent vector $|\alpha\rangle := \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{k=0}^{\infty} \frac{\alpha^k}{\sqrt{k!}} |k\rangle$ and $|k\rangle$ is the number vector on $L^2(\mathbb{R})$. Now, we propose a suitable PVM. This PVM not only attains the optimal exponents of the second error probability, but also has an intuitive physical meaning. When the null hypothesis is the state $\rho_{\theta_0}^{\otimes n}$ and the alternative hypothesis is the state $\rho_{\theta_1}^{\otimes n}$, our PVM is constructed as follows. First, we perform the unitary evolution as

$$\rho_\theta^{\otimes n} \mapsto \rho_{\theta-\theta_1}^{\otimes n}.$$

Second, we perform the unitary evolutions

$$\rho_{\theta-\theta_1}^{\otimes n} \mapsto \rho_{\sqrt{n}(\theta-\theta_1)} \otimes \rho_0^{\otimes(n-1)}.$$

The constructions of these unitary evolutions were mentioned in appendix F of [6]. Finally, we perform a number detection $\{|k\rangle\langle k|\}_{k=0}^{\infty}$ to the system whose state is $\rho_{\sqrt{n}(\theta-\theta_1)}$ and denote the final data (this PVM) by k $\left(M_{\theta_1}^n\right)$. Following section 7.1 in [6] and its appendices, we obtain

$$P_{\rho_{\theta_0}^{\otimes n}}^{M_{\theta_1}^n} \left\{ \left| \sqrt{\frac{k}{n}} - |\theta_0 - \theta_1| \right| > \epsilon \right\} \to 0$$

$$\frac{-1}{n} \log P_{\rho_{\theta_1}^{\otimes n}}^{M_{\theta_1}^n} \left\{ \sqrt{\frac{k}{n}} \geqslant |\theta_0 - \theta_1| \right\} \to |\theta_0 - \theta_1|^2 \log\left(1 + \frac{1}{N}\right) = D\big(\rho_{\theta_0} \big\| \rho_{\theta_1}\big)$$

for any $\epsilon > 0$, and any $\theta \in \mathbb{C}$. Therefore, when we choose the acceptance region as $\left\{ \left| \sqrt{\frac{k}{n}} - |\theta_0 - \theta_1| \right| > \epsilon \right\}$, the optimal exponent of the second error probability can be approximately attained. Note that this measurement depends on the alternative hypothesis ρ_{θ_1}, and is almost independent of the null hypothesis ρ_{θ_0}. This optimality is guaranteed because the converse part by Ogawa and Nagaoka [10] is valid in this case. Thus, it implies

$$B^{\vec{M}_{\theta_1}}\left(\rho_{\theta_0} \big\| \rho_{\theta_1}\right) = D\big(\rho_{\theta_0} \big\| \rho_{\theta_1}\big)$$

where $\vec{M}_{\theta_1} := \left\{ M_{\theta_1}^n \right\}$.

6. Application of information-spectrum method

In order to prove theorems 3 and 4, we have to treat the general sequences of probabilities having no structure like a Markov chain because the sequence of two probabilities $P^{M^n}_{\rho^{\otimes n}}$ and $P^{M^n}_{\sigma^{\otimes n}}$ generally has no structure. In the classical information theory, Han [12, 13] introduced the information-spectrum method in order to treat a general sequence of information sources. In this section, we simply review the information-spectrum method in classical hypothesis testing, and by using this method, we characterize $B^{\vec{M}}(\rho\|\sigma)$ and $B^{\dagger,\vec{M}}(\rho\|\sigma)$. Given two general sequences of probabilities $\vec{p} = \{p_n\}$ and $\vec{q} = \{q_n\}$ on the same probability sets $\{\Omega_n\}$, we may define the general hypothesis testing problem with $\vec{p} = \{p_n\}$ as the null hypothesis and $\vec{q} = \{q_n\}$ as the alternative hypothesis. In this situation, any *classical test* is described by a function $T^n : \Omega_n \to [0, 1]$. For any test T^n, the error probabilities of the first and the second are, respectively, defined by

$$\alpha_n(T^n) := \sum_{\omega_n\in\Omega_n} (1 - T^n(\omega_n))p_n(\omega_n) \qquad \beta_n(T^n) := \sum_{\omega_n\in\Omega_n} T^n(\omega_n)q_n(\omega_n).$$

We focus on the following two quantities:

$$B(\vec{p}\|\vec{q}) := \sup\left\{\lambda \left| \exists\vec{A},\ \lim_{n\to\infty} \alpha_n(T^n) = 0, \limsup_{n\to\infty} \frac{1}{n}\log\beta_n(T^n) \leqslant -\lambda \right.\right\}$$

$$B^{\dagger}(\vec{p}\|\vec{q}) := \sup\left\{\lambda \left| \exists\vec{A},\ \liminf_{n\to\infty} \alpha_n(T^n) < 1, \limsup_{n\to\infty} \frac{1}{n}\log\beta_n(T^n) \leqslant -\lambda \right.\right\}$$

which can be regarded as generalizations of $B^{\vec{M}}(\rho\|\sigma)$ and $B^{\dagger,\vec{M}}(\rho\|\sigma)$. In the independent identical distributed (i.i.d.) case of p and q, as is known as Stein's lemma, these two values $B(\vec{p}\|\vec{q})$ and $B^{\dagger}(\vec{p}\|\vec{q})$ coincide with the relative entropy (Kullback–Leibler divergence) $D(p\|q)$. Since the relative entropy $D(p\|q)$ is the expectation of the variable $\log\frac{p}{q}$ under the distribution p, these two values $B(\vec{p}\|\vec{q})$ and $B^{\dagger}(\vec{p}\|\vec{q})$ seem related to the variable $\frac{1}{n}\log\frac{p_n}{q_n}$ under the distribution p_n. In order to characterize the asymptotic behaviour of the variable $\frac{1}{n}\log\frac{p_n}{q_n}$ under the distribution p_n, we define the other two values by

$$\underline{D}(\vec{p}\|\vec{q}) := \sup\left\{\lambda \left| \lim_{n\to\infty} p_n\left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} < \lambda \right.\right\} = 0 \right.\right\}$$

$$\overline{D}(\vec{p}\|\vec{q}) := \inf\left\{\lambda \left| \lim_{n\to\infty} p_n\left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} > \lambda \right.\right\} = 0 \right.\right\}.$$

As guaranteed by the following lemma, two values $B(\vec{p}\|\vec{q})$ and $B(\vec{p}\|\vec{q})$ are characterized by the asymptotic behaviour of the variable $\frac{1}{n}\log\frac{p_n}{q_n}$ under the distribution p_n.

Lemma 5 (Han [12], Verdú [19], Nagaoka [14, 15]). *We can show the relations*

$$B(\vec{p}\|\vec{q}) = \underline{D}(\vec{p}\|\vec{q}) \tag{7}$$

$$B^{\dagger}(\vec{p}\|\vec{q}) = \overline{D}(\vec{p}\|\vec{q}). \tag{8}$$

Defining a test $T^n(\lambda)$ as the test with the acceptance region $S_n(\lambda)$:

$$S_n(\lambda) := \left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} \geqslant \lambda \right.\right\}$$

we obtain

$$\alpha_n(T^n(\underline{D}(\vec{p}\|\vec{q}) - \epsilon)) \to 0 \tag{9}$$

$$\beta_n(T^n(\underline{D}(\vec{p}\|\vec{q}) - \epsilon)) \leqslant \mathrm{e}^{-n(\underline{D}(\vec{p}\|\vec{q})-\epsilon)} \tag{10}$$

$$\liminf_{n\to\infty} \alpha_n(T^n(\overline{D}(\vec{p}\|\vec{q}) - \epsilon)) < 1 \tag{11}$$

$$\beta_n(T^n(\overline{D}(\vec{p}\|\vec{q}) - \epsilon)) \leqslant \mathrm{e}^{-n(\overline{D}(\vec{p}\|\vec{q})-\epsilon)} \tag{12}$$

for any $\epsilon > 0$. Thus, the tests $T^n(\underline{D}(\vec{p}\|\vec{q}) - \epsilon)$ and $T^n(\overline{D}(\vec{p}\|\vec{q}) - \epsilon)$ approximately attain the bounds $\underline{D}(\vec{p}\|\vec{q})$ and $\overline{D}(\vec{p}\|\vec{q})$, respectively.

Equation (7) was proved in chapter 4 by Han [12]. He referred to [19]. Equation (8) was derived by Nagaoka [14, 15]. Lemma 5 and a comprehensive review of the information-spectrum method are presented in [16]. For the reader's convenience, we give a proof in appendix A.

Next, in order to apply lemma 5 to the characterization of $B^{\vec{M}}(\rho\|\sigma)$ and $B^{\dagger,\vec{M}}(\rho\|\sigma)$, we define

$$\underline{D}^{\vec{M}}(\rho\|\sigma) := \underline{D}\big(\{P^{M^n}_{\rho^{\otimes n}}\}\big\|\{P^{M^n}_{\sigma^{\otimes n}}\}\big) \qquad \overline{D}^{\vec{M}}(\rho\|\sigma) := \overline{D}\big(\{P^{M^n}_{\rho^{\otimes n}}\}\big\|\{P^{M^n}_{\sigma^{\otimes n}}\}\big)$$

for any sequence $\vec{M}$ of POVMs. From lemma 5, we have

$$\underline{D}^{\vec{M}}(\rho\|\sigma) = B^{\vec{M}}(\rho\|\sigma) \qquad \overline{D}^{\vec{M}}(\rho\|\sigma) = B^{\dagger,\vec{M}}(\rho\|\sigma).$$

Therefore, a sequence $\vec{M}$ of POVMs is optimal in the sense of Stein's lemma if and only if

$$\underline{D}^{\vec{M}}(\rho\|\sigma) = D(\rho\|\sigma).$$

In the following, we discuss $\underline{D}^{\vec{M}}(\rho\|\sigma)$ and $\overline{D}^{\vec{M}}(\rho\|\sigma)$ instead of $B^{\vec{M}}(\rho\|\sigma)$ and $B^{\dagger,\vec{M}}(\rho\|\sigma)$.

In this paper, we use only lemma 5 among several results regarding the information-spectrum method, and this lemma is sufficient for our current issue. This paper treats only an application of the classical information-spectrum method to quantum hypothesis testing, while Nagaoka and Hayashi [16] discussed a quantum analogue of the information-spectrum method. The references, Han [12, 13] and Nagaoka and Hayashi [16], may be useful for the reader who is interested in other related topics concerning the information-spectrum method.

7. Proof of the main result

In this section, by using lemma 5, we prove theorems 3 and 4. In our proofs, we use lemma 6 and several fundamental inequalities given in section 8.

Proof of theorem 3. As guaranteed by lemma 5, it is sufficient to show

$$\underline{D}^{\vec{M}}(\rho\|\sigma) = D(\rho\|\sigma).$$

First, we prove that the variable $\frac{1}{n}\log P^{M^n}_{\rho^{\otimes n}}$ converges to $\operatorname{Tr}\rho\log\rho$ in probability. We can calculate

$$\begin{aligned}
\sum_i P^{M^n}_{\rho^{\otimes n}}(i)\left(\frac{1}{n}\log P^{M^n}_{\rho^{\otimes n}}(i) - \operatorname{Tr}\rho\log\rho\right)^2 &= \operatorname{Tr}\mathcal{E}_{M^n}(\rho^{\otimes n})\left(\frac{1}{n}\log\mathcal{E}_{M^n}(\rho^{\otimes n}) - \operatorname{Tr}\rho\log\rho\right)^2 \\
&= \operatorname{Tr}\rho^{\otimes n}\left(\frac{1}{n}\log\mathcal{E}_{M^n}(\rho^{\otimes n}) - \frac{1}{n}\operatorname{Tr}\rho\log\rho\right)^2 \\
&\leqslant 2\operatorname{Tr}\rho^{\otimes n}\left(\frac{1}{n}\log\mathcal{E}_{M^n}(\rho^{\otimes n}) - \frac{1}{n}\log\rho^{\otimes n}\right)^2 \\
&\quad + 2\operatorname{Tr}\rho^{\otimes n}\left(\frac{1}{n}\log\rho^{\otimes n} - \operatorname{Tr}\rho\log\rho\right)^2
\end{aligned}$$

$$\leqslant 8\left(\frac{(k-1)\log(n+1)}{n}\right)^2 + 2\,\mathrm{Tr}\,\rho^{\otimes n}\left(\frac{1}{n}(\log\rho)^{(n)} - \mathrm{Tr}\,\rho\log\rho\right)^2$$
$$= 8\left(\frac{(k-1)\log(n+1)}{n}\right)^2 + \frac{2}{n}\,\mathrm{Tr}\,\rho\,(\log\rho - \mathrm{Tr}\,\rho\log\rho)^2$$

where the last inequality follows from lemma 7 given in section 8. Thus, the variable $\frac{1}{n}\log P^{M^n}_{\rho^{\otimes n}}$ converges to $\mathrm{Tr}\,\rho\log\rho$ in probability.

Since $\frac{1}{n}\log\frac{P^{M^n}_{\rho^{\otimes n}}}{P^{M^n}_{\sigma^{\otimes n}}} = \frac{1}{n}\log P^{M^n}_{\rho^{\otimes n}} - \frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}}$ and $D(\rho\|\sigma) = \mathrm{Tr}\,\rho\log\rho - \mathrm{Tr}\,\rho\log\sigma$, condition (13) is equivalent to

$$-\,\mathrm{Tr}\,\rho\log\sigma = \sup\left\{\lambda\left|\lim_{n\to\infty} P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} < \lambda\right\} = 0\right.\right\}. \tag{13}$$

Since

$$\sup\left\{\lambda\left|\lim_{n\to\infty} P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} < \lambda\right\} = 0\right.\right\} \leqslant \inf\left\{\lambda\left|\lim_{n\to\infty} P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} > \lambda\right\} = 0\right.\right\}$$

it follows from lemma 6 that condition (13) is equivalent to

$$-\,\mathrm{Tr}\,\rho\log\sigma = \sup\left\{\lambda\left|\lim_{n\to\infty} P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} < \lambda\right\} = 0\right.\right\}$$
$$= \inf\left\{\lambda\left|\lim_{n\to\infty} P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} > \lambda\right\} = 0\right.\right\}.$$

Therefore, if and only if the variable $-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}}$ converges to $-\,\mathrm{Tr}\,\rho\log\sigma$ in probability, relation (13) holds. □

Next, we give lemma 6 with a proof, which is used in our proof of theorem 3.

Lemma 6. *Under the same assumption as theorem 3, we obtain*

$$-\,\mathrm{Tr}\,\rho\log\sigma \geqslant \inf\left\{\lambda\left|\lim_{n\to\infty} P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} > \lambda\right\} = 0\right.\right\}. \tag{14}$$

Proof. We discuss the asymptotic behaviour of the variable $\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}}$. From the Markov inequality, we have

$$p\{X \geqslant a\} \leqslant \mathrm{e}^{-\Lambda(X,p,a)} \qquad \Lambda(X,p,a) := \sup_{0\leqslant t\leqslant 1}\left(at - \log\int \mathrm{e}^{tX(\omega)}p(\mathrm{d}\omega)\right).$$

We can calculate

$$\Lambda\left(-\log P^{M^n}_{\sigma^{\otimes n}}, P^{M^n}_{\rho^{\otimes n}}, an\right) = \sup_{0\leqslant t\leqslant 1} ant - \log\mathrm{Tr}(\mathcal{E}_{M^n}(\rho^{\otimes n})(\mathcal{E}_{M^n}(\sigma^{\otimes n}))^{-t})$$
$$= \sup_{0\leqslant t\leqslant 1} ant - \log\mathrm{Tr}(\rho^{\otimes n}(\mathcal{E}_{M^n}(\sigma^{\otimes n}))^{-t})$$
$$\geqslant \sup_{0\leqslant t\leqslant 1} ant - (t\log w(E^n) + \log\mathrm{Tr}\,\rho^{\otimes n}(\sigma^{\otimes n})^{-t})$$
$$= \sup_{0\leqslant t\leqslant 1} n\left(at - t\frac{\log w(E^n)}{n} - \log\mathrm{Tr}\,\rho\sigma^{-t}\right)$$

where the inequality $\geqslant$ follows from lemma 10 given in section 8. If $a > -\,\mathrm{Tr}\,\rho\log\sigma$, then there exists a real number t_0 such that $0 < t_0 < 1$ and $\frac{\log\mathrm{Tr}\,\rho\sigma^{-t_0}}{t_0} < a$ because

$\lim_{t\to 0}\frac{\log \operatorname{Tr}\rho\sigma^{-t}}{t} = -\operatorname{Tr}\rho\log\sigma$. Therefore the inequalities

$$\begin{aligned}
\liminf_{n\to\infty}\frac{-1}{n}\log P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} > a\right\} &\geqslant \liminf_{n\to\infty}\Lambda\left(-\log P^{M^n}_{\sigma^{\otimes n}}, P^{M^n}_{\rho^{\otimes n}}, an\right)\\
&\geqslant \lim_{n\to\infty}\sup_{0\leqslant t\leqslant 1}\left(at - t\frac{(k+1)\log(n+1)}{n} - \log\operatorname{Tr}\rho\sigma^{-t}\right)\\
&= \sup_{0\leqslant t\leqslant 1}(at - \log\operatorname{Tr}\rho\sigma^{-t}) \geqslant t_0\left(a - \frac{\log\operatorname{Tr}\rho\sigma^{-t_0}}{t_0}\right) > 0
\end{aligned}$$

hold, i.e.,

$$\lim P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} > a\right\} = 0.$$

Thus, inequality (14) holds. □

Proof of theorem 4. From equation (4) and lemma 5, it is sufficient to prove the inequality

$$D(\rho\|\sigma) \geqslant \overline{D}^{\vec{M}}(\rho\|\sigma)$$

for any sequence $\vec{M} = \{M^n\}$ such that $M^n \geqslant E^n$ and $w(M^n) = 1$. As shown in our proof of theorem 3, the variable $\frac{1}{n}\log P^{M^n}_{\rho^{\otimes n}}$ converges to $\operatorname{Tr}\rho\log\rho$ in probability. Therefore,

$$\begin{aligned}
\overline{D}^{\vec{M}}(\rho\|\sigma) &= \operatorname{Tr}\rho\log\rho + \inf\left\{\lambda\left|\lim_{n\to\infty}P^{M^n}_{\rho^{\otimes n}}\left\{-\frac{1}{n}\log P^{M^n}_{\sigma^{\otimes n}} > \lambda\right\} = 0\right.\right\}\\
&\leqslant \operatorname{Tr}\rho\log\rho - \operatorname{Tr}\rho\log\sigma = D(\rho\|\sigma)
\end{aligned}$$

where the inequality follows from (14). □

8. Fundamental inequalities

In this section, we give some fundamental inequalities used in our proofs of theorem 3 and lemma 6.

Lemma 7. *If PVMs E, M satisfy $M \geqslant E$ and a state ρ is commutative with E and $w(E) \geqslant 3$, then the inequality*

$$\operatorname{Tr}\rho(\log\rho - \log\mathcal{E}_M(\rho))^2 \leqslant 4(\log w(E))^2 \tag{15}$$

holds.

Proof. Define $a_i := \operatorname{Tr}E_i\rho E_i$, $\rho_i := \frac{1}{a_i}E_i\rho E_i$, then the equations $\rho = \sum_i a_i\rho_i$, $\mathcal{E}_M(\rho) = \sum_i a_i\mathcal{E}_M(\rho_i)$ hold. Using the operator inequality $(A+B)^2 \leqslant 2(A^2+B^2)$, we have

$$\begin{aligned}
\operatorname{Tr}\rho(\log\rho - \log\mathcal{E}_M(\rho))^2 &= \sum_i a_i\operatorname{Tr}\rho_i(\log\rho_i - \log\mathcal{E}_M(\rho_i))^2\\
&\leqslant \sup_i\operatorname{Tr}\rho_i(\log\rho_i - \log\mathcal{E}_M(\rho_i))^2 \leqslant \sup_i\operatorname{Tr}\rho_i 2((\log\rho_i)^2 + (\log\mathcal{E}_M(\rho_i))^2)\\
&= 2\sup_i(\operatorname{Tr}\rho_i(\log\rho_i)^2 + \operatorname{Tr}\mathcal{E}_M(\rho_i)(\log\mathcal{E}_M(\rho_i))^2) \leqslant 4\sup_i(\log\dim E_i)^2
\end{aligned}$$

where the last inequality follows from lemma 8. We obtain (15). □

Lemma 8 (Nagaoka [20], Osawa [21]). *The equation*

$$\max\left\{\sum_{i=1}^{k} p_i(\log p_i)^2 \,\middle|\, p_i \geqslant 0, \sum_{i=1}^{k} p_i = 1\right\}$$
$$= \begin{cases} (\log k)^2 & \text{if} \quad k \geqslant 3 \\ \frac{1-\sqrt{1-\frac{4}{e^2}}}{2}\left(\log\frac{1-\sqrt{1-\frac{4}{e^2}}}{2}\right)^2 + \frac{1+\sqrt{1-\frac{4}{e^2}}}{2}\left(\log\frac{1+\sqrt{1-\frac{4}{e^2}}}{2}\right)^2 & \text{if} \quad k = 2 \end{cases} \tag{16}$$

holds.

Its proof is given in appendix B.

Lemma 9. *Let k be the dimension of $\mathcal{H}$. For any state $\rho \in \mathcal{S}(\mathcal{H})$ and any PVM M, the inequality $\rho \leqslant k\mathcal{E}_M(\rho)$ holds.*

Proof. The relations

$$\langle\psi|(\mathcal{E}_M(|\phi\rangle\langle\phi|)k - |\phi\rangle\langle\phi|)|\psi\rangle = k\sum_{i=1}^{k}\langle\psi|M_i|\phi\rangle\langle\phi|M_i|\psi\rangle - \left|\sum_{i=1}^{k}\langle\psi|M_i|\phi\rangle\right|^2 \geqslant 0$$

hold for $\forall\phi, \forall\psi \in \mathcal{H}$, where the inequality follows from Schwartz' inequality about vectors $\{\langle\psi|M_i|\phi\rangle\}_{i=1}^{k}, \{1\}_{i=1}^{k}$. Thus, we obtain $|\phi\rangle\langle\phi| \leqslant k\mathcal{E}_M(|\phi\rangle\langle\phi|)$. Any state ρ can be decomposed as $\rho = \sum_i s_i|\phi_i\rangle\langle\phi_i|$. Thus,

$$\rho = \sum_i s_i|\phi_i\rangle\langle\phi_i| \leqslant \sum_i s_i k\mathcal{E}_M(|\phi_i\rangle\langle\phi_i|) = k\mathcal{E}_M(\rho).$$

The proof is completed. □

Lemma 10. *Let ρ be a state commuting with the PVM E. If the PVM M satisfies $M \geqslant E$, the operator inequality*

$$w(E)^t\rho^{-t} \geqslant (\mathcal{E}_M(\rho))^{-t} \tag{17}$$

holds for $0 < t \leqslant 1$ when ρ^{-1} is bounded.

Proof. Based on the same notation as in our proof of lemma 7, it follows from lemma 9 that

$$\rho = \sum_i a_i\rho_i \leqslant \sum_i a_i \operatorname{rank} E_i\mathcal{E}_M(\rho_i) \leqslant \sum_i a_i w(E)\mathcal{E}_M(\rho_i) = w(E)\mathcal{E}_M(\rho).$$

Since the map $u \to -u^{-t}$ $(0 < t \leqslant 1)$ is an operator monotone function in $(0, \infty)$ [22], the operator inequality (17) holds. □

9. Relation between $\rho^{\otimes n}, \sigma^{\otimes n}$ and group representation

In this section, by using the representation theory, we prove the existence of a PVM E^n such that the PVM E^n is commutative with states $\sigma^{\otimes n}$ and $\rho^{\otimes n}$ and $w(E^n) \leqslant (n+1)^{k-1}$. In subsection 9.1, for this purpose, we consider the relation between irreducible representations and PVMs. In subsection 9.2, we discuss n-tensor product states from a group theoretical viewpoint and prove the desired existence.

9.1. Group representation and its irreducible decomposition

Let V be a finite-dimensional vector space over the complex numbers $\mathbb{C}$. A map π from a group G to the generalized linear group of a vector space V is called a *representation* on V if the map π is a homomorphism, i.e. $\pi(g_1)\pi(g_2) = \pi(g_1g_2), \forall g_1, g_2 \in G$. A subspace W of V is called *invariant* with respect to a representation π if the vector $\pi(g)w$ belongs to the subspace W for any vector $w \in W$ and any element $g \in G$. A representation π is called *irreducible* if there is no proper nonzero invariant subspace of V with respect to π. Let π_1 and π_2 be representations of a group G on V_1 and V_2, respectively. The *tensored* representation $\pi_1 \otimes \pi_2$ of G on $V_1 \otimes V_2$ is defined as $(\pi_1 \otimes \pi_2)(g) = \pi_1(g) \otimes \pi_2(g)$, and the *direct sum* representation $\pi_1 \oplus \pi_2$ of G on $V_1 \oplus V_2$ is also defined as $(\pi_1 \oplus \pi_2)(g) = \pi_1(g) \oplus \pi_2(g)$.

In the following, we treat a representation π of a group G on a finite-dimensional Hilbert space $\mathcal{H}$; The following facts are crucial for the latter arguments. There exists an irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ such that the irreducible components are orthogonal to one another if for any element $g \in G$ there exists an element $g^* \in G$ such that $\pi(g)^* = \pi(g^*)$, where $\pi(g)^*$ denotes the adjoint of the linear map $\pi(g)$. We can regard the irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ as the PVM $\left\{P_{\mathcal{H}_i}\right\}_{i=1}^{l}$, where $P_{\mathcal{H}_i}$ denotes the projection to $\mathcal{H}_i$. If two representations π_1 and π_2 satisfy the preceding condition, then the tensored representation $\pi_1 \otimes \pi_2$ also satisfies it. Note that, in general, an irreducible decomposition of a representation satisfying the preceding condition is not unique. In other words, we cannot uniquely define the PVM from such a representation.

9.2. Relation between the tensored representation and PVMs

Let the dimension of the Hilbert space $\mathcal{H}$ be k. Concerning the natural representation $\pi_{SL(\mathcal{H})}$ of the special linear group $SL(\mathcal{H})$ on $\mathcal{H}$, we consider its nth tensored representation $\pi_{SL(\mathcal{H})}^{\otimes n} := \underbrace{\pi_{SL(\mathcal{H})} \otimes \cdots \otimes \pi_{SL(\mathcal{H})}}_{n}$ on the tensored space $\mathcal{H}^{\otimes n}$ [23, 24]. For any element $g \in SL(\mathcal{H})$, the relation $\pi_{SL(\mathcal{H})}(g)^* = \pi_{SL(\mathcal{H})}(g^*)$ holds where the element $g^* \in SL(\mathcal{H})$ denotes the adjoint matrix of the matrix g. Consequently, there exists an irreducible decomposition of $\pi_{SL(\mathcal{H})}^{\otimes n}$ regarded as a PVM and we denote the set of such PVMs by $Ir^{\otimes n}$.

From Weyl's dimension formula ((7.1.8) or (7.1.17) in [24]), the nth symmetric tensored space is the maximum-dimensional space in the irreducible subspaces with respect to the nth tensored representation $\pi_{SL(\mathcal{H})}^{\otimes n}$. Its dimension equals the repeated combination ${}_kH_n$ evaluated by ${}_kH_n = \binom{n+k-1}{k-1} = \binom{n+k-1}{n} = {}_{n+1}H_{k-1} \leqslant (n+1)^{k-1}$. Thus, any element $E^n \in Ir^{\otimes n}$ satisfies $w(E^n) \leqslant (n+1)^{k-1}$.

Lemma 11. *A PVM $E^n \in Ir^{\otimes n}$ is commutative with the nth tensored state $\rho^{\otimes n}$ of any state ρ on $\mathcal{H}$.*

Proof. If $\det \rho \neq 0$, then this lemma is trivial from the fact that $\det(\rho)^{-1}\rho \in SL(\mathcal{H})$. If $\det \rho = 0$, there exists a sequence $\{\rho_i\}_{i=1}^{\infty}$ such that $\det \rho_i \neq 0$ and $\rho_i \to \rho$ as $i \to \infty$. We have $\rho_i^{\otimes n} \to \rho^{\otimes n}$ as $i \to \infty$. Because a PVM $E^n \in Ir^{\otimes n}$ is commutative with $\rho_i^{\otimes n}$, it is also commutative with $\rho^{\otimes n}$. □

Therefore, the existence of a desired PVM is proved. In particular, in the spin-1/2 system, E^n corresponds to the measurement of the total momentum. Therefore, $E^n \times E(\sigma^{\otimes n})$ can be regarded as a simultaneous measurement of the total momentum and a momentum of the specified direction.

10. Conclusion

We discuss quantum measurements from the viewpoint of quantum hypothesis testing. We characterize a sequence of quantum measurements whose second error exponent attains the quantum relative entropy in theorem 3 in the finite-dimensional case. As a by-product, we give another proof of quantum Stein's lemma. This characterization is closely related to the irreducible decomposition of the tensored representation of the group $SL(\mathcal{H})$. In our proof of the main theorem, the information-spectrum method plays an important role. In the further research of quantum information, this method seems a powerful and useful technique. In addition, as a special case of the infinite-dimensional case, we treat the quantum Gaussian states. The photon counting measurement is used in the construction of our asymptotically optimal measurement, and this fact indicates its importance.

Acknowledgment

The author wishes to thank Professor H Nagaoka, Mr S Osawa, Dr T Ogawa and Dr Y Watanabe for useful comments.

Appendix A. Proof of lemma 5

We simplify $\underline{D}(\vec{p}\|\vec{q})$ and $\overline{D}(\vec{p}\|\vec{q})$ by $\underline{D}$ and $\overline{D}$, respectively.

Direct part $\geqslant$ of (7). For any $\epsilon > 0$, we have

$$\alpha_n(T^n(\underline{D}-\epsilon)) = p_n(S_n(\underline{D}-\epsilon)^c) = p_n\left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} < \underline{D}-\epsilon\right.\right\} \to 0$$

and

$$\begin{aligned}\beta_n(T^n(\underline{D}-\epsilon)) &= q_n\left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} \geqslant \underline{D}-\epsilon\right.\right\}\\ &\leqslant \mathrm{e}^{-n(\underline{D}-\epsilon)} p_n\left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} \geqslant \underline{D}-\epsilon\right.\right\} \leqslant \mathrm{e}^{-n(\underline{D}-\epsilon)}\end{aligned}$$

which imply (9) and (10). Thus,

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta_n(T^n(\underline{D}-\epsilon)) \leqslant -(\underline{D}-\epsilon).$$

Direct part $\geqslant$ of (8). Note that

$$\overline{D} = \sup\left\{\lambda \left| \liminf_{n\to\infty} p_n\left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} \leqslant \lambda\right.\right\} < 1\right.\right\}.$$

For any $\epsilon > 0$, similarly, we have

$$\liminf_{n\to\infty}\alpha_n(T^n(\overline{D}-\epsilon)) = \liminf_{n\to\infty} p_n\left\{\omega_n \left| \frac{1}{n}\log\frac{p_n(\omega_n)}{q_n(\omega_n)} < \overline{D}-\epsilon\right.\right\} < 1$$

$$\beta_n(T^n(\overline{D}-\epsilon)) \leqslant \mathrm{e}^{-n(\overline{D}-\epsilon)}$$

which imply (11) and (12). Thus,

$$\liminf_{n\to\infty}\frac{1}{n}\log\beta_n(T^n(\overline{D}-\epsilon)) \leqslant -(\overline{D}-\epsilon).$$

Converse part $\leqslant$ of (7). Assume that $\alpha_n(T^n) \to 0$ as $n \to \infty$ and

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_n(T^n) = -R.$$

For any $\epsilon > 0$, the inequality

$$\alpha_n(T^n(R-\epsilon)) + \mathrm{e}^{n(R-\epsilon)}\beta_n(T^n(R-\epsilon)) = 1 + \sum_{\omega_n}(\mathrm{e}^{n(R-\epsilon)}q_n(\omega_n) - p_n(\omega_n))T^n(R-\epsilon)(\omega_n)$$

$$\leqslant 1 + \sum_{\omega_n}(\mathrm{e}^{n(R-\epsilon)}q_n(\omega_n) - p_n(\omega_n))T^n(\omega_n) = \alpha_n(T^n) + \mathrm{e}^{n(R-\epsilon)}\beta_n(T^n) \qquad \text{(A.1)}$$

holds. Since the RHS goes to 0 and $\mathrm{e}^{n(R-\epsilon)}\beta_n(T^n(R-\epsilon)) \geqslant 0$, the relation

$$p_n\left\{\omega_n \left| \frac{1}{n} \log \frac{p_n(\omega_n)}{q_n(\omega_n)} < R - \epsilon \right.\right\} = \alpha_n(T^n(R-\epsilon)) \to 0$$

holds. It implies that $R - \epsilon < \underline{D}$.

Converse part $\leqslant$ of (8). Assume that $\liminf_{n\to\infty} \alpha_n(T^n) < 1$ and

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_n(T^n) = -R. \qquad \text{(A.2)}$$

For any $\epsilon > 0$, from (A.1) and (A.2), we have

$$\liminf_{n\to\infty} p_n\left\{\omega_n \left| \frac{1}{n} \log \frac{p_n(\omega_n)}{q_n(\omega_n)} < R - \epsilon \right.\right\} = \liminf_{n\to\infty} \alpha_n(T^n(R-\epsilon)) \leqslant \liminf_{n\to\infty} \alpha_n(T^n) < 1.$$

It implies that $R - \epsilon < \overline{D}$.

Appendix B. Proof of lemma 8

In the cases $k = 2, 3$, equation (16) is checked by a calculation. Now, we prove (16) by induction in the case $k \geqslant 4$. Let a_k be the RHS of (16). The inequality $a_k \geqslant (\log k)^2$ is trivial. From the assumption of the induction, if $a_k = \sum_{i=1}^{k} p_i(\log p_i)^2$, then $p_i > 0\ (i = 1, \ldots, k)$. Using the Lagrange multiplier method, we have $(\log p_i)^2 + 2\log p_i - \lambda' = 0$, where λ' is the Lagrange multiplier. The solution is written as $\log p_i = -1 \pm \lambda$, where $\lambda := \sqrt{1+\lambda'}$. Without loss of generality, we can assume that there exists $0 \leqslant r \leqslant k$ such that

$$\log p_i = \begin{cases} -1+\lambda & \text{if} \quad r \geqslant i \\ -1-\lambda & \text{if} \quad r < i. \end{cases}$$

Since $\sum_i p_i = 1$, we have

$$1 = r\,\mathrm{e}^{-1+\lambda} + (k-r)\,\mathrm{e}^{-1-\lambda}$$

which is equivalent to the quadratic equation

$$rx^2 - \mathrm{e}x + k - r = 0$$

where $x := \mathrm{e}^{\lambda}$. Since the discriminant is greater than 0, we have

$$\mathrm{e}^2 - 4r(k-r) \geqslant 0$$

which is solved as

$$r \leqslant \frac{k - \sqrt{k^2 - \mathrm{e}^2}}{2} \qquad \frac{k + \sqrt{k^2 - \mathrm{e}^2}}{2} \leqslant r. \qquad \text{(B.1)}$$

The function $c(x) := \frac{x - \sqrt{x^2 - \mathrm{e}^2}}{2}$ is monotone decreasing in (e, ∞), and $c(4) < 1$. Thus, condition (B.1) implies that $r = 0$ or k. Thus, we have $p_i = 1/k$, i.e., (16).

References

[1] Ogawa T 2000 A study on the asymptotic property of the hypothesis testing and the channel coding in quantum mechanical systems *PhD Dissertation* University of Electro-Communications (in Japanese)

[2] Nagaoka H 2001 Strong converse theorems in quantum information theory *Proc. ERATO Workshop on Quantum Information Science 2001* p 33

[3] Hayashi M and Nagaoka H 2002 General formulas for capacity of classical-quantum channels *Preprint* quant-ph/0206186

Hayashi M and Nagaoka H 2002 A general formula for the classical capacity of a general quantum channel *Proc. 2002 IEEE Int. Symp. on Information Theory* p 71

[4] Ogawa T and Nagaoka H 2002 A new proof of the channel coding theorem via hypothesis testing in quantum information theory *Preprint* quant-ph/0208139

Ogawa T and Nagaoka H 2002 *Proc. 2002 IEEE Int. Symp. on Information Theory* p 73

[5] Rains E M 2001 A semidefinite program for distillable entanglement *IEEE Trans. Inform. Theory* **47** 2921–33 (*Preprint* quant-ph/0008047)

[6] Hayashi M 2002 Two quantum analogues of Fisher information from a large deviation viewpoint of quantum estimation *J. Phys. A: Math. Gen.* **35** 7689–727 (*Preprint* quant-ph/0202003)

[7] Hayashi M and Matsumoto K 2002 Quantum universal variable-length source coding *Phys. Rev.* A **66** 022311 (*Preprint* quant-ph/0202001)

[8] Mayers D, Salvail L and Chiba-Kohno Y 1999 Unconditionally secure quantum coin tossing *Preprint* quant-ph/9904078

[9] Hiai F and Petz D 1991 *Commun. Math. Phys.* **143** 99–114

[10] Ogawa T and Nagaoka H 2000 Strong converse and Stein's lemma in the quantum hypothesis testing *IEEE Trans. Inform. Theory* **46** 2428–33 (*Preprint* quant-ph/9906090)

[11] Ogawa T and Hayashi M 2002 On error exponents in quantum hypothesis testing *Preprint* quant-ph/0206151

[12] Han T S 1998 *Information-Spectrum Methods in Information Theory* (Tokyo: Baifukan-Press) (in Japanese), (Engl. Transl. New York: Springer)

[13] Han T S 2000 Hypothesis testing with the general source *IEEE Trans.* **46** 2415–27 (*Preprint* math.PR/0004121)

[14] Nagaoka H 1998 On asymptotic theory of quantum hypothesis testing *Proc. Symp. Statistical Inference Theory and Its Information Theoretical Aspect* pp 49–52 (in Japanese)

[15] Nagaoka H 1999 Information spectrum theory of quantum hypothesis testing *Proc. 22nd Symp. on Information Theory and Its Applications (SITA99)* pp 245–47 (in Japanese)

[16] Nagaoka H and Hayashi M 2002 An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses *Preprint* quant-ph/0206185

[17] Petz D 1994 *J. Funct. Anal.* **120** 82–97

[18] Hayashi M 2001 Asymptotics of quantum relative entropy from a representation theoretical viewpoint *J. Phys. A: Math. Gen.* **34** 3413 (*Preprint* quant-ph/9704040)

[19] Verdú S 1994 Private communication to T S Han

[20] Nagaoka H 2001 Private communication

[21] Osawa S 2001 Private communication

[22] Bhatia R 1997 *Matrix Analysis* (New York: Springer)

[23] Weyl H 1939 *The Classical Groups, Their Invariants and Representations* (Princeton, NJ: Princeton University Press)

[24] Goodman R and Wallach N 1998 *Representations and Invariants of the Classical Groups* (Cambridge: Cambridge University Press)