

## Upper bound on our knowledge about noncommuting observables for a qubit system

Takahiro Sagawa<sup>1</sup>, Yuji Kurotani<sup>2</sup>, and Masahito Ueda<sup>1,3</sup>

<sup>1</sup>*Department of Physics, The University of Tokyo,  
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan,*

<sup>2</sup>*Bridgestone corporation, and*

<sup>3</sup>*ERATO Macroscopic Quantum Control Project, JST,  
2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan.*

### Abstract

We formulate the accuracy of quantum measurement for a qubit system in terms of a 3 by 3 matrix. This matrix, which we refer to as the accuracy matrix, can be calculated from a POVM corresponding to the quantum measurement. Based on the accuracy matrix, we derive new trade-off relations between the measurement accuracy of two or three noncommuting observables of a qubit system. These trade-off relations offer a quantitative information-theoretic representation of Bohr's principle of complementarity. We also show that the accuracy matrix is closely related to the maximum-likelihood estimation and the Fisher information matrix for a finite number of samples; the accuracy matrix tells us how accurately we can estimate the probability distributions of observables of an unknown state by quantum measurement.

## I. INTRODUCTION

Accessible information about a quantum system is restricted by the noncommutability of observables. The nature of this restriction can be classified at least into two categories: fluctuations inherent in a quantum system and the error caused by the process of measurement. These two aspects of uncertainty constitute the distinctive feature of quantum mechanics.

The Kennard-Robertson uncertainty relation such as  $\Delta x \Delta p \geq \hbar/2$  describes quantum fluctuations independent of the measurement process. According to Bell's theorem, this type of quantum fluctuations prohibits us from presupposing any "element of reality" behind the probability distributions of observables. The measurement error, on the other hand, is determined by the process of measurement which is characterized by a POVM. In the idealized error-free limit, quantum measurement is described by projection operators which, however, cannot always be implemented experimentally.

The information about more than one observable can be obtained from a single POVM in simultaneous measurement of two noncommuting observables and quantum state tomography. It is known that, in simultaneous measurements, at least one of the observables cannot be measured without incurring a measurement error. Furthermore, various uncertainty relations between the measurement errors of noncommuting observables have been studied.

In our paper [1, 2], we quantify the measurement accuracy and the measurement error of observables in terms of given POVM  $\mathbf{E} = \{\hat{E}_k\}$ , by introducing  $3 \times 3$  accuracy matrix  $\chi(\mathbf{E})$  calculated from POVM  $\mathbf{E}$ . Based on this accuracy matrix, we derive the trade-off relations between the measurement accuracy of two or three observables, which can be interpreted as the uncertainty relations between the measurement errors of noncommuting observables by generalized simultaneous measurements. In the following, we review our main results without proofs.

## II. QUANTUM MEASUREMENT OF A QUBIT SYSTEM

We consider a quantum measurement described by POVM  $\mathbf{E} = \{\hat{E}_k\}$  ( $k = 1, 2, \dots, m$ ) on state  $\hat{\rho}$  of a qubit (i.e. spin-1/2) system, where  $k$  denotes the outcome of the measurement. POVM  $\mathbf{E}$  satisfies  $\sum_k \hat{E}_k = \hat{I}$ , with  $\hat{I}$  being the identity operator, and can be parameterized as

$$\hat{E}_k = r_k(\hat{I} + \mathbf{v}_k \cdot \hat{\boldsymbol{\sigma}}), \quad (1)$$

where  $\hat{\boldsymbol{\sigma}} \equiv (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$  represents the Pauli matrices. The requirements that the sum of  $\hat{E}_k$ 's equals the identity operator and that all of them be nonnegative are met if and only if

$$\sum_k r_k = 1, \quad \sum_k r_k \mathbf{v}_k = \mathbf{0}, \quad r_k > 0, \quad |\mathbf{v}_k| \leq 1 \text{ for all } k. \quad (2)$$

We can also parameterize density operator  $\hat{\rho}$  as

$$\hat{\rho} = \frac{1}{2}(\hat{I} + \mathbf{s}_0 \cdot \hat{\boldsymbol{\sigma}}), \quad (3)$$

where  $\mathbf{s}_0$  is the Bloch vector satisfying  $|\mathbf{s}_0| \leq 1$ . Conversely, for given  $\hat{\rho}$ ,  $\mathbf{s}_0$  is calculated as  $\mathbf{s}_0 = \text{tr}(\hat{\rho}\hat{\boldsymbol{\sigma}})$ . The probability of obtaining measurement outcome  $k$  is then given by

$$q_k \equiv \text{tr}(\hat{E}_k \hat{\rho}) = r_k(1 + \mathbf{v}_k \cdot \mathbf{s}_0). \quad (4)$$

Any observable  $\hat{O}$  of the qubit system can be diagonalized as

$$\hat{O} = \lambda_+ \hat{P}(+; \mathbf{n}) + \lambda_- \hat{P}(-; \mathbf{n}), \quad (5)$$

where  $\lambda_+$  and  $\lambda_-$  are the corresponding eigenvalues and  $\hat{P}(+; \mathbf{n})$  and  $\hat{P}(-; \mathbf{n})$  are projection operators with  $\mathbf{n}$  being a three-dimensional unit vector, and

$$\hat{P}(\pm; \mathbf{n}) = \frac{1}{2}(\hat{I} \pm \mathbf{n} \cdot \hat{\boldsymbol{\sigma}}). \quad (6)$$

The probability distribution of observable  $\hat{O}$  is then given by

$$p(\pm; \mathbf{n}) \equiv \text{tr}(\hat{P}(\pm; \mathbf{n}) \hat{\rho}) = \frac{1}{2}(1 \pm \mathbf{n} \cdot \mathbf{s}_0). \quad (7)$$

If we are not interested in eigenvalues of the observables but are only concerned with the directions ( $\pm$ ) of the outcome, we can replace  $\hat{O}$  with  $\mathbf{n} \cdot \hat{\boldsymbol{\sigma}}$  by setting  $\lambda_{\pm} = \pm 1$ . In the following analysis, we identify observable  $\lambda_+ \hat{P}(+; \mathbf{n}) + \lambda_- \hat{P}(-; \mathbf{n})$  with observable  $\mathbf{n} \cdot \hat{\boldsymbol{\sigma}}$ , and refer to the probability distribution in Eq. (7) as that in the direction of  $\mathbf{n}$ .

### III. ACCURACY MATRIX

#### A. Definition of the Accuracy Matrix

We will characterize the accuracy of arbitrary observable in such a manner that it depends only on the process of measurement and not on measured state  $\hat{\rho}$ . We first define the accuracy matrix:

**Definition 1** (*Accuracy matrix*) The  $3 \times 3$  accuracy matrix  $\chi(\mathbf{E})$  characterizes the measurement accuracy of observables in terms of POVM  $\mathbf{E}$ , and is defined as

$$\chi(\mathbf{E})_{ij} \equiv \sum_k r_k (\mathbf{v}_k)_i (\mathbf{v}_k)_j, \quad (8)$$

where  $(\mathbf{v}_k)_i$  denotes the  $i$ th component of real vector  $\mathbf{v}_k$  and  $ij$  shows indexes of matrix elements of  $\chi(\mathbf{E})$ . We introduce notation  $\mathbf{v}\mathbf{v}^T$  with  $\mathbf{v} \in \mathbf{R}^3$  as

$$(\mathbf{v}\mathbf{v}^T)_{ij} \equiv (\mathbf{v})_i (\mathbf{v})_j; \quad (9)$$

that is,  $\mathbf{v}\mathbf{v}^T$  denotes the projection matrix onto direction  $\mathbf{v}$  in  $\mathbf{R}^3$  whose  $ij$  matrix element is given by  $(\mathbf{v})_i (\mathbf{v})_j$ . We can then rewrite (8) in matrix form as

$$\chi(\mathbf{E}) \equiv \sum_k r_k \mathbf{v}_k \mathbf{v}_k^T. \quad (10)$$

Note that  $\chi(\mathbf{E})$  is positive semidefinite and Hermitian, and can therefore be diagonalized by an orthonormal transformation.

The physical meaning and useful properties of the accuracy matrix will be investigated subsequently, and its foundation from an information-theoretic point of view will be established in terms of the maximum-likelihood estimation of the probability distribution of observables in Sec. VI. In fact, the accuracy matrix is closely related to Fisher information matrix (41) or (42), although physical quantities such as the measurement error can be directly derived from the accuracy matrix without resort to Fisher information.

Noting that  $\sum_k r_k |\mathbf{v}_k|^2 \leq \sum_k r_k = 1$ , we can obtain the following fundamental inequality which forms the basis of trade-off relations to be discussed later.

**Theorem 1** Three eigenvalues  $\{\chi_1, \chi_2, \chi_3\}$  of  $\chi(\mathbf{E})$  satisfy

$$\chi_1 + \chi_2 + \chi_3 \leq 1, \quad (11)$$

or equivalently,

$$\text{Sp}(\chi(\mathbf{E})) \leq 1, \quad (12)$$

where we denote the trace of  $3 \times 3$  matrix as  $\text{Sp}(\dots)$  to reserve symbol  $\text{tr}(\dots)$  for the trace of a quantum-mechanical  $2 \times 2$  matrix. The equality  $\chi_1 + \chi_2 + \chi_3 = 1$ , or  $\text{Sp}(\chi) = 1$ , holds if and only if  $|\mathbf{v}_k| = 1$  for all  $k$ .

The following corollary follows from the positivity of  $\chi(\mathbf{E})$ .

**Corollary 1** The accuracy matrix satisfies the following matrix inequality:

$$0 \leq \chi(\mathbf{E}) \leq I_3, \quad (13)$$

where  $I_3$  is the  $3 \times 3$  identity matrix, and  $\chi(\mathbf{E}) \leq I_3$  means that all eigenvalues of  $I_3 - \chi(\mathbf{E})$  are non-negative.

The following two examples illustrate the physical meaning of the accuracy matrix.

**Example 1** (*Nonideal measurement*) We consider POVM  $\mathbf{E}$  consisting of two positive operators which can be parameterized as

$$\hat{E}(+; \mathbf{n}) = r(\hat{I} + \varepsilon_1 \mathbf{n} \cdot \hat{\sigma}), \quad \hat{E}(-; \mathbf{n}) = (1 - r)(\hat{I} - \varepsilon_2 \mathbf{n} \cdot \hat{\sigma}), \quad (14)$$

where  $\mathbf{n}$  is a unit vector,  $r\varepsilon_1 - (1 - r)\varepsilon_2 = 0$ ,  $0 < r < 1$ ,  $-1 \leq \varepsilon_1 \leq 1$ , and  $-1 \leq \varepsilon_2 \leq 1$ . This POVM corresponds to a nonideal measurement of observable  $\mathbf{n} \cdot \hat{\sigma}$ . It can be reduced to projection measurement  $\{\hat{P}(+; \mathbf{n}), \hat{P}(-; \mathbf{n})\}$  if and only if  $\varepsilon_1 = \varepsilon_2 = 1$  and  $r = 1/2$ . On the other hand, the POVM is trivial (i.e.,  $\hat{E}_+ = r\hat{I}$  and  $\hat{E}_- = (1 - r)\hat{I}$ ) if and only if  $\varepsilon_1 = \varepsilon_2 = 0$ ; we cannot then obtain any information about  $\hat{\rho}$ . Equation (14) can be rewritten as

$$\begin{pmatrix} \hat{E}_+ \\ \hat{E}_- \end{pmatrix} = F \begin{pmatrix} \hat{P}(+; \mathbf{n}) \\ \hat{P}(-; \mathbf{n}) \end{pmatrix}, \quad (15)$$

where  $F$  is the  $2 \times 2$  transition-probability matrix

$$F = \begin{pmatrix} r(1 + \varepsilon_1) & r(1 - \varepsilon_1) \\ (1 - r)(1 - \varepsilon_2) & (1 - r)(1 + \varepsilon_2) \end{pmatrix}, \quad (16)$$

satisfying  $\sum_i F_{ij} = 1$  and  $0 \leq |\det F|^2 \leq 1$ . Note that  $F$  describes a binary symmetric channel if and only if  $r = 1/2$  and  $\varepsilon_1 = \varepsilon_2$ . It follows from Eq. (15) that any measurement

process described by a POVM consisting of two positive operators is formally equivalent to a measurement process in which a classical error is added to the projection measurement. The physical origin of this error, however, lies in the quantum-mechanical interaction. We can rewrite Eq.(14) as

$$\hat{E}_1 = r(\hat{I} + \mathbf{v}_1 \cdot \hat{\boldsymbol{\sigma}}), \quad \hat{E}_2 = (1 - r)(\hat{I} + \mathbf{v}_2 \cdot \hat{\boldsymbol{\sigma}}), \quad (17)$$

where  $\mathbf{v}_1 = \varepsilon_1 \mathbf{n}$  and  $\mathbf{v}_2 = -\varepsilon_2 \mathbf{n}$ . The accuracy matrix can then be represented by

$$\chi(\mathbf{E}) \equiv r\mathbf{v}_1\mathbf{v}_1^T + (1 - r)\mathbf{v}_2\mathbf{v}_2^T = \chi_{11}\mathbf{n}\mathbf{n}^T, \quad (18)$$

where  $\chi_{11}$  is the eigenvalue of  $\chi$  corresponding to the eigenvector  $\mathbf{n}$ , and is given by

$$\chi_{11} = r|\mathbf{v}_1|^2 + (1 - r)|\mathbf{v}_2|^2. \quad (19)$$

We can also write  $\chi_{11}$  in terms of the transition-probability matrix as

$$\chi_{11} = \frac{|\det F|^2}{4r} + \frac{|\det F|^2}{4(1 - r)} = \frac{|\det F|^2}{4r(1 - r)}. \quad (20)$$

Accuracy parameter  $\chi_{11}$  satisfies

$$0 \leq \chi_{11} \leq 1, \quad (21)$$

where  $\chi_{11} = 1$  holds if and only if  $|\mathbf{v}_1| = |\mathbf{v}_2| = 1$  and  $r = 1/2$ ; that is,  $\mathbf{E}$  describes the projection measurement of observable  $\mathbf{n} \cdot \hat{\boldsymbol{\sigma}}$ . Note that  $\chi(\mathbf{E}) = \mathbf{n}\mathbf{n}^T$  holds in this case. On the other hand,  $\chi_{11} = 0$  holds if and only if  $|\mathbf{v}_1| = |\mathbf{v}_2| = 0$ . In this case,  $\chi(\mathbf{E}) = O$  holds, and we cannot obtain any information about  $\hat{\rho}$ . Nonzero eigenvalue  $\chi_{11}$  thus characterizes the measurement accuracy of  $\mathbf{n} \cdot \hat{\boldsymbol{\sigma}}$ ; the larger  $\chi_{11}$ , the more information we can extract about  $\mathbf{n} \cdot \hat{\boldsymbol{\sigma}}$  from the measurement outcome. These properties can be generalized for an arbitrary POVM as shown below.

**Example 2** (*Probabilistic measurement*) Suppose that a nonideal measurement of  $\hat{A} = \mathbf{n}_A \cdot \hat{\boldsymbol{\sigma}}$  is performed with probability  $\xi$  ( $0 < \xi < 1$ ) and  $\hat{B} = \mathbf{n}_B \cdot \hat{\boldsymbol{\sigma}}$  is performed with probability  $1 - \xi$ . The POVM corresponding to this probabilistic measurement consists of four operators:

$$\mathbf{E} = \{\xi\hat{E}(\pm; \mathbf{n}_A), (1 - \xi)\hat{E}(\pm; \mathbf{n}_B)\}. \quad (22)$$

As the number of measured samples increases, this measurement asymptotically approaches the measurements on  $N$  identically prepared samples which are divided into

two groups in the ratio  $\xi : 1 - \xi$ , with  $\hat{A}$  being measured for the first group and  $\hat{B}$  for the second group. Suppose that the accuracy matrix of the measurement of  $\hat{A}$  is given by  $\chi_A \mathbf{n}_A \mathbf{n}_A^T$  and that of  $\hat{B}$  is given by  $\chi_B \mathbf{n}_B \mathbf{n}_B^T$ . The accuracy matrix of the probabilistic measurement is given by

$$\chi(\mathbf{E}) = \xi \chi_A \mathbf{n}_A \mathbf{n}_A^T + (1 - \xi) \chi_B \mathbf{n}_B \mathbf{n}_B^T. \quad (23)$$

This representation suggests that the measurement accuracy concerning  $\hat{A}$  is degraded by a factor of  $\xi$  compared with the single nonideal measurement of  $\hat{A}$ , because we cannot observe  $\hat{A}$  with probability  $1 - \xi$ . A similar discussion applies to  $\hat{B}$  as well.

Equation (23) shows that  $\chi(\mathbf{E})$  is the convex combination of the accuracy matrix of the POVMs measuring  $\hat{A}$  and  $\hat{B}$ , where coefficients  $\xi$  and  $1 - \xi$  give the probabilities of measuring  $\hat{A}$  and  $\hat{B}$ , respectively.

This relationship can be generalized as follows. Let us consider three POVMs:  $\mathbf{E}' = \{\hat{E}_1, \hat{E}_2, \dots, \hat{E}_m\}$ ,  $\mathbf{E}'' = \{\hat{E}_{m+1}, \hat{E}_{m+2}, \dots, \hat{E}_n\}$ , and  $\mathbf{E} = \{\xi \hat{E}_1, \dots, \xi \hat{E}_m, (1 - \xi) \hat{E}_{m+1}, \dots, (1 - \xi) \hat{E}_n\}$  with  $0 < \xi < 1$ . POVM  $\mathbf{E}$  describes the probabilistic measurement of  $\mathbf{E}'$  with probability  $\xi$  and of  $\mathbf{E}''$  with probability  $1 - \xi$ . According to the definition of the accuracy matrix, we obtain the following theorem.

**Theorem 2** (*Linearity*)

$$\chi(\mathbf{E}) = \xi \chi(\mathbf{E}') + (1 - \xi) \chi(\mathbf{E}''), \quad (24)$$

or more symbolically,

$$\chi(\xi \mathbf{E}' + (1 - \xi) \mathbf{E}'') = \xi \chi(\mathbf{E}') + (1 - \xi) \chi(\mathbf{E}''). \quad (25)$$

## B. Accuracy Parameter in a Specific Direction

We next parameterize the measurement accuracy of a particular observable. We introduce the subspace  $V(\mathbf{E})$  of  $\mathbf{R}^3$  spanned by the eigenvectors of  $\chi(\mathbf{E})$  corresponding to nonzero eigenvalues.

**Definition 2** (*Measurement Accuracy*) Accuracy parameter  $\chi(\mathbf{n}; \mathbf{E})$  in direction  $\mathbf{n} \in V(\mathbf{E})$  is defined as

$$\chi(\mathbf{n}; \mathbf{E}) \equiv \frac{1}{\mathbf{n} \cdot (\chi(\mathbf{E})^{-1}) \mathbf{n}}, \quad (26)$$

where  $\chi(\mathbf{E})^{-1}$  is assumed to act only on subspace  $V(\mathbf{E})$ . If  $\mathbf{n} \notin V(\mathbf{E})$ , we set  $\chi(\mathbf{n}; \mathbf{E}) \equiv 0$ .

This definition is closely related to the Fisher information of a particular direction in Eq. (46) of Sec. VI. The inverse of  $\chi(\mathbf{n}; \mathbf{E})$  provides a measure of the measurement error.

**Definition 3** (*Measurement Error*) The error parameter of the measurement in direction  $\mathbf{n}$  is defined as

$$\varepsilon(\mathbf{n}; \mathbf{E}) \equiv \frac{1}{\chi(\mathbf{n}; \mathbf{E})} - 1 = \mathbf{n} \cdot (\chi(\mathbf{E})^{-1})\mathbf{n} - 1. \quad (27)$$

These parameters satisfy the following inequalities.

**Theorem 3**

$$0 \leq \chi(\mathbf{n}; \mathbf{E}) \leq 1, \quad (28)$$

$$0 \leq \varepsilon(\mathbf{n}; \mathbf{E}) \leq \infty. \quad (29)$$

Equality  $\chi(\mathbf{n}; \mathbf{E}) = 1$ , or equivalently  $\varepsilon(\mathbf{n}; \mathbf{E}) = 0$ , holds if and only if the measurement described by  $\mathbf{E}$  is equivalent to a projection measurement in direction  $\mathbf{n}$ . In this case, the measurement involves no measurement error. The other limit of  $\chi(\mathbf{n}; \mathbf{E}) = 0$ , or equivalently  $\varepsilon(\mathbf{n}; \mathbf{E}) = \infty$ , holds if and only if  $\mathbf{n} \notin V(\mathbf{E})$ . In this case, we cannot obtain any information about direction  $\mathbf{n}$  from the measurement.

Let  $\mathbf{n}_1$ ,  $\mathbf{n}_2$ , and  $\mathbf{n}_3$  be the eigenvectors of  $\chi(\mathbf{E})$ , and  $\chi_1$ ,  $\chi_2$ , and  $\chi_3$  be the corresponding eigenvalues. It can be shown that

$$\chi(\mathbf{n}_i; \mathbf{E}) = \chi_i, \quad (i = 1, 2, 3). \quad (30)$$

According to Theorem 1, we cannot simultaneously measure the three directions corresponding to the eigenvectors with the maximum accuracy  $\chi_i = 1$  for all  $i$ . This trade-off relation represents the uncertainty relation between the measurement errors.

The nonideal measurement of  $\hat{A} = \mathbf{n}_A \cdot \hat{\sigma}$  with POVM  $\mathbf{E}_A$  is characterized with accuracy matrix  $\chi_A \mathbf{n}_A \mathbf{n}_A^T$ . In this case, we can show that  $V(\mathbf{E}_A) = \{\mathbf{a} \mid \mathbf{a} = \lambda \mathbf{n}_A, \lambda \in \mathbf{R}\}$ . It follows that  $\chi(\pm \mathbf{n}_A; \mathbf{E}) = \chi_A$  and  $\chi(\mathbf{n}; \mathbf{E}) = 0$  for  $\mathbf{n} \neq \pm \mathbf{n}_A$ .

The probabilistic measurement of  $\mathbf{E}_A$  and  $\mathbf{E}_B$  in characterized with accuracy matrix of the joint POVM  $\mathbf{E}$  given in (23), so  $V(\mathbf{E})$  contains the subspace spanned by  $\mathbf{n}_A$  and  $\mathbf{n}_B$ . A straightforward calculation shows that

$$\chi(\mathbf{n}_A; \mathbf{E}) = \xi \chi_A, \quad \chi(\mathbf{n}_B; \mathbf{E}) = (1 - \xi) \chi_B. \quad (31)$$



### C. Reconstructive subspace

We next introduce the concept of “reconstructive subspace” and “reconstructive direction”. We begin by the following theorem.

**Theorem 4**  $V(\mathbf{E})$  corresponds to the subspace spanned by the set of basis vectors  $\{\mathbf{v}_k\}$  of accuracy matrix (10).

Suppose that we perform measurement  $\{\hat{E}_k\}$  and obtain probability distribution  $\{q_k\}$  for each outcome  $k$ . Can we then reconstruct pre-measurement distribution  $\{p(\mathbf{n}; \pm)\}$  of the system from  $\{q_k\}$ ? The answer is given by the following theorem.

**Theorem 5** (*Reconstructive subspace and reconstructive direction*) We can reconstruct probability distribution  $\{p(\pm; \mathbf{n})\}$  from measured distribution  $\{q_k\}$ , if and only if  $\mathbf{n} \in V(\mathbf{E})$ . We thus refer to  $V(\mathbf{E})$  as a reconstructive subspace, and to a unit vector in  $V(\mathbf{E})$  as a reconstructive direction.

In the case of the nonideal measurement in Example 1, the reconstructive subspace is given by  $V(\mathbf{E}) = \{\lambda \mathbf{n} \mid \lambda \in \mathbf{R}\}$ . On the other hand, in the case of the probabilistic measurement in Example 2, the reconstructive subspace is two-dimensional:  $V(\mathbf{E}) = \{\lambda_A \mathbf{n}_A + \lambda_B \mathbf{n}_B \mid (\lambda_A, \lambda_B) \in \mathbf{R}^2\}$ .

## IV. TRADE-OFF RELATIONS FOR GENERALIZED SIMULTANEOUS MEASUREMENT OF A QUBIT SYSTEM

We now derive the general trade-off relations between the measurement errors of non-commuting observables, which are the main results of this paper.

Let  $\mathbf{n}_1$ ,  $\mathbf{n}_2$ , and  $\mathbf{n}_3$  be the respective eigenvectors of  $\chi(\mathbf{E})$  corresponding to the eigenvalues  $\chi_1$ ,  $\chi_2$ , and  $\chi_3$ , where  $\chi_i = \chi(\mathbf{n}_i; \mathbf{E})$  ( $i = 1, 2, 3$ ). We define the error parameters as  $\varepsilon_i \equiv \varepsilon(\mathbf{n}_i; \mathbf{E}) = (\chi_i)^{-1} - 1$ . Inequality (11) or (12) in Theorem 1 can be rewritten in terms of the error parameters as

$$\varepsilon_1 \varepsilon_2 \varepsilon_3 \geq \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + 2. \quad (32)$$

Considering two eigenvalues alone (i.e.,  $\chi_1 + \chi_2 \leq 1$ ), we can simplify the trade-off relation:

$$\varepsilon_1 \varepsilon_2 \geq 1. \quad (33)$$

The trade-off relations (32) and (33) can be generalized to the case of arbitrary directions. We first consider the case of two observables.

**Theorem 8** (*Trade-off relation*) We consider a simultaneous measurement in two directions  $\mathbf{n}_A$  and  $\mathbf{n}_B$  ( $\mathbf{n}_A \cdot \mathbf{n}_B = \cos \theta$ ) described by POVM  $\mathbf{E}$ . We assume  $\mathbf{n}_A \in V(\mathbf{E})$  and  $\mathbf{n}_B \in V(\mathbf{E})$ , and define  $\varepsilon_\alpha \equiv \varepsilon(\mathbf{n}_\alpha; \mathbf{E})$  and  $\chi_\alpha \equiv \chi(\mathbf{n}_\alpha; \mathbf{E})$  ( $\alpha = A, B$ ). Then trade-off relation

$$\varepsilon_A \varepsilon_B \geq \sin^2 \theta, \quad (34)$$

or equivalently,

$$\chi_A + \chi_B - \chi_A \chi_B \cos^2 \theta \leq 1 \quad (35)$$

holds. The equality in (34) or (35) holds if and only if  $\varepsilon_1 \cos \theta_A \cos \theta_B + \varepsilon_2 \sin \theta_A \sin \theta_B = 0$ . In the case of  $\varepsilon_A = \varepsilon_B$  (i.e., the measurement errors are symmetric), the equality holds if and only if  $\sin(\theta_A + \theta_B) \sin(\theta_A - \theta_B) \cos(\theta_A + \theta_B) = 0$ .

The accessible regime for  $\chi_A$  and  $\chi_B$  is illustrated in FIG.1 for the case of  $\theta = \pi/2$ ,  $\theta = \pi/6$ , and  $\theta = 0$ .

The trade-off relations (34) and (35) can be interpreted as the uncertainty relations between measurement errors. They offer a rigorous representation of Bohr's principle of complementarity which dictates "the mutual exclusion of any two experimental procedures" to measure two noncommuting observables.

The trade-off relation between three observables can be formulated as follows:

**Theorem 9** We consider a simultaneous measurement in three directions  $\mathbf{n}_A$ ,  $\mathbf{n}_B$ , and  $\mathbf{n}_C$  described by POVM  $\mathbf{E}$ . Let us assume that  $\mathbf{n}_A$ ,  $\mathbf{n}_B$ , and  $\mathbf{n}_C$  are linearly independent. We set the notation  $\varepsilon_\alpha \equiv \varepsilon(\mathbf{n}_\alpha; \mathbf{E})$  and  $\chi_\alpha \equiv \chi(\mathbf{n}_\alpha; \mathbf{E})$ , where  $\alpha = A, B, C$ . Then inequality

$$\varepsilon_A \varepsilon_B \varepsilon_C \geq 8 \{ \mathbf{n}_A \cdot (\mathbf{n}_B \times \mathbf{n}_C) \}^2 \quad (36)$$

holds. The equality in (36) holds if and only if  $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 2$  and  $\{ \mathbf{n}_\alpha \}$  are orthogonal.

## V. MAXIMUM-LIKELIHOOD ESTIMATION

In this section, we point out the close connection between the accuracy matrix and the Fisher information.

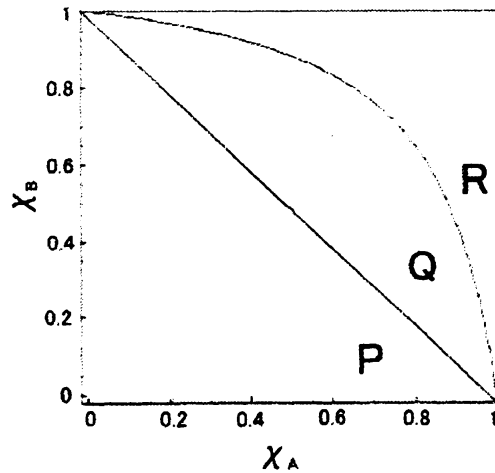


FIG. 1: Trade-off relation for the accuracy of noncommuting observables. P indicates the regimes satisfying the inequality for the case of  $\theta = \pi/2$ , P and Q indicate the regimes satisfying inequality (35) for the case of  $\theta = \pi/6$ , and P, Q and R indicate the regimes satisfying the inequality for the case of  $\theta = 0$ . We can only access regime Q through simultaneous measurement for the case of  $\theta = \pi/6$ .

We first formulate quantum measurements in terms of the maximum-likelihood estimation. We consider the quantum measurements described by POVM  $\mathbf{E} = \{\hat{E}_k\}$  for each of  $N$  ( $< \infty$ ) samples prepared in the same unknown state  $\hat{\rho}$ . Note that  $\hat{E}_k = r_k(\hat{I} + \mathbf{v}_k \cdot \hat{\sigma})$ . Our task is to estimate the Bloch vector  $\mathbf{s}_0$  by maximum-likelihood estimation. Suppose that we obtain outcome “ $k$ ”  $N_k$  times. The likelihood function then becomes

$$L(\mathbf{s}) \equiv \sum_k N_k \ln f_k(\mathbf{s}), \quad (37)$$

where  $f_k(\mathbf{s}) \equiv r_k(1 + \mathbf{v}_k \cdot \mathbf{s})$ . We denote  $\mathbf{s}^*$  as the maximum-likelihood estimator of  $\mathbf{s}_0$  from  $N$  measurement outcomes;  $L(\mathbf{s})$  takes the maximum value with  $\mathbf{s} = \mathbf{s}^*$  under condition  $|\mathbf{s}| \leq 1$ .

If we can determine unique estimator  $\mathbf{s}^*$ , then we can calculate the maximum-likelihood estimator of  $p(\pm; \mathbf{n})$  for arbitrary direction  $\mathbf{n}$ :

$$p(\pm; \mathbf{n})^* = \frac{1}{2}(1 \pm \mathbf{n} \cdot \mathbf{s}^*). \quad (38)$$

In general, however, we cannot uniquely determine estimator  $\mathbf{s}^*$  from measurement outcomes;  $L(\mathbf{s}^*) = L(\mathbf{s}^* + \mathbf{a})$  holds for all  $\mathbf{a} \in V(\mathbf{E})^\perp$ , where  $V(\mathbf{E})^\perp$  is the orthogonal complement of  $V(\mathbf{E})$ , because  $\mathbf{v}_k \cdot \mathbf{s}^* = \mathbf{v}_k \cdot (\mathbf{s}^* + \mathbf{a})$  holds for all  $k$ . Corresponding to this indeterminacy, we cannot estimate the probability distribution of all observables  $\mathbf{n} \cdot \hat{\sigma}$  in general. In fact, we can uniquely estimate the probability distribution of the observable  $\mathbf{n} \cdot \hat{\sigma}$  only if  $\mathbf{n} \in V(\mathbf{E})$ , because  $\mathbf{n} \cdot \mathbf{s}^* = \mathbf{n} \cdot (\mathbf{s}^* + \mathbf{a})$  holds for all  $\mathbf{a} \in V(\mathbf{E})^\perp$  only if  $\mathbf{n} \in V(\mathbf{E})$ . In other words, we can uniquely estimate  $p(\pm; \mathbf{n})^*$  only if  $\mathbf{n} \in V(\mathbf{E})$ .

We next discuss the asymptotic behavior of the maximum-likelihood estimation.

In general, it is known that maximum-likelihood estimation is asymptotically unbiased; in the present context, it holds that

$$\lim_{N \rightarrow \infty} \mathbf{s}^* = \mathbf{s}_0, \quad (39)$$

in the case of  $V(\mathbf{E}) = \mathbf{R}^3$ ; we can precisely evaluate  $\mathbf{s}_0$  if an infinite number of samples are available.

In the case of  $V(\mathbf{E}) \neq \mathbf{R}^3$ , however, we cannot determine component of  $V(\mathbf{E})^\perp$  of  $\mathbf{s}_0$ . In other words, we can uniquely determine  $\mathbf{n} \cdot \mathbf{s}_0$  only if  $\mathbf{n} \in V(\mathbf{E})$ . We can thus generalize Eq. (39) for the case that  $V(\mathbf{E}) \neq \mathbf{R}^3$ :

$$\lim_{N \rightarrow \infty} \mathbf{n} \cdot \mathbf{s}^* = \mathbf{n} \cdot \mathbf{s}_0, \quad \forall \mathbf{n} \in V(\mathbf{E}). \quad (40)$$

Theorem 5 in Sec. III has revealed the necessary and sufficient condition for direction  $\mathbf{n}$  about which we can reconstruct the true probability distribution  $p(\pm; \mathbf{n})$  from the measurement outcomes in the limit of  $N \rightarrow \infty$ .

The asymptotic accuracy of maximum-likelihood estimation is characterized by the Fisher information. The Fisher information takes the matrix form in our situation, given by

$$I_{ij} \equiv - \sum_k q_k \frac{\partial^2 \ln f_k(\mathbf{s})}{\partial(\mathbf{s})_i \partial(\mathbf{s})_j} \Big|_{\mathbf{s}=\mathbf{s}_0} = \sum_k \frac{r_k^2}{q_k} (\mathbf{v}_k)_i (\mathbf{v}_k)_j, \quad (41)$$

or equivalently,

$$I = \sum_k \frac{r_k^2}{q_k} \mathbf{v}_k \mathbf{v}_k^\top. \quad (42)$$

Note that  $I$  is a  $3 \times 3$  positive and Hermite matrix, so it can be diagonalized by orthonormal eigenvectors.

It can be shown in the same manner as Theorem 4 that  $V(\mathbf{E})$  corresponds to the subspace of  $\mathbf{R}^3$  spanned by the eigenvectors of  $I$  corresponding to nonzero eigenvalues; the Fisher information matrix has the information about the reconstructive subspace.

A remarkable property of the Fisher information is that it characterizes the asymptotic variance of the estimator  $\mathbf{s}^*$ . To see this, we introduce the  $3 \times 3$  correlation matrix of the probability variable  $\mathbf{s}^*$  given by

$$T_{ij} \equiv \langle \{(\mathbf{s}^*)_i - (\mathbf{s}_0)_i\} \{(\mathbf{s}^*)_j - (\mathbf{s}_0)_j\} \rangle, \quad (43)$$

where  $\langle \dots \rangle$  describes the statistical average. Note that  $\langle \mathbf{s}^* \rangle = \mathbf{s}_0$ . Let  $T(\mathbf{n})$  be the variance of  $\mathbf{n} \cdot \mathbf{s}^*$ :  $T(\mathbf{n}) = \langle (\mathbf{n} \cdot \mathbf{s}^*)^2 \rangle - (\mathbf{n} \cdot \mathbf{s}_0)^2$ . It can be easily shown that

$$T(\mathbf{n}) = \mathbf{n} \cdot T \mathbf{n}. \quad (44)$$

We first consider the case of  $V(\mathbf{E}) = \mathbf{R}^3$ . In this case, it is well-known that the probability distribution of  $\mathbf{s}^*$  approaches the normal distribution with average  $\mathbf{s}_0$  and correlation matrix  $T = (IN)^{-1}$  as the number of samples  $N$  increases. This theorem does not hold if  $\hat{\rho}$  is a pure state (e.g.  $|\mathbf{s}_0| = 1$ ) which, however, does not affect the following discussions. Focusing on particular direction  $\mathbf{n}$ , we can reduce the foregoing theorem to the following form: the distribution of  $\mathbf{n} \cdot \mathbf{s}^*$  approaches the normal distribution with average  $\mathbf{n} \cdot \mathbf{s}_0$  and variance

$$T(\mathbf{n}) = \frac{1}{I(\mathbf{n})N}, \quad (45)$$

as the number of samples  $N$  increases. Here,

$$I(\mathbf{n}) \equiv \frac{1}{\mathbf{n} \cdot I^{-1} \mathbf{n}} \quad (46)$$

is the Fisher information in direction  $\mathbf{n}$ .

We next consider the case of  $V(\mathbf{E}) \neq \mathbf{R}^3$  and  $V(\mathbf{E}) \neq \{0\}$ . We can meaningfully consider  $I^{-1}$  even in this case by restricting the domain of  $I$  to  $V(\mathbf{E})$ ;  $\mathbf{n} \cdot I^{-1} \mathbf{n}$  is well-defined. We can thus generalize the above theorem: the distribution of  $\mathbf{n} \cdot \mathbf{s}^*$  approaches the normal distribution with average  $\mathbf{n} \cdot \mathbf{s}_0$  and variance  $T(\mathbf{n}) = 1/I(\mathbf{n})N$  as the number of samples  $N$  increases.

The greater the Fisher information, the more information we can extract from the measurement outcome. In the case of  $I(\mathbf{n}) = 0$ , the variance of estimator  $p(\pm; \mathbf{n})$  diverges,

so we cannot gain any information about the probability distribution in direction  $\mathbf{n}$ . This is the case of  $\mathbf{n}$  not being in any reconstructive direction.

Replacing  $q_k$  by  $r_k$  in the Fisher information (41) or (42), we can obtain the accuracy matrix in Eq. (8) or (10). Note that  $r_k$  is the average of  $q_k$  over the entire Bloch sphere. We can thus say that the accuracy matrix characterizes a certain average of the asymptotic accuracy of maximum-likelihood estimation. Although the Fisher information matrix does not characterize the asymptotic accuracy when  $\hat{\rho}$  is a pure state, the average is meaningful because the Lebesgue measure of the pure states is zero. The trade-off relations (34), (35), and (36) can thus be interpreted as the trade-off relations between the asymptotic accuracy of maximum-likelihood estimation of the probability distributions of observables. A finite number of samples only give us an imperfect information about the probability distribution of an observable for an unknown state. As we have shown, this imperfection further deteriorates in the case of simultaneous estimation due to the noncommutability of the observables.

## VI. CONCLUSION

We have considered the accuracy matrix of the most general class of measurements of a qubit system. Our main result is trade-off relations (34), (35), and (36) between the accuracy parameters and the error parameters. We can interpret them as the uncertainty relations between measurement errors in generalized simultaneous measurements; the more information we obtain about an observable, the less information we can obtain about another noncommuting observable. We have also pointed out the close relationship between the accuracy matrix and the Fisher information; the accuracy matrix is an average of the Fisher information matrix over the measured states. We have shown that the trade-off relations can be interpreted as being between the accuracy of the maximum-likelihood estimators of the probability distributions of noncommuting observables.

- 
- [1] Y. Kurotani, T. Sagawa, and M. Ueda, Phys. Rev. A **76**, 022325 (2007).
  - [2] T. Sagawa and M. Ueda, Phys. Rev. A **77**, 012313 (2008).