# Optimal estimation of a physical observable's expectation value for pure states

PHYSICAL REVIEW A **73**, 062322 (2006)

# Optimal estimation of a physical observable's expectation value for pure states

A. Hayashi, M. Horibe, and T. Hashimoto
*Department of Applied Physics, University of Fukui, Fukui 910-8507, Japan*
(Received 5 December 2005; published 14 June 2006)

We study the optimal way to estimate the quantum expectation value of a physical observable when a finite number of copies of a quantum pure state are presented. The optimal estimation is determined by minimizing the squared error averaged over all pure states distributed in a unitary invariant way. We find that the optimal estimation is "biased" though the optimal measurement is given by successive projective measurements of the observable. The optimal estimate is not the sample average of observed data, but the arithmetic average of observed and "default nonobserved" data, with the latter consisting of all eigenvalues of the observable.

DOI: 10.1103/PhysRevA.73.062322 PACS number(s): 03.67.Hk

## I. INTRODUCTION

One of the fundamental tasks in quantum physics is to determine the expectation value of a physical observable of an unknown quantum state. With only a single copy of the quantum state given, we cannot determine the expectation value of a physical observable because of the statistical nature of quantum measurement. Suppose we are presented with a certain finite number $N$ of copies of an unknown quantum state. We cannot increase the number of the copies since the no-cloning theorem [1] forbids it. Then what is the optimal way to determine the expectation value of the observable for a given $N$? An intuitively plausible optimal estimate is given by the arithmetic average of the data produced by successive projective measurements of the observable on the individual systems.

This problem, however, is by no means trivial. Given a quantum system composed of subsystems, we can consider two types of measurement. One is separate measurements: a sequence of measurements on the individual subsystems, possibly dependent on the outcomes of earlier measurements. The other is joint measurement: a single measurement on the system as a whole. Recent studies on quantum-state discrimination and estimation [2,3] provide considerable instances in which joint measurements perform better than separate measurements even for a state composed of mutually uncorrelated subsystems.

Peres and Wootters showed that a certain set of three bipartite product states can be better distinguished by a joint measurement [4,5] (see also [6–8]). An even stronger example was provided by Bennett *et al.* [9], which shows that a certain orthogonal set of bipartite product states cannot be reliably distinguished by any separate measurement though a joint measurement perfectly distinguishes them because of their mutual orthogonality. The superiority of joint measurement has also been discussed in the problem of quantum-state estimation for identically prepared copies of an unknown state (see [10–14,16], for example).

In [15], D'Ariano, Giovannetti, and Perinotti raised the question of whether the standard procedure of averaging the outcomes of repeated measurements of an observable over equally prepared systems is the best way of estimating the expectation value of the observable, or whether a joint measurement can improve the estimation. They showed that the standard procedure is indeed optimal if one is restricted to the class of unbiased estimation for any generally mixed state. Here an estimator is said to be unbiased if the average over many independent estimates gives the true value to be estimated.

An unbiased result is certainly one of the desirable properties for estimation but not a necessary condition. A natural question is then whether a "biased" estimation performs better than the standard unbiased estimation. Let us take a simple example, in which we estimate the expectation value of the observable $\sigma_z$ for a single-qubit system in an unknown pure state. We assume that the state of the qubit is chosen according to the uniform distribution on the Bloch sphere. Suppose that the projective measurement of $\sigma_z$ produced the outcome 1, which means the sample average is 1. Now, one can ask if it is reasonable to conclude that the expectation value of $\sigma_z$ is most likely equal to 1. Note that the expectation value of $\sigma_z$ is 1 only if the qubit lies exactly at the north pole of the Bloch sphere. On the other hand, the measurement of $\sigma_z$ can produce the outcome 1 with some probability unless the qubit is exactly at the south pole. Therefore, it is more reasonable to consider that the expectation value of $\sigma_z$ is not 1, but somewhere between 0 and 1. In fact, the optimal estimate turns out to be $1/3$ in this case, as we will see in the next section.

In this paper, without assuming unbiasedness of the estimation, we study the optimal procedure for the expectation value of a physical observable of an unknown pure state, when $N$ copies of the state are presented. We assume that the unknown pure state is chosen from the pure-state space according to a unitary invariant *a priori* distribution. The optimal estimation is determined by minimizing the squared error averaged over the *a priori* distribution.

## II. OPTIMAL ESTIMATION

We determine the optimal way to estimate the expectation value of a physical observable $\Omega$ when $N$ copies of an unknown pure state $\rho=|\phi\rangle\langle\phi|$ on a $d$-dimensional Hilbert space $\mathcal{H}$ are given. Let $\{E_a\}$ be a positive-operator-valued measure (POVM) on the total system $\mathcal{H}^{\otimes N}$, with outcome labeled $a$ providing an estimate $\omega_a$ for the expectation value given by

 

$\mathrm{tr}[\rho\Omega]$. For a given $\rho$, the mean squared error in the estimate is written as

$$\Delta(\rho) = \sum_a \mathrm{tr}[E_a \rho^{\otimes N}](\omega_a - \mathrm{tr}[\rho\Omega])^2. \quad (1)$$

We will first average this $\Delta(\rho)$ over all pure states $\rho$ and then minimize it with respect to the POVM $\{E_a\}$ and the estimate $\{\omega_a\}$.

The distribution of the pure states $\rho$ is specified in the following way. Expand a pure state as $|\phi\rangle = \Sigma_{i=1}^{d} c_i |i\rangle$ in terms of an orthonormal base $\{|i\rangle\}$ of $\mathcal{H}$. The distribution is then defined to be the one in which the $2d$-component real vector $\{x_i = \mathrm{Re}\, c_i, \quad y_i = \mathrm{Im}\, c_i\}$ is uniformly distributed on the $(2d-1)$-dimensional hypersphere of radius 1. The distribution is unitary invariant in the sense that it is independent of the orthonormal base $\{|i\rangle\}$ chosen to define it. Let us denote the average over this distribution by $\langle\cdots\rangle$. All we need in the following calculation is a useful relation for the average of $\rho^{\otimes n}$ given in Ref. [16], that is,

$$\langle \rho^{\otimes n} \rangle = \frac{\mathcal{S}_n}{d_n}, \quad (2)$$

where $\mathcal{S}_n$ is the projection operator onto the totally symmetric subspace of $\mathcal{H}^{\otimes n}$ and $d_n$ is its dimension given by $d_n = \mathrm{tr}[\mathcal{S}_n] = {}_{n+d-1}C_{d-1}$. It may be instructive to see how this formula comes out in some simple cases of qubits ($d=2$), in which the above distribution means that the Bloch vector $\boldsymbol{n}$ is uniformly distributed on the surface of the Bloch sphere. Then we can easily verify

$$\frac{1}{4\pi}\int d\boldsymbol{n}\left(\frac{1+\boldsymbol{n}\cdot\boldsymbol{\sigma}}{2}\right)^{\otimes 2} = \frac{1}{24}[\boldsymbol{\sigma}(1)+\boldsymbol{\sigma}(2)]^2 = \begin{cases} 1/3 & (S=1), \\ 0 & (S=0), \end{cases} \quad (3)$$

where $S$ is the eigenvalue of the total spin. This is a special case of the formula (2), since the state is symmetric if $S=1$ and antisymmetric if $S=0$. The case of three qubits provides another example.

$$\frac{1}{4\pi}\int d\boldsymbol{n}\left(\frac{1+\boldsymbol{n}\cdot\boldsymbol{\sigma}}{2}\right)^{\otimes 3} = \begin{cases} 1/4 & (S=3/2), \\ 0 & (S=1/2). \end{cases} \quad (4)$$

Now going back to the general-dimensional case, we expand Eq. (1) and perform averaging over $\rho$ by use of the formula (2):

$$\begin{aligned}\langle\Delta\rangle &= \sum_a \langle \mathrm{tr}[E_a\rho^{\otimes N}]\{\omega_a^2 - 2\omega_a\,\mathrm{tr}[\rho\Omega] + (\mathrm{tr}[\rho\Omega])^2\}\rangle \\ &= \langle \Delta_1(\rho) + \Delta_2(\rho) + \Delta_3(\rho)\rangle, \end{aligned} \quad (5)$$

where we denote the three terms in $\Delta(\rho)$ by $\Delta_1(\rho)$, $\Delta_2(\rho)$, and $\Delta_3(\rho)$, and we evaluate each separately. The first term $\langle\Delta_1\rangle$ is readily calculated as

$$\langle\Delta_1\rangle = \frac{1}{d_N}\sum_a \omega_a^2\,\mathrm{tr}[E_a\mathcal{S}_N]. \quad (6)$$

For $\langle\Delta_3\rangle$, we first use the completeness of the POVM by summing over $a$ and perform the average in the following way:

$$\begin{aligned}\langle\Delta_3\rangle &= \langle(\mathrm{tr}[\rho\Omega])^2\rangle = \langle \mathrm{tr}[\rho^{\otimes 2}\Omega(1)\Omega(2)]\rangle \\ &= \frac{1}{d_2}\mathrm{tr}[\mathcal{S}_2\Omega(1)\Omega(2)] = \frac{1}{d(d+1)}[(\mathrm{tr}\Omega)^2 + \mathrm{tr}\Omega^2], \end{aligned} \quad (7)$$

where $\rho^{\otimes 2}$ is understood to be the tensor product of two $\rho$'s in spaces 1 and 2, and the space on which the operator $\Omega$ acts is specified by the number in the parentheses. Hereafter we will use this convention in more general cases, namely,

$$\Omega(n) \equiv \mathbf{1}^{\otimes(n-1)} \otimes \Omega \otimes \mathbf{1} \otimes \mathbf{1} \otimes \cdots. \quad (8)$$

Evaluation of the second term $\langle\Delta_2\rangle$ is more involved. Introducing another system on $\mathcal{H}$, which we call system $N+1$, we have

$$\begin{aligned}\langle\Delta_2\rangle &= -2\sum_a \omega_a\langle \mathrm{tr}[E_a\rho^{\otimes N}]\mathrm{tr}[\rho\Omega]\rangle \\ &= -2\sum_a \omega_a \langle \mathrm{tr}[E_a\rho^{\otimes(N+1)}\Omega(N+1)]\rangle \\ &= -\frac{2}{d_{N+1}}\sum_a \omega_a\,\mathrm{tr}[E_a\mathcal{S}_{N+1}\Omega(N+1)], \end{aligned} \quad (9)$$

where the traces in the second and third equalities is understood to be over systems $1,2,\ldots,N$, and $N+1$. The operator $\Omega(N+1)$ acts on system $N+1$. The projection operator $\mathcal{S}_{N+1}$ is the sum of all permutation operators of $N+1$ systems divided by a factor of $(N+1)!$. Any permutation of $N+1$ objects is either just a permutation among the first $N$ objects or the product of a permutation among the first $N$ objects and the transposition between the $(N+1)$th object and one of the first $N$ objects. With this observation we find, for any operator $\Omega$,

$$\mathrm{tr}_{N+1}[\mathcal{S}_{N+1}\Omega(N+1)] = \frac{\mathcal{S}_N}{N+1}\left(\mathrm{tr}\,\Omega + \sum_{n=1}^{N}\Omega(n)\right), \quad (10)$$

where $\mathrm{tr}_{N+1}$ is the trace over the $(N+1)$st system. We use this formula to trace out the newly introduced system $(N+1)$ in the expression $\Delta_2$ given by Eq. (9). The result is given by

$$\Delta_2 = -\frac{2}{d_N}\sum_a \omega_a\,\mathrm{tr}[E_a\mathcal{S}_N\hat{\Omega}], \quad (11)$$

where we define the symmetric one-body operator $\hat{\Omega}$ to be

$$\hat{\Omega} \equiv \frac{1}{N+d}\left(\mathrm{tr}\,\Omega + \sum_{n=1}^{N}\Omega(n)\right). \quad (12)$$

Combining the three averages $\langle\Delta_1\rangle$, $\langle\Delta_2\rangle$, and $\langle\Delta_3\rangle$, we obtain

$$\langle\Delta\rangle = \frac{1}{d_N}\sum_a \mathrm{tr}[E_a\mathcal{S}_N(\omega_a^2 - 2\omega_a\hat{\Omega})] + \frac{1}{d(d+1)}[(\mathrm{tr}\,\Omega)^2 + \mathrm{tr}\,\Omega^2]. \tag{13}$$

To minimize $\langle\Delta\rangle$ we complete the square with respect to $\omega_a$ in this expression. Owing to the completeness of the POVM, this is reduced to the calculation of $\mathrm{tr}[\mathcal{S}_N\hat{\Omega}^2]$, which can be performed by using the following formulas:

$$\mathrm{tr}[\mathcal{S}_N\Omega(n)] = \frac{d_N}{d}\mathrm{tr}\,\Omega, \tag{14}$$

$$\mathrm{tr}[\mathcal{S}_N\Omega(n)\Omega(m)] = \begin{cases} \dfrac{d_N}{d}\mathrm{tr}[\Omega^2] & (n=m), \\ \dfrac{d_N}{d(d+1)}(\mathrm{tr}[\Omega^2] + (\mathrm{tr}\,\Omega)^2) & (n\neq m). \end{cases} \tag{15}$$

After some calculation we find

$$\mathrm{tr}[\mathcal{S}_N\hat{\Omega}^2] = \frac{d_N}{d(d+1)(N+d)}\{N\,\mathrm{tr}[\Omega^2] + (N+d+1)(\mathrm{tr}\Omega)^2\}. \tag{16}$$

We thus finally obtain the mean squared error in the completed square form

$$\langle\Delta\rangle = \frac{1}{d_N}\sum_a \mathrm{tr}[E_a\mathcal{S}_N(\omega_a - \hat{\Omega})^2] + \frac{1}{d(d+1)(N+d)}(d\,\mathrm{tr}[\Omega^2] - (\mathrm{tr}\,\Omega)^2). \tag{17}$$

Now note that the first term in Eq. (17) is positive. This is because $\hat{\Omega}$ is symmetric under exchange of component subsystems and therefore $\mathcal{S}_N(\omega_a-\hat{\Omega})^2=\mathcal{S}_N(\omega_a-\hat{\Omega})^2\mathcal{S}_N$ is a positive operator. The $\Delta$ has a lower bound given by the second term of Eq. (17). Let us denote the eigenvalue of $\Omega$ by $\Omega_i$ $(i=1,\ldots,d)$ and the corresponding eigenstate by $|i\rangle$. It is then readily seen that this lower bound can be achieved if the index $a$ of the POVM element collectively represents the set of $\{i_1,i_2,\ldots,i_N\}$, the POVM element is taken to be the projector

$$E_{i_1,i_2,\ldots,i_N} = |i_1 i_2\cdots i_N\rangle\langle i_1 i_2\cdots i_N|, \tag{18}$$

and the estimate $\omega_a$ to be the corresponding eigenvalue of $\hat{\Omega}$,

$$\omega_{i_1,i_2,\ldots,i_N} = \frac{1}{N+d}\left(\mathrm{tr}\,\Omega + \sum_{n=1}^{N}\Omega_{i_n}\right). \tag{19}$$

Thus we conclude that the mean squared error $\langle\Delta\rangle$ in the estimation for the expectation value of the observable $\Omega$ takes its minimum value given by

$$\Delta_{\mathrm{opt}} = \frac{1}{d(d+1)(N+d)}(d\,\mathrm{tr}[\Omega^2] - (\mathrm{tr}\,\Omega)^2), \tag{20}$$

if one measures the observable $\Omega$ independently for each system and makes the estimate given by

$$\omega_{\mathrm{opt}} \equiv \frac{1}{N+d}\left(\mathrm{tr}\,\Omega + \sum_{n=1}^{N}\Omega_{i_n}\right), \tag{21}$$

where $\{\Omega_{i_1},\Omega_{i_2},\ldots,\Omega_{i_N}\}$ are the data observed by the measurement.

The optimal estimate $\omega_{\mathrm{opt}}$ is not the arithmetic average of observed data (the sample average), though the optimal measurement is projective and independent. For a finite $N$, it is not unbiased either since

$$\sum_a \omega_a\,\mathrm{tr}[E_a\rho^{\otimes N}] = \mathrm{tr}[\hat{\Omega}\rho^{\otimes N}] = \frac{1}{N+d}(\mathrm{tr}\,\Omega + N\,\mathrm{tr}[\rho\Omega]), \tag{22}$$

which only asymptotically approaches $\mathrm{tr}[\rho\Omega]$. In Sec. IV we will discuss the biasedness of $\omega_{\mathrm{opt}}$ and present an interpretation of its structure.

What do we obtain for the mean squared error if we take the sample average of the values of $\Omega$ observed by the successive measurements on each copy? In this case the POVM is given by Eq. (18) and the estimate by

$$\omega_{i_1,i_2\ldots,i_N} = \frac{1}{N}\sum_{n=1}^{N}\Omega_{i_n} \equiv \omega_{\mathrm{av}}, \tag{23}$$

which can be easily shown to be unbiased. The squared error for a given $\rho$ given in Eq. (1) takes the form

$$\Delta_{\mathrm{av}}(\rho) = \frac{1}{N}(\mathrm{tr}[\rho\Omega^2] - (\mathrm{tr}[\rho\Omega])^2). \tag{24}$$

After the average over $\rho$ we have

$$\Delta_{\mathrm{av}} = \langle\Delta_{\mathrm{av}}(\rho)\rangle = \frac{1}{d(d+1)N}[d\,\mathrm{tr}\,\Omega^2 - (\mathrm{tr}\,\Omega)^2], \tag{25}$$

where we used Eq. (7).

Comparing $\Delta_{\mathrm{opt}}$ and $\Delta_{\mathrm{av}}$, we find that the only difference between them is in the factor in the denominators, $N+d$ in $\Delta_{\mathrm{opt}}$ and $N$ in $\Delta_{\mathrm{av}}$. While $\Delta_{\mathrm{opt}}$ is certainly less than $\Delta_{\mathrm{av}}$, both show the same asymptotics when the number of copies goes to infinity. The difference becomes important when the number of copies is comparable to the dimension of the system.

Let us examine the example discussed in Sec. I, in which $\sigma_z$ is measured with the result 1 for a single qubit in an unknown pure state ($d=2$ and $N=1$). In this case the observed data is $\{1\}$. The estimate by the sample average gives $\omega_{\mathrm{av}}=1$ for the expectation value of $\sigma_z$ with the mean squared error $\Delta_{\mathrm{av}}=2/3$, whereas the optimal estimation predicts $\omega_{\mathrm{opt}}=1/3$ with the mean squared error $\Delta_{\mathrm{opt}}=2/9$.

## III. ESTIMATION WITH THE UNBIASEDNESS CONDITION

In Ref. [15], D'Ariano, Giovannetti, and Perinotti considered the estimation for the expectation of observables under the unbiasedness condition for any generally mixed state $\rho^{\otimes N}$ and showed that the optimal estimate under the constraint is given by the sample average obtained by the independent successive measurement of the observable on each copy. In this section we briefly discuss the same problem in the pure state case and show the same conclusion holds.

The unbiasedness condition is written as

$$\sum_a \omega_a \,\mathrm{tr}[E_a \rho^{\otimes N}] = \mathrm{tr}[\Omega\rho]. \tag{26}$$

Note that $\mathrm{tr}[\Omega\rho]$ on the right-hand side can be expressed as

$$\mathrm{tr}[\Omega\rho] = \mathrm{tr}[\hat{\Omega}_{\mathrm{av}}\rho^{\otimes N}], \tag{27}$$

$$\hat{\Omega}_{\mathrm{av}} \equiv \frac{1}{N}\sum_{n=1}^{N} \Omega(n). \tag{28}$$

If the unbiasedness condition (26) is assumed for any generally mixed state $\rho$, then it can be shown [15] that

$$\sum_a \omega_a E_a = \hat{\Omega}_{\mathrm{av}} \tag{29}$$

for any permutation-invariant POVM $\{E_a\}$. If we require the unbiasedness condition for any pure state $\rho$, we can still show that the relation (29) holds in the totally symmetric subspace of $\mathcal{H}^{\otimes N}$, namely,

$$\mathcal{S}_N\left(\sum_a \omega_a E_a - \hat{\Omega}_{\mathrm{av}}\right)\mathcal{S}_N = 0. \tag{30}$$

This follows from a lemma for an operator $A$ on $\mathcal{H}^{\otimes N}$:

$$\mathrm{tr}[A\rho^{\otimes N}] = 0 \quad \text{for any pure state } \rho,$$

$$\text{if and only if } \mathcal{S}_N A \mathcal{S}_N = 0.$$

The "if" part is trivial and we sketch the proof of the "only if" part. We write $|\phi\rangle = \Sigma_{i=1}^{d} c_i|i\rangle$ in terms of a basis $\{|i\rangle\}$ of $\mathcal{H}$, where $\rho = |\phi\rangle\langle\phi|$. Then we have

$$\begin{aligned}\mathrm{tr}[A\rho^{\otimes N}] &= \sum_{i_1\cdots i_N, j_1\cdots j_N} c_{i_1}^{*} c_{i_2}^{*}\cdots c_{i_N}^{*} c_{j_1} c_{j_2}\cdots c_{j_N} \\ &\quad\times \langle i_1 i_2\cdots i_N|A|j_1 j_2 \ldots j_N\rangle \\ &= \sum_{n_i, m_i} c_1^{*n_1} c_2^{*n_2}\cdots c_d^{*n_d} c_1^{m_1} c_2^{m_2}\cdots c_d^{m_d} \\ &\quad\times \langle \psi_{n_1 n_2\cdots n_d}|A|\psi_{m_1 m_2\cdots m_d}\rangle = 0,\end{aligned} \tag{31}$$

where the summation over integers $n_i \geqslant 0$ and $m_i \geqslant 0$ should be taken under the conditions $\Sigma_i n_i = \Sigma_i m_i = N$, and the state $|\psi_{n_1 n_2\cdots n_d}\rangle$ is the occupation-number representation of symmetric states (generally not normalized), with $n_i$ being the occupation number of state $i$. Equation (31) should hold for any complex $c_i$, implying $\langle\psi_{n_1 n_2\cdots n_d}|A|\psi_{m_1 m_2\cdots m_d}\rangle = 0$.

The difference between the two unbiased conditions (29) and (30) is the projection operator $\mathcal{S}_N$ in the pure-state case. This, however, does not hamper the subsequent argument since the support of the operator $\rho^{\otimes N}$ for pure $\rho$ is the totally symmetric subspace.

We go back to the expanded form of $\Delta(\rho)$ as in Eq. (5), but before being averaged over $\rho$. By using the unbiased condition (30) we readily find $\Delta_2(\rho) = -2\Delta_3(\rho)$ so that we have

$$\begin{aligned}\Delta(\rho) &= \sum_a \omega_a^2 \mathrm{tr}\Delta(\rho) = \sum_a \omega_a^2\,\mathrm{tr}[E_a\rho^{\otimes N}] - (\mathrm{tr}[\rho\Omega])^2.[E_a\rho^{\otimes N}] \\ &\quad - (\mathrm{tr}[\rho\Omega])^2.\end{aligned} \tag{32}$$

It can be shown that

$$\sum_a \omega_a^2 \mathrm{tr}[E_a\rho^{\otimes N}] \geqslant \mathrm{tr}\,[\hat{\Omega}_{\mathrm{av}}^2\rho^{\otimes N}], \tag{33}$$

since in the symmetric subspace we have

$$0 \leqslant \sum_a (\omega_a - \hat{\Omega}_{\mathrm{av}})E_a(\omega_a - \hat{\Omega}_{\mathrm{av}}) = \sum_a \omega_a^2 E_a - \hat{\Omega}_{\mathrm{av}}^2. \tag{34}$$

It is evident that the equality holds if the POVM element $E_a$ is the projector of the eigenstate of $\hat{\Omega}_{\mathrm{av}}$ and the estimate $\omega_a$ is the corresponding eigenvalue, which is the sample average of the observed values of $\Omega$ for each copy. Thus the minimum value of the squared error in the unbiased estimation is given by

$$\mathrm{tr}[\hat{\Omega}_{\mathrm{av}}^2\rho^{\otimes N}] - (\mathrm{tr}[\rho\Omega])^2 = \frac{1}{N}(\mathrm{tr}[\rho\Omega^2] - (\mathrm{tr}[\rho\Omega])^2) = \Delta_{\mathrm{av}}(\rho), \tag{35}$$

which shows that the conclusion of Ref. [15] holds if we restrict ourselves to the pure-state input ensemble. Averaging over $\rho$ gives $\Delta_{\mathrm{av}}$ given in Eq. (25).

## IV. DISCUSSION AND CONCLUDING REMARKS

We have seen that the optimal estimation of the expectation value of a physical observable is biased, though the optimal measurement is given by the successive projective measurement of the observable. The optimal estimate $\omega_{\mathrm{opt}}$ is not given by the arithmetic average of observed data.

We can interpret the expression (21) of the optimal estimate $\omega_{\mathrm{opt}}$ in the following way. First of all, we should remember that we have full knowledge on properties of the observable $\Omega$ including its eigenvalues. Otherwise we cannot perform a measurement associated with $\Omega$. Then what can we expect for outcomes of the $\Omega$ measurement *before* performing the measurement? The state $\rho$ is given to us according to the unitary invariant distribution on the pure-state space, implying that we expect that each eigenvalue $\Omega_i$ occurs with equal probabilities as the outcome of the $\Omega$ measurement. This *a priori* knowledge should be somehow taken into account in the estimation. We can see that this *a priori* knowledge is incorporated into the optimal estimate $\omega_{\mathrm{opt}}$ in a natural way. It is just the arithmetic average of $N$ observed data points $\{\Omega_{i_n}\}_{n=1}^{N}$ and the $d$ "default nonobserved" data points $\{\Omega_i\}_{i=1}^{d}$, the latter of which add up to the trace of the observable.

One may still wonder why the weights of the average for the observed and nonobserved data are equal. Actually this is a feature of the pure-state ensemble considered in this paper. To see this, let us take the simplest example of $d=2$ and $N=1$, but this time the state $\rho$ is generally mixed. We assume that the Bloch vector $\boldsymbol{n}$ is distributed isotropically inside the Bloch sphere. The ensemble is characterized by the average $\langle n^2 \rangle$, which is 1 for the pure-state ensemble, but generally less than 1.

After some calculation, the mean squared error turns out to be

$$\langle \Delta \rangle = \frac{1}{2} \sum_a \mathrm{tr}[E_a(\omega_a - \hat{\Omega})^2] + \frac{\langle n^2 \rangle}{12}\left(1 - \frac{\langle n^2 \rangle^2}{3}\right)(2\,\mathrm{tr}[\Omega^2] - (\mathrm{tr}\,\Omega)^2), \tag{36}$$

where

$$\hat{\Omega} = \frac{1}{3}\left(\frac{3 - \langle n^2 \rangle}{2}\mathrm{tr}\,\Omega + \langle n^2 \rangle \Omega\right). \tag{37}$$

This implies that the optimal measurement is the projective measurement of $\Omega$, and the optimal estimate is given by

$$\omega_{\mathrm{opt}} = \frac{1}{3}\left(\frac{3 - \langle n^2 \rangle}{2}\mathrm{tr}\,\Omega + \langle n^2 \rangle \Omega_{i_1}\right), \tag{38}$$

where $\Omega_{i_1}$ is the observed eigenvalue of $\Omega$. The minimal mean squared error is given by the second term of Eq. (36). We can see that the weight for the observed data decreases as the degree of mixing of the ensemble increases. When $\langle n^2 \rangle = 0$, this $\omega_{\mathrm{opt}}$ implies we should disregard the observed data. The reason is that we know that the expectation value is given by $\mathrm{tr}\,\Omega/2$ for a completely mixed state.

The generalization of our analysis to an ensemble of mixed states, including the details of the above discussion, will be presented elsewhere.

[1] W. K. Wootters and W. H. Zurek, Nature (London) **299**, 802 (1982).
[2] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[3] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[4] A. Peres and W. K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).
[5] W. K. Wootters, e-print quant-ph/0506149.
[6] M. Ban, K. Kurokawa, R. Momose, and O. Hirota, Int. J. Theor. Phys. **36**, 1269 (1997).
[7] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, Phys. Rev. A **58**, 146 (1998).
[8] Y. C. Eldar and G. D. Forney, Jr., IEEE Trans. Inf. Theory **47**, 858 (2001).
[9] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **59**, 1070 (1999).
[10] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).
[11] R. Derka, V. Bužek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998).
[12] E. Bagan, M. Baig, and R. Muñoz-Tapia, Phys. Rev. A **64**, 022305 (2001); Phys. Rev. Lett. **87**, 257903 (2001).
[13] M. Hayashi, J. Phys. A **31**, 4633 (1998).
[14] A. S. Holevo, e-print quant-ph/0307225.
[15] G. M. D'Ariano, V. Giovannetti, and P. Perinotti, J. Math. Phys. **47**, 022102 (2006).
[16] A. Hayashi, T. Hashimoto, and M. Horibe, Phys. Rev. A **72**, 032325 (2005).