Fourier analytic approach to phase estimation in quantum systems

Hiroshi Imai\textsuperscript{1} and Masahito Hayashi\textsuperscript{2}

\textsuperscript{1} National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
\textsuperscript{2} Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai 980-8579, Japan
E-mail: himai@nii.ac.jp and hayashi@math.is.tohoku.ac.jp

Abstract. For a unified analysis of phase estimation, we focus on the limiting distribution. It is shown that the limiting distribution can be expressed as the absolute square of the Fourier transform of an $L^2$ function whose support belongs to $[-1, 1]$. Using this result, we study the relation between the variance of the limiting distribution and its tail probability. We prove that the protocol which minimizes the asymptotic variance does not minimize the tail probability. We derive the estimation protocol minimizing the tail probability outside a given interval, depending on the width of the interval. Such an optimal protocol is given by a prolate spheroidal wave function, which often appears in wavelet or time-limited Fourier analysis. Also, within the framework of interval estimation, we derive the minimum confidence interval that guarantees a given confidence coefficient.
1. Introduction

Estimating and/or identifying an unknown unitary operator is discussed in both research fields of quantum computation [1, 2] and quantum statistical inference, therefore, it is a fundamental topic in quantum information. In quantum computation, the unknown unitary operator is given as an oracle, and it is discussed how many applications are required for identifying the given unknown oracle with a given precision. In quantum statistical inference, by contrast, many researchers optimize the average fidelity or mean square error between the true unitary operator and the obtained guess [3]–[11]. In both research areas, we optimize the state inputting the unknown unitary operator as well as the measurement, and quadratic speedup is reported by researchers in both fields. However, in each field, researchers discuss this same topic based on different criteria, and there is an example for which quadratic speedup appears or not depending on the choice of the criterion. Therefore, since the relation between these different criteria is not clear, it is desirable to treat this problem within a common framework.

In the present paper, as the most typical example, we focus on phase estimation, for which quadratic speedup has been demonstrated experimentally [12, 13]. Kitaev was the first to deal with the phase estimation problem from the quantum computation viewpoint [14]. Since it appears in Shor’s factorization, it is considered as a fundamental topic in quantum computation as well as physics. In order to treat the quadratic speedup in more depth, we focus on the limiting distribution, which describes the stochastic behavior of the estimate in a neighborhood of the true parameter. That is, it provides the distribution of the random variable $n(\hat{\theta} - \theta)$ when the estimate, the true parameter, and the number of applications are given as $\hat{\theta}$, $\theta$ and $n$. While the limiting distribution is a common concept in statistics and has been studied in the context of estimation of a quantum state [15]–[17], it has not been studied systematically in the context of estimation of an unknown unitary operator.

The concept of ‘limiting distribution’ is very useful for the following four reasons. Firstly, the variance of the limiting distribution provides the asymptotic first-order coefficient of the
mean square error. Secondly, the tail probability of the limiting distribution for a given interval provides the tail probability of the interval when the width of the interval is of order $1/n$. Thirdly, using the limiting distribution, we can discuss phase estimation within the framework of interval estimation, as we explain later. Fourthly, using this concept, we can treat the number of applications required for attaining a given accuracy, i.e. for dealing with the error probability and error bars in the asymptotic framework. In fact, the four advantages above correspond to particular criteria. Therefore, the limiting distribution provides a unified framework for these criteria. The first three criteria are familiar in statistics, and the fourth criterion is familiar in computer science.

In the present paper, we analyze the limiting distribution for phase estimation systematically, and show that the limiting distribution can be expressed as the Fourier transform of a square integrable function on the closed interval $[-1, 1]$, which approximately yields the input state in the asymptotic setting.

In a realistic setting, the optimization corresponding to the fourth criterion is more appropriate than that of the first criterion, i.e. the optimization of the mean square error or the average fidelity. However, only the estimator with several special input states have been treated in the fourth formulation, and its optimization has not been discussed, whereas optimization with respect to the mean square error and the average fidelity has already been investigated [4, 5, 18].

In statistics, in order to treat this problem, interval estimation is considered, in which our estimate is given as an interval. Indeed, there are two formulations in statistics: one is point estimation, in which our estimate is given as a single point, and the other is interval estimation. In point estimation, it is not easy to guarantee the quality of the estimate because the estimated value always has statistical fluctuation. In order to resolve this problem, in interval estimation, for given data and confidence coefficient, our estimate is given as an interval, known as a confidence interval. In this formulation, a confidence interval of smaller width is better. As is mentioned in section 2, when the number $n$ of data values is sufficiently large, the confidence interval can be provided by using the limiting distribution [19].

Further, we analyze the variance and the tail probability of limiting distribution. The result we prove is that the limiting distribution that minimizes the asymptotic variance does not minimize the tail probability. We also show that the limiting distribution that minimizes the tail probability depends on the width of the interval. The definition of the tail probability also depends on the width of the interval. Such an optimal input state is given by a prolate spheroidal wave function [20], which often appears in wavelet or time-limited Fourier analysis. This function is a solution of the linear differential equation [21]

$$\frac{d}{dx}(1-x^2)\frac{df}{dx} + (\xi(R) - R^2 x^2) f = 0.$$ 

Originally, prolate spheroidal wave functions appeared in analysis of the Helmholtz equation in electromagnetics [22] or the determination of laser modes [23]. Employing this wave function, Slepian and Pollak [20] extended Shannon’s sampling theorem to the case where the time interval is limited as well as the bandwidth, whereas Shannon’s original sampling theorem dealt with the limited bandwidth case.

Further, using these facts, we study optimal interval estimation. We provide the estimation protocol minimizing the width of the confidence interval that guarantees the given confidence coefficient. In this case, the optimal estimation protocol depends on a given confidence
coefficient. That means we must choose the input state appropriately, depending on the confidence coefficient.

The paper is organized as follows. In section 2, the formulation of phase estimation is given and the limiting distribution is introduced with an explanation of its meaning. In section 3, we clarify the relation between the limiting distribution and the Fourier transform. In section 4, we analyze the variance of the limiting distribution. This problem is reduced to finding the minimum eigenvalue of a operator for the Dirichlet problem. In section 5, the tail probability of the limiting distribution is discussed. It is shown that the limiting distribution minimizing the variance does not provide a small tail probability. In section 6, we treat the interval estimation problem. This problem can be analyzed by a prolate spheroidal wave function and the eigenvalue of the defining differential equation. In section 7, phase estimation with a single copy is discussed for the continuous system. The discussions in sections 4 and 5 can be applied to this formulation under a deterministic energy constraint. In section 8, we present a short note on the asymptotic Cramér–Rao lower bound.

2. Limiting distribution

Let us consider the estimation problem for an unknown phase shift $\theta$ with an $n$-fold unitary evolution $V_n^\theta$ of unitary operator $V^\theta := \begin{bmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{bmatrix}$, for which our estimating protocol is given by a combination of an appropriate input state $|\phi_{0,n}\rangle$ and a suitable measurement $M^n$ (see figure 1).

As is discussed later, this formulation essentially contains the most general framework with $n$ applications of the unknown unitary operator. In the following discussion, the true parameter is denoted by $\theta$ and our estimate is denoted by $\hat{\theta}$.

As is shown in appendix A, this problem is equivalent to estimating the parameter $\theta$ of the unitary operation

$$U^n_\theta = \sum_{k=0}^{n} e^{i(k-n/2)\theta} |k\rangle \langle k|. \quad (1)$$

Our scheme for estimating $\theta$ is as follows. Firstly, prepare an input state $|\phi_{0,n}\rangle = \sum_{k=0}^{n} a_k |k\rangle$ such that the coefficients $\tilde{a}^n = \{a_k\}_{k=0}^{n}$ satisfy the normalizing condition $\sum_{k} |a_k|^2 = 1$.\(^3\) Secondly, allow the input state $|\phi_{0,n}\rangle$ to evolve by means of the unitary evolution $U^n_\theta$. Lastly, perform a measurement defined by the positive operator value measure (POVM) $M^n = M^n(\hat{\theta})d\hat{\theta}$. Then, the estimate $\hat{\theta}$ obeys the probability distribution

$$P_{\theta,\tilde{a}^n}(\hat{\theta}) := \langle \phi_\theta | M^n(\hat{\theta}) | \phi_\theta \rangle,$$

\(^3\) As is shown in appendix A, even if input state is entangled with the ancilla, there is no improvement.

---

\(^3\) As is shown in appendix A, even if input state is entangled with the ancilla, there is no improvement.

New Journal of Physics 11 (2009) 043034 (http://www.njp.org/)
where

$$|\phi_{0,n}\rangle := U_0^n |\phi_{0,n}\rangle = \sum_{k=0}^{n} a_k e^{i(k-n/2)\theta} |k\rangle.$$  

When our error function is given by \( R(\theta, \hat{\theta}) \), we optimize the mean error \( D_\theta(M^n, \hat{a}^n) := \int_0^{2\pi} R(\theta, \hat{\theta}) P_{\theta, \hat{\theta}}(\hat{\theta}) d\hat{\theta} \). We consider only the covariant framework, i.e. the situation where the error function \( R(\theta, \hat{\theta}) \) is assumed to be given as a function of the difference \( (\theta - \hat{\theta}) \text{Mod} 2\pi \). For example, when we focus on the gate fidelity \( |\text{Tr} V_\theta V_\hat{\theta}^{-1} / 2|^2 = \sin^2(\theta - \hat{\theta}) / 2 \) satisfies such a covariant condition.

Then, our measurement may be restricted into a group covariant measurement

$$M^n_{\theta}(d\hat{\theta}) := U_\theta^n |t\rangle \langle t| U_\theta^{n\dagger} \frac{d\hat{\theta}}{2\pi}, \tag{2}$$

where

$$|t\rangle = \sum_{k=0}^{n} e^{ik\theta} |k\rangle. \tag{3}$$

This is because the minimum of the Bayesian average value \( \min_{M^n} 1/2\pi \int_0^{2\pi} D_\theta(M^n, \hat{a}^n) d\theta \) under the invariant prior and the mini-max value \( \min_{M^n} \max_\theta D_\theta(M^n, \hat{a}^n) \) can be attained by the same group covariant measurement [24]. Therefore, we restrict our measurement to covariant measurements in the following discussion. In this situation, our protocol is described by the pair of the coefficient of the input state \( \hat{a}^n \) and the vector \( |t\rangle \) given in (3).

Further, without loss of generality, we can restrict our protocol to the pair \((\hat{a}^n, |t_0\rangle)\) as follows, where

$$|t_0\rangle = \sum_{k=0}^{n} |k\rangle. \tag{4}$$

For any protocol \((\hat{a}^n, |t\rangle)\), we define \( \hat{a}^n = \{a_k\} \) by

$$a'_k := a_k e^{-ik\theta}. \tag{5}$$

Then, as explained below, the protocol \((\hat{a}'^n, |t_0\rangle)\) has the same performance as the protocol \((\hat{a}^n, |t_0\rangle)\).

$$P_{\theta, \hat{a}^n}^{M^n_{\theta}(0)} (\hat{\theta}) = |\langle \phi_0 | U_\theta^n |t\rangle|^2 = \left| \sum_{k=0}^{n} a_k |k\rangle \langle k| U_\theta^{n\dagger} |t\rangle \right|^2 = \left| \sum_{k=0}^{n} a_k e^{ik\theta} e^{i(k-n/2)(\theta - \hat{\theta})} \right|^2 = \left| \sum_{k=0}^{n} a'_k e^{ik(\theta - \hat{\theta})} \right|^2 = P_{\theta, \hat{a}'^n}^{M^n_{\theta}(0)} (\hat{\theta}).$$
Therefore, the choice of our protocol is essentially given by the choice of input state. Dam et al [25] essentially showed that, even if adaptive $n$ applications of the unitary $V_\theta$ are allowed, such a protocol can be simulated by the unitary evolution operator $U_\theta^n$ (1) with the above error function. This fact implies that such a protocol can be simulated by an $n$-fold unitary evolution operator $V_\theta^\otimes n$. Chiribella et al [26] showed this type reduction with a general group covariant framework more clearly.

The main target of the present paper is to analyze the asymptotic behavior of the output distribution for a sequence of input states $\mathcal{M} := \{\hat{a}^n\}$. For this purpose, we treat the distribution concerning the parameter $z_n = n(\hat{\theta}_n - \theta)/2$ because the estimate $\hat{\theta}_n$ approaches the true parameter $\theta$ to the order $1/n$ when an appropriate measurement and an appropriate input state are used. When the random variable $z_n$ converges to a random variable $z$ in probability, the distribution $P(\mathcal{M})$ of $z$ is called the limiting distribution of the sequence of input states $\mathcal{M}$. In the case of state estimation including the classical case, if we apply a suitable estimator, the limiting distribution is the Gaussian distribution under a suitable regularity condition. In the classical case, more precisely, the asymptotic sufficient statistics for the given parameter obey this Gaussian distribution. That is, any estimator can be expressed as a function of statistics obeying the Gaussian distribution, asymptotically. This Gaussian distribution is characterized only by the variance. Even for the quantum case of state estimation, the estimation problem can be reduced to that of quantum Gaussian states family in the asymptotic sense [15]–[17]. In particular, if we treat the estimation of a one-parameter model, we obtain the same conclusion as that obtained in the classical case. Since the Gaussian distribution with mean zero is characterized only by the variance, it is sufficient to evaluate the variance when we are considering the limiting distribution. That is, there is no variation of the limiting distribution for state estimation because the limiting distribution is uniquely determined as a Gaussian distribution. The main problem in the present paper, on the other hand, is to determine whether or not there exists a variety of limiting distributions in the case of phase estimation.

When the cost function $R(\theta, \hat{\theta})$ has the form $R(\theta, \hat{\theta}) \cong c(\theta - \hat{\theta})^2 + o((\theta - \hat{\theta})^2)$, the average error behaves as $D_0(M_{\theta|\theta}, \hat{a}^n) \cong (c/n^2) \sqrt{V}$, where $V$ is the variance of the limiting distribution. As an example, for $R(\theta, \hat{\theta}) = \sin^2(\theta - \hat{\theta})/2$, the constant $c$ is $1/4$. Hence, analysis of the limiting distribution yields the asymptotic analysis for average gate fidelity. Further, analysis of the limiting distribution provides the asymptotic analysis of phase estimation from another aspect. For example, the tail probability of the sequence of input states $\mathcal{M}$ can be described as follows:

$$ P_{\theta, \hat{a}^n}^{M_{\theta|\theta}} \left\{ |\hat{\theta}_n - \theta| > \frac{A}{n} \right\} \to P(\mathcal{M})\{|z| > A\}. \quad (6) $$

Using this relation, we can evaluate the required number of applications of the unknown unitary gate for a given allowable error width $B$ and given allowable error probability $\epsilon$ as follows. First, we choose $A$ satisfying $P(\mathcal{M})\{|z| > A\} = \epsilon$. Next, we choose $n$ such that $A/n = B$, i.e. $n = A/B$. Thus, the required number of applications is equal to $A/B$ if we use the sequence of input states $\mathcal{M}$. The above discussions clarify that analysis of the limiting distribution yields various types of asymptotic analysis for phase estimation.

Hence, in the present paper, for a deeper and more unified asymptotic analysis of phase estimation, we analyze the limiting distribution of the sequence of input states $\mathcal{M}$.

New Journal of Physics 11 (2009) 043034 (http://www.njp.org/)
3. Relation with square integrable functions

In this section, we present a remarkable relationship between limiting distributions and square integrable functions on $[-1, 1]$. In the following, we denote the set of square integrable functions $f \in L^2([-1, 1])$ satisfying the normalizing condition $\int_{-\infty}^\infty |f(x)|^2 \, dx = 1$ by $L^2([-1, 1])$.

**Theorem 1.** For any square integrable function $f \in L^2([-1, 1])$, there exists a sequence of input states $\mathcal{M}$ such that

\[ P(\mathcal{M}) = P^f, \quad (7) \]

where $P^f(dy) = |F(f)(y)|^2 \, dy$ and $F(f)$ is the Fourier transform on $L^2(\mathbb{R})$ of $f$, i.e.

\[ F(f)(y) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty f(x)e^{i\pi y} \, dx. \]

Conversely, for any small real number $\epsilon > 0$ and any large real number $R > 0$, there exists a sufficiently large integer $N$ satisfying the following condition: For any $n \geq N$ and any input state $\hat{a}^n$, there exists a square integrable function $f \in L^2([-1, 1])$ such that

\[ \left| \int_B P_{\theta, \hat{a}^n}^{M_n} \left( \theta + \frac{2y}{n} \right) \frac{2}{n} \, dy - P^f(B) \right| \leq \epsilon \quad (8) \]

for any subset $B$ of $[-R, R]$.

Due to this relationship, we can convert the analysis of limiting distributions to the analysis of wave functions on the interval $[-1, 1]$. That is, our problem is reduced to Fourier analysis on the interval $[-1, 1]$.

**Proof.** As a first step in the proof, for a given input state $\hat{a}^n$, we construct a function $f \in L^2([-1, 1])$ satisfying (8). For this purpose, we define a function $f \in L^2([-1, 1])$ by

\[ f_n(x) := \frac{\hat{a}^n}{\sqrt{n}} e^{i\pi x/n} \quad (9) \]

for $x \in (x_k - 1/(n+1), x_k + 1/(n+1)]$, where $x_k := (2k - n)/(n+1)$. In the following, the set of the above $L^2$ functions is denoted by $L^2_n$. The parameter $z_n = [n(\theta_n - \theta)]/2$ can be replaced by the parameter $y_n = [(n+1)(\theta_n - \theta)]/2$ because the ratio $y_n/z_n \to 1$. Since

\[ \int_{x_k - 1/(n+1)}^{x_k + 1/(n+1)} f_n(x)e^{i\pi y_n} \, dx = \frac{\hat{a}^n}{\sqrt{n}} \sqrt{n+1} \frac{2}{y_n} \sin \frac{y_n}{n+1}, \]

we have

\[ P_{\theta, \hat{a}^n}^{M_n}(\theta_n) \frac{1}{2\pi} d\theta_n \, dym_n = \frac{1}{\pi(n+1)} \left| \sum_{k=0}^n \frac{a_k^m e^{i(k-n/2)(\theta_n - \theta)}}{\sqrt{n+1}} \right|^2 \, dy_n \]

\[ = \frac{1}{2\pi} \left| \sum_{k=0}^n \frac{1}{\sqrt{n+1}} a_k^m e^{i\pi k} e^{i\pi y_n} \right|^2 \, dy_n \]

\[ = \frac{1}{2\pi} \left| \int_{-1}^1 f_n(x)e^{i\pi y_n} \, dx \right|^2 \left( \frac{\sqrt{n}}{\sin \frac{\pi y_n}{n+1}} \right)^2 \, dy_n. \quad (10) \]
Then, a function \( f \) over all functions \( f \) satisfies \( (7) \). Therefore, the problem is therefore reduced to finding the minimum eigenvalue of the operator \( P \). Due to \( (10) \) and \( (11) \), the sequence of input states \( M_f := \{ \hat{a}^n \} \) satisfies \( (7) \).

For example, when the input state is \( \sum_{k=0}^{n} \sqrt{1/(n+1)} |k \rangle \), the function \( f \) is the constant \( \sqrt{1/n} \). Since its Fourier transformation is given by \( (1/\sqrt{2\pi}) \) (\( \sin y/\pi \)), the limiting distribution is given by

\[
P_f(dy) = \left( \frac{\sin y}{y} \right)^2 \frac{dy}{2\pi}.
\]

4. Variance of the limiting distribution

In the previous section, we showed that limiting distributions \( P_f \) of outcomes can be acquired through Fourier transforms of wave functions \( f \in L^2([-1, 1]) \); they correspond to sequences of coefficients \( a_k^f \) of input states. In this section, let us determine the input state minimizing the variance by utilizing this fact. As was mentioned in section 2, optimizing the first-order coefficient of the variance is equivalent to minimizing the variance \( V(f) := \int_{-\infty}^{\infty} y^2 P_f(dy) \) over all functions \( f \in L^2([-1, 1]) \).

Define the multiplication operator \( Q \) and the momentum operator \( P = -i(d/dx) \) on \( L^2(\mathbb{R}) \). Then, a function \( f \in L^2([-1, 1]) \) satisfies

\[
V(f) = \langle f | F^\dagger Q^2 F | f \rangle = \langle f | P^2 | f \rangle.
\]

Under the natural embedding from \( L^2([-1, 1]) \) to \( L^2(\mathbb{R}) \), the minimum value of \( V(f) \) is given by

\[
\min_{f \in L^2([-1, 1])} \langle f | P^2 | f \rangle,
\]

which is nothing but the Dirichlet problem because the restriction of the operator \( P^2 \) on \( L^2([-1, 1]) \) is equivalent to the square of the operator \( -i(d/dx) \) on \( \{ f \in L^2([-1, 1]) \cap C^1([-1, 1]) | f(-1) = f(1) = 0 \} \).

That is, the problem is therefore reduced to finding the minimum eigenvalue of the operator \( P^2 \). Its eigenvalues are \( \pi^2 m^2 (m = 1, 2, \ldots) \) and corresponding eigenfunctions are \( \phi_m(x) = 2\sqrt{2} \sin \pi m ((x + 1)/2) / C_m \) where \( C_m \) is a normalizing constant.
Note that careful treatment is required for the operator $P^2$ when it is treated as an operator on $L^2([-1, 1])$. When the function $f \in L^2([-1, 1])$ does not satisfy the Dirichlet condition $f(-1) = f(1) = 0$, the function $f$ has a discontinuity at 1 or $-1$ as an element of $L^2(\mathbb{R})$. Hence, the variance $\langle f | P^2 | f \rangle$ is infinite. For example, in the case of $f = \sqrt{1/2}$, the variance $\int_{-\infty}^{\infty} (\sin^2 y/y^2) \ (dy/2\pi)$ diverges. In this case, the limiting distribution is obtained concerning the variable $n(\hat{\theta} - \theta)/2 [1]$ whereas the mean square error goes to zero only in the order $1/n$, i.e. we have a quadratic speedup with respect to the limiting distribution, but no quadratic speedup with respect to mean square error. This fact is closely related to the divergence of the integral $\int_{-\infty}^{\infty} (\sin^2 y/y^2) \ (dy/2\pi)$.

5. Tail probability of the limiting distribution

In the asymptotic statistics, the behavior of the tail probability of the limiting distribution is one of the most important topics [28] because it yields the performance of interval estimation and the power of the one (or two)-sided test. Thus, we consider the tail probability of the limiting distribution $P^{m\phi}$.

In the independent and identically distributed (i.i.d.) case, the minimum tail probability and minimum variance among limiting distributions can be realized by the same Gaussian distribution. However, in our setting, the Gaussian distribution does not minimize the variance. Hence, it is not clear whether the minimum tail probability can be attained by the same distribution as that which minimizes the variance.

In the following, we evaluate the tail probability of the limiting distribution $P^{m\phi}$, which is acquired via the Fourier transforms

$$ P^{m\phi}(y) = |F(\phi_m)(y)|^2. $$

Since

$$ \int_{-1}^{1} C_m \phi_m(x)e^{ixy} \ dx 
= \frac{1}{\sqrt{2i}} \int_{0}^{1} e^{i(y+m\pi)x} - e^{i(y-m\pi)x} \ dx 
= \frac{1}{\sqrt{2i}} \left\{ \left[ \frac{e^{i(y+m\pi)x}}{i(y + m\pi)} \right]_{x=0}^{1} - \left[ \frac{e^{i(y-m\pi)x}}{i(y - m\pi)} \right]_{x=0}^{1} \right\} 
= \frac{1}{\sqrt{2}} (1 - (-1)^m e^{iy}) \left( \frac{1}{y + m\pi} - \frac{1}{y - m\pi} \right) 
= \frac{1}{\sqrt{2}} (-1 + (-1)^m e^{iy}) \frac{2m\pi}{y^2 - m^2\pi^2}, $$

the limiting distributions are

$$ P^{m\phi}(y) = |F(\phi_m)(y)|^2 = \frac{2m^2\pi (1 - (-1)^m \cos y)}{(y^2 - m^2\pi^2)^2 C_m}. $$

Thus, the tail probability of $P^{m\phi}$ decreases with order $O(y^{-4})$. In order to improve the tail probability, we focus on the well-known fact that the Fourier transform converts only a rapidly
Figure 2. The wave function minimizing the variance \( \phi_1 \) (thin dashed), the rapidly decreasing function \( g_3 \) whose support is included in \([-1, 1]\) (thin solid), prolate spheroidal wave function \( \psi_2 \) (thick dashed), and prolate spheroidal wave function \( \psi_{10} \) (thick solid).

descending function to a rapidly decreasing function. In our problem, the support of the original wave function \( f \) is included in \([-1, 1]\). Under this condition, \( f \) is a rapidly decreasing function if and only if \( f \) is a smooth function. Note that a rapidly decreasing function does not decrease ‘suddenly’; that is, smoothness is an essential requirement. For example, the function \( \phi_m \) is not smooth at \(-1\) or \(1\) (see figure 2). In the following, we construct a rapidly decreasing wave function \( f \) whose support is included in \([-1, 1]\). In this construction, smoothing at \(-1\) and \(1\) is essential.

First, functions \( g_0 \), \( g_1 \) and \( g_2 \) are defined by

\[
g_0(x) := \begin{cases} 
\frac{2 \exp(-1/x)}{\sqrt{x}}, & \text{if } x > 0, \\
0, & \text{otherwise},
\end{cases}
\]

\[
g_1(x) := g(x + 1),
\]

\[
g_2(x) := g(1 - x).
\]

Using these functions, we define a rapidly decreasing \( g_3 \) whose support is included in \([-1, 1]\), thus:

\[
g_3(x) = g_1(x)g_2(x)/C,
\]

where \( C \) is the normalizing constant. As can be checked numerically (see figure 4), this function improves the tail probability.

Now, we analyze the decreasing speed of the tail probabilities of \( P^{e^1([-R, R])} \). Their Fourier transformations are

\[
F(g_0)(y) = \frac{1}{\sqrt{2}} \frac{\exp(-\sqrt{2}|y|)}{\sqrt{|y|}} \exp \text{sgn}(y)i\left(\sqrt{2}|y| + \frac{\pi}{4}\right),
\]

\[
F(g_1)(y) = e^{-iy}F(g_0)(y),
\]
\[ F(g_2)(y) = -e^{-iy} F(g_0)(-y), \]
\[ F(g_3)(y) = \frac{1}{C} \sqrt{2\pi} F(g_1) \ast F(g_2)(y), \]
where \( F(g_1) \ast F(g_2) \) is the convolution of \( F(g_1) \) and \( F(g_2) \). When \( y \) is sufficiently large, \( |F(g_i)(y)|^2 \cong e^{-2\sqrt{2}\pi|y|} \) for \( i = 1, 2 \), i.e.
\[ \lim_{y \to \infty} -\frac{1}{\sqrt{|y|}} \log |F(g_i)(y)|^2 = 2\sqrt{2}, \quad i = 1, 2. \]

Then, as is shown in appendix B, we obtain
\[ \lim_{y \to \infty} -\frac{1}{\sqrt{|y|}} \log 2\pi |F(g_1) \ast F(g_2)(y)|^2 \geq 2\sqrt{2}. \] (12)

Therefore, there exists a function \( f \) such that the tail probability of \( P^f \) is exponentially small and the support is included in \([-1, 1]\). Note that the above wave function \( g_3 \) does not minimize the variance \( V(f) \). This fact tells us that the input state minimizing the variance is not optimal with respect to the tail probability of the limiting distribution. That is, the optimal input state depends on the choice of criterion.

Next, we consider maximization of the probability \( P^f([-R, R]) \). For this purpose, we denote the natural projection from \( L^2(\mathbb{R}) \) to \( L^2([-R, R]) \) by \( \Pi_R \). By using the operator \( F_R := F \dagger \Pi_R F \), this probability has the form
\[ \langle f | F_R | f \rangle. \]
That is, our aim is the following maximization:
\[ \max_{f \in L^2([-1,1])} \langle f | F_R | f \rangle = \max_{g \in L^2(\mathbb{R})} \frac{\langle g | \Pi_1 F_R \Pi_1 | g \rangle}{\|g\|^2} = \max_{g \in L^2(\mathbb{R})} \frac{\| \Pi_R F \Pi_1 g \|^2}{\|g\|^2}. \]

This problem is equivalent to evaluation of the maximum eigenvalue of \( \Pi_1 F_R \Pi_1 \).

Slepian and Pollak [20] showed that the eigenfunction \( \psi_R \) of \( \Pi_1 F_R \Pi_1 \) associated with the maximum eigenvalue can be obtained as the solution of a certain linear differential equation, called the prolate spheroidal wave function
\[ \frac{d}{dx} (1 - x^2) \frac{df}{dx} + (\xi(R) - R^2 x^2) f = 0, \]
where \( \xi(R) \) is chosen depending on the number \( R \).

Slepian [21] showed that the maximum eigenvalue \( \lambda(R) \) of \( \Pi_1 F_R \Pi_1 \) behaves as
\[ 1 - \lambda(R) \cong 4\sqrt{\pi} Re^{-2R} \left( 1 - \frac{3}{32R} + O(R^{-2}) \right). \] (13)

Footnote 4: For the relation between \( \xi(R) \) and \( R \), see Slepian and Pollak [20].
when $R$ is sufficiently large. The numerical calculation of this minimum probability
\[ \min_{f \in C([-1, 1])} P_f([-R, R]^c) \] is presented in figure 4. Thus, the minimum probability
\[ \min_{f \in C([-1, 1])} P_f([-R, R]^c) \] can be evaluated as
\[ \lim_{R \to \infty} \frac{-1}{R} \log \min_{f \in C([-1, 1])} P_f([-R, R]^c) = 2. \]

That is, the minimum tail probability \( \min_{f \in C([-1, 1])} P_f([-R, R]^c) \) goes to zero with exponential rate 2. This optimal value is attained when the input state is given by the eigenfunction \( \psi_R \) of
\[ P_1 F_R \Pi_1 \] associated with the maximum eigenvalue \( \lambda(R) \).

Now, we numerically compare the function \( \psi_R \) in the case \( R = 2, 10 \) with the functions \( \phi_2 \) and \( g_3 \). The density functions of the distributions \( P_{\phi_2} \), \( P_{g_3} \), \( P_{\psi_2} \) and \( P_{\psi_{10}} \) are plotted in figure 3. Their tail probabilities are plotted in figure 4. The tail probabilities \( P_{\psi_2}([-y, y]^c) \) (thick dashed) and \( P_{\psi_{10}}([-y, y]^c) \) (thick solid) attain the minimum tail probability \( \min_{f \in C([-1, 1])} P_f([-y, y]^c) \) only at 2 and 10, respectively. The distributions \( P_{\phi_2} \) and \( P_{\psi_2} \) are concentrated in the range \([-2, 2] \); however, their tail probabilities do not decrease as rapidly as those of the distributions \( P_{g_3} \) and \( P_{\psi_{10}} \). This comparison indicates that the optimizations of the concentration and the tail probability are not compatible. That is, the distributions of the Fourier transforms of the functions \( g_3 \) and \( \psi_{10} \) have a small tail probability (figure 3). These functions are smooth at \(-1 \) and 1. This indicates that smoothness is closely related to the tail probability.

Next, we generalize this problem slightly, i.e. we maximize the probability \( P_f([R_1, R_2]) \). In this case, the maximum value coincides with that of \( P_f([\frac{R_2 - R_1}{2}, \frac{(R_1 - R_2)}{2}]) \), and its maximum is attained by the function \( e^{\psi(R_1 + R_2)/2} \psi(R_2 - R_1)/2(x) \).

Since the function \( R \mapsto P_f([-R, R]) \) is a strictly monotonic increasing function, the inverse function \( \beta \mapsto R(\beta) \) is also a strictly monotonic increasing function. Thus,
\[ \min_{f \in C([-1, 1])} \min \{ R | P_f([-R, R]) \geq \beta \} = R(\beta). \]
Figure 4. Logarithm of tail probabilities $\log P_{\psi_1}([-y, y]^c)$ (thin dashed), $\log P_{\psi_2}([-y, y]^c)$ (thin solid), $\log P_{\psi_3}([-y, y]^c)$ (thick dashed) and $\log P_{\psi_4}([-y, y]^c)$ (thick solid), and logarithm of the minimum tail probability $\log \min_{f \in \mathcal{C}([-1, 1])} P_f([-y, y]^c)$ (thick dotted).

Further, the lhs coincides with

$$\min_{f \in \mathcal{C}([-1, 1])} \min \{ R | P_f([-R + a, R + a]) \geq \beta \}$$

for any real number $a$.

### 6. Interval estimation

Now, we treat the phase estimation problem with interval estimation. In interval estimation, given a confidence coefficient $\beta$, we estimate the confidence interval $[L, U]$, in which the unknown parameter $\theta$ is guaranteed to lie with probability $\beta$. Here, since our parameter space is the torus $\mathbb{R}/2\pi \mathbb{Z}$, a careful treatment is required for the confidence interval $[L, U]$. That is, for $L, U \in [0, 2\pi)$, the confidence interval $[L, U]$ is defined as a subset of $\mathbb{R}/2\pi \mathbb{Z}$ by

$$[L, U] := \begin{cases} [L, U], & \text{if } L < U, \\ [L, U + 2\pi], & \text{otherwise}, \end{cases}$$

and its width is defined by

$$|[L, U]| := \begin{cases} U - L, & \text{if } L < U, \\ U + 2\pi - L, & \text{otherwise}. \end{cases}$$

In interval estimation, the upper bound $U$ and the lower bound $L$ of the interval are chosen from the outcome $\omega$ obeying the distribution $P^{M^\theta}_{\tilde{a}^\theta}$ when the application of the unknown unitary $U^\theta_n$ is available. Since a smaller width $|[L(\omega), U(\omega)]|$ is better, we minimize the width $|[L(\omega), U(\omega)]|$ under the condition $P^{M^\theta}_{\tilde{a}^\theta} \{ \omega | \theta \in [L(\omega), U(\omega)] \} \geq \beta$ for any $\theta \in [0, 2\pi)$. That is, we consider

$$\min_{U, L, M^\theta, \tilde{a}^\theta} \left\{ \max_{\omega} |[L(\omega), U(\omega)]| P^{M^\theta}_{\tilde{a}^\theta} \{ \omega | \theta \in [L(\omega), U(\omega)] \} \geq \beta, \quad \forall \theta \in [0, 2\pi) \right\}$$

$$= \min_{U, L, M^\theta, \tilde{a}^\theta} \max_{\omega} \left( \max \left\{ |[L(\omega), U(\omega)]| \right\} P^{M^\theta}_{\tilde{a}^\theta} \{ \omega | \theta \in [L(\omega), U(\omega)] \} \geq \beta \right).$$

(14)
The value (14) has a mini-max form of the cost \( \max_{\omega, \theta} |[L(\omega), U(\omega)]| P^{M_{\theta, \hat{\omega}}} (\omega | \theta \in [L(\omega), U(\omega)]) \geq \beta \), which has a covariant form. Thus, we can restrict our measurement to covariant measurement (2). Hence, our problem is reduced to

\[
\min_{t, \hat{a}} \left( \max_{\omega} \left\{ |[L(\omega), U(\omega)]| P^{M_{t, \hat{a}}} (\omega | \theta \in [L(\omega), U(\omega)]) \geq \beta \right\} \right).
\]

However, since it is quite difficult to treat this optimization for finite \( n \), we consider the asymptotic setting as follows:

\[
\lim_{n \to \infty} n \min_{t, \hat{a}} \left( \max_{\omega} \left\{ |[L(\omega), U(\omega)]| P^{M_{t, \hat{a}}} (\omega | \theta \in [L(\omega), U(\omega)]) \geq \beta \right\} \right)
\]
\[
= \min_{f \in \mathcal{C}^2([-1, 1])} \max_{\omega} \left\{ |[L(\omega), U(\omega)]| P^{f} (\omega | \theta \in [L(\omega), U(\omega)]) \geq \beta \right\}
\]
\[
= \min_{f \in \mathcal{C}^2([-1, 1])} \min (2R | P^{f} ([−R, R]) \geq \beta)
\]
\[
= 2R(\beta),
\]

where \( R(\beta) \) is defined as the inverse function of \( R \mapsto P^{f} ([−R, R]) \) at the end of the previous section. This optimal value is attained when the input state constructed by the wave function \( |\psi(\theta)\rangle \) and the measurement is given by the covariant measurement (2) with the vector \( |t_0\rangle \). That is, there exists a pair of functions \( U \) and \( L \) such that \( |[L(\omega), U(\omega)]| \leq 2R(\beta) \) and \( P^{M_{t, \hat{a}}} (\omega | \theta \in [L(\omega), U(\omega)]) \geq \beta \). The optimal input state depends on the choice of the confidence coefficient \( \beta \).

7. Continuous case with single copy

Concerning estimation of unitary action, several papers (e.g. [11]) employ the Cramér–Rao approach. This approach is essentially equivalent to the minimization of variance at the one point with locally unbiased condition, which is discussed by several researchers, e.g. Giovannetti et al [29]. Hence, it is necessary to clarify the relation between our discussion and the Cramér–Rao lower bound.

As the first step for this comparison, we treat phase estimation for the continuous case with single copy, in which by inputting the wave function \( f \), we estimate the parameter \( \theta \) in a group-covariant model \( \rho_\theta = e^{i\theta \hat{Q}} |f\rangle \langle f| e^{-i\theta \hat{Q}} \) on the space \( L^2(\mathbb{R}) \).

It is known that when the shift-covariance condition is assumed for estimators, our estimator is restricted to measurement of the observable \( P \) [30]. Then, the outcome \( \hat{\theta} \) obeys the distribution \( P^{f} \), and the variance of the outcome is given by \( \langle f | \Delta P | f \rangle \), which is abbreviated by \( \langle \Delta P | \Delta P \rangle \).

If we can input any wave function \( f \), the variance can be reduced infinitesimally. Hence, it is natural to assume a constraint for an input wave function \( f \). Here, we assume that the potential is given as a monotonic function of the absolute value \( |Q| \). While we often assume a constraint on the average potential, we consider here a deterministic condition for the potential. That is, the wave packet of \( f \) is assumed to exist only in the region where the potential is less than a given constant. In the following, for simplicity of analysis, we assume that the input wave

New Journal of Physics 11 (2009) 043034 (http://www.njp.org/)
function belongs to $L^2([-1, 1])$. Hence, the discussion in sections 4 and 5 can be applied to this problem.

Here, it is meaningful to consider the relation with the Cramér–Rao bound. It is known in general that the Fisher information $J_\theta$ for a group-covariant model $\rho_\theta = e^{i\theta}O|f\rangle\langle f|e^{-i\theta}O$ is given by $J_\theta = \langle \Delta Q^2 \rangle$ because the symmetric logarithmic derivative (SLD) is given by $Q = \langle Q \rangle$ [30].

Since the operator $P$ has a commutation relation with $Q$, we have the Heisenberg limit $\langle \Delta P^2 \rangle \langle \Delta Q^2 \rangle \geq 1/4$, which is equivalent to the Cramér–Rao inequality:

$$\langle \Delta P^2 \rangle \geq \frac{1}{2} J_\theta^{-1}.$$  

In particular, if and only if $f$ is a squeezed state satisfying $\langle \Delta P^2 \rangle = c$ and $\langle \Delta Q^2 \rangle = c^{-1}$, the above inequality is achievable because its attainability is equivalent to that of $\langle \Delta P^2 \rangle \langle \Delta Q^2 \rangle \geq 1/4$. Thus, if $f$ is not a squeezed state, the Cramér–Rao lower bound $1/4 J_\theta^{-1}$ cannot be attained uniformly in the one-copy case. As is shown in the next section, our asymptotic case is essentially equivalent to the above group-covariant model under the restriction $\text{supp} f \subset [-1, 1]$.

8. Asymptotic Cramér–Rao lower bound

In this section, using the discussion in the above section, we treat the relation between our discussion and the Cramér–Rao lower bound. When we apply the sequence of protocols $\mathcal{M} := \{\vec{a}^n\}$, phase estimation can be treated as an estimation problem for the state family $\{|\phi_{\theta,n}\rangle \langle \phi_{\theta,n}|\theta \in [0, 2\pi]\}$, where $\langle \phi_{\theta,n} | = \sum_k a_k^n e^{i\theta} |k\rangle$. Let us calculate the SLD Fisher information. From the group covariance of the output state, it suffices to calculate the SLD Fisher information at $\theta = 0$. Let $|l_{\theta,n}\rangle := (1 - |\phi_{\theta,n}\rangle \langle \phi_{\theta,n}|)(\partial/\partial \theta)|\phi_{\theta,n}\rangle$. The SLD Fisher information $J_{0,n}$ is given by $(1/4) J_{0,n} = \langle l_{0,n} | l_{0,n} \rangle$.

$$\frac{J_{0,n}}{4} = \langle l_{0,n} | l_{0,n} \rangle = \langle \phi_{0,n} | \phi_{0,n} \rangle - |\langle \phi_{0,n} | \phi_{0,n} \rangle|^2 = \sum_{k=0}^n k^2 |a_k^n|^2 - \left( \sum_{k=0}^n k |a_k^n|^2 \right)^2$$

where $|\phi_{0,n}'\rangle = (\partial/\partial \theta)|\phi_{0,n}\rangle|_{\theta=0}$. Choosing a smooth function $f_n$ by (9), we have

$$\frac{J_{0,n}}{4(n+1)^2} = \sum_{k=0}^n (x_k)^2 |f_n(x_k)|^2 - \left( \sum_{k=0}^n (x_k) |f_n(x_k)|^2 \right)^2.$$

When $f_n$ converges to $f$,

$$\lim_{n \to \infty} \frac{J_{0,n}}{4(n+1)^2} = \int_{-1}^1 x^2 |f(x)|^2 \, dx - \left( \int_{-1}^1 x |f(x)|^2 \, dx \right)^2 = \langle f | \Delta Q^2 | f \rangle.$$
Since the variance of the limiting distribution is \( \langle f \mid \Delta P^2 \mid f \rangle \), we obtain the limiting distribution version of the Cramér–Rao inequality as
\[
\langle f \mid \Delta P^2 \mid f \rangle \geq \lim_{n \to \infty} \frac{1}{\frac{1}{4(n+1)^2} J_{0,n}} = \frac{1}{\langle f \mid \Delta Q^2 \mid f \rangle}.
\]
The equality holds if and only if the wave function \( f \) is a squeezed state, i.e. its Wigner function is a Gaussian distribution. However, since the support \( f \) belongs to \([-1, 1]\), the equality above cannot be attained. This fact indicates that the Cramér–Rao approach does not yield the attainable bound in the estimation of unitary action even in the asymptotic formulation, whereas this approach generally yields the attainable bound in the estimation of the quantum states. This point is the essential difference between state estimation and unitary estimation.

9. Conclusion

As a unified approach to asymptotic analysis of phase estimation, we have treated the limiting distribution of a sequence of estimators because we can recover various asymptotic performance measures of the estimation protocols from the limiting distribution.

As the first step, we found a one-to-one correspondence between a limiting distribution and a wave function on \( L^2([-1, 1]) \). That is, we have shown that any limiting distribution is given by the absolute square of the Fourier transform of a wave function \( f \in L^2([-1, 1]) \). Due to this correspondence, it is sufficient to optimize a distribution given in terms of the square of a Fourier transform on \( L^2([-1, 1]) \).

As a next step, minimization of the variance among the above distributions was dealt with by treating the Dirichlet problem in a similar way to Buzek et al [5]. We also considered its tail probability. In order to guarantee a small error probability outside the given interval, it is better that the limiting distribution be rapidly decreasing. However, it was established that the limiting distribution minimizing the variance is not necessarily rapidly decreasing. In order to construct such a limiting distribution, we employed a smoothing method in order to construct a rapidly decreasing function whose support is included in \([-1, 1]\). It was numerically checked that this function improves the tail probability remarkably.

Further, the tail probability for a given interval was minimized among these limiting distributions by employing Slepian and Pollak’s analysis of signal processing [20]. The optimal limiting distribution depends on the width of this interval. Using this optimization, we treated interval estimation in the asymptotic setting.

Next, we treated the relation with phase estimation in the continuous system with a one copy setting. In this case, Heisenberg’s uncertainty relation is equivalent to the Cramér–Rao inequality. Using this equivalent relation, we obtained the condition for attainability of the Cramér–Rao inequality. Further, we applied this relation to asymptotic analysis of the variance of the phase estimation. Then, we established that the Cramér–Rao bound cannot be attained in our framework.

Throughout these discussions, it has been established that optimization of asymptotic phase estimation cannot be characterized by a single parameter whereas this problem can be characterized by a single parameter, i.e. the variance, in state estimation of a single parameter model with a regularity condition due to asymptotic normality [15]–[17]. This property is the biggest difference between state estimation and phase estimation.
Indeed, a similar property can be expected for general unitary estimation. It is a future problem to investigate the limiting distribution in the estimation of unitary operators in a more general sense.

Acknowledgments

We thank Professor Toshiyuki Sugawa, Professor Fumio Hiai and Professor Fuminori Sakaguchi for discussions regarding the Fourier analysis. We also thank Professor Michele Mosca for his discussions regarding quantum circuits. We thank the referees and the editor for helpful comments concerning this manuscript.

This research was partially supported by a Grant-in-Aid for scientific research on the Priority Area ‘Deepening and Expansion of Statistical Mechanical Informatics (DEX-SMI)’, no. 18079014 and an MEXT Grant-in-Aid for Young Scientists (A) no. 20686026.

Appendix A. Elimination of multiplicity

The unitary operator $V^{\otimes n}_\theta$ can be written as the form

$$U^{n'}_\theta = \sum_{k=0}^{n} \sum_{j=1}^{m_k} e^{i(k-n/2)\theta} |k, j\rangle \langle k, j|$$

with multiplicity $m_k = \binom{n}{k}$. When the unitary operator $U^{n'}_\theta$ acts on the input state $\sum_{k=0}^{n} \sum_{j=1}^{m_k} a_{k,j} |k, j\rangle$, the final state is given by $\sum_{k=0}^{n} \sum_{j=1}^{m_k} e^{i(k-n/2)\theta} a_{k,j} |k, j\rangle$ where $a_k \equiv \sqrt{\sum_{j=1}^{m_k} |a_{k,j}|^2}$ and $|k\rangle \equiv \frac{1}{\sqrt{a_k}} \sum_{j=1}^{m_k} a_{k,j} |k, j\rangle$. Then, the estimation problem of $U^{n'}_\theta$ can be reduced to that of $U^n_\theta$ given in (1).

Further, when the input state is entangled with the ancilla, our estimation problem can be reduced to the estimation of unitary action with multiplicity. So, due to the above discussion, there is no improvement even if the input system is entangled with the ancilla.

Appendix B. Proof of (12)

Now, we prove (12). Assume that $y > 0$. For a given integer $N$

$$\frac{\sqrt{2\pi}}{C} \left| \int_{-\infty}^{\infty} F(g_1)(y')F(g_2)(y-y')dy' \right|$$

$$\leq \sum_{k=1}^{N} \frac{\sqrt{2\pi}}{C} \max_{y' \in [y(k-1)/N, y(k)/N]} |F(g_1)(y')| \cdot |F(g_2)(y-y')| \frac{y}{N}$$

$$+ \frac{\sqrt{2\pi}}{C} \int_{-\infty}^{0} |F(g_1)(y')| \cdot |F(g_2)(y-y')|dy'$$

$$+ \frac{\sqrt{2\pi}}{C} \int_{y}^{\infty} |F(g_1)(y')| \cdot |F(g_2)(y-y')|dy'.$$
Since $|F(g_1)(y')|$ is bounded

$$\sqrt{2\pi} \frac{C}{\int_{-\infty}^{0} |F(g_1)(y')| \cdot |F(g_2)(y-y')|dy'} \approx O \left( \int_{-\infty}^{0} |F(g_2)(y-y')|dy' \right) \approx O \left( e^{-\sqrt{2}\sqrt{\gamma}} \right).$$

Similarly

$$\sqrt{2\pi} \frac{C}{\int_{y}^{\infty} |F(g_1)(y')| \cdot |F(g_2)(y-y')|dy'} \approx O \left( e^{-\sqrt{2}\sqrt{\gamma}} \right)$$

Further

$$\sqrt{2\pi} \frac{C}{\max_{y' \in [y(k-1)/N,y(k)/N]} |F(g_1)(y')| \cdot |F(g_2)(y-y')| \frac{y}{N}} \approx O \left[ e^{-\sqrt{2}\sqrt{\gamma}y(k-1)/N} \cdot e^{-\sqrt{2}\sqrt{y-y(k)/N}} \right] \leq O \left( e^{-\sqrt{2}\sqrt{\gamma}(1-1/N)} \right).$$

Therefore

$$\sqrt{2\pi} \left| \int_{-\infty}^{\infty} F(g_1)(y') F(g_2)(y-y')dy' \right| \leq O \left( e^{-\sqrt{2}\sqrt{\gamma}(1-1/N)} \right).$$

Since $N$ is arbitrary

$$\sqrt{2\pi} \left| \int_{-\infty}^{\infty} F(g_1)(y') F(g_2)(y-y')dy' \right| \leq O \left( e^{-\sqrt{2}\sqrt{\gamma}} \right).$$

Taking the square, we obtain (12).

In the case of $y < 0$, we can show (12) by replacing $y'$ by $-y'$.

References

[1] Cleve R, Ekert A, Macchiavello C and Mosca M 1998 Quantum Algorithm Revisited Proc. Soc R. Lond. A 454 339
[2] Kitaev A Y, Shen A H and Vyalyi M N 2002 Classical and Quantum Computation (Graduate Studies in Mathematics vol 47) (Providence, RI: American Mathematical Society)
[3] Giovannetti V, Lloyd S and Maccone L 2004 Quantum-enhanced measurements: beating the standard quantum limit Science 306 1330–6
[4] Luis A and Perina J 1996 Phys. Rev. A 54 4564
[5] Buzek V, Derka R and Massar S 1999 Optimal quantum clocks Phys. Rev. Lett. 82 2207
[6] Bagan E, Baig M and Muñoz-Tapia R 2004 Entanglement-assisted alignment of reference frames using a dense covariant coding Phys. Rev. A 69 050303
[7] Bagan E, Baig M and Muñoz-Tapia R 2004 Quantum reverse-engineering and reference-frame alignment without nonlocal correlations Phys. Rev. A 70 030301

New Journal of Physics 11 (2009) 043034 (http://www.njp.org/)
[8] Chiribella G, D’Ariano G M, Perinotti P and Sacchi M F 2004 Efficient use of quantum resources for the transmission of a reference frame Phys. Rev. Lett. **93** 180503
[9] Chiribella G, D’Ariano G M and Sacchi M F 2005 Optimal estimation of group transformations using entanglement Phys. Rev. A **72** 042338
[10] Hayashi M 2006 Parallel treatment of estimation of SU(2) and phase estimation Phys. Lett. A **354** 183–9
[11] Imai H and Fujiwara A 2007 Geometry of optimal estimation scheme for SU(D) channels J. Phys. A: Math. Theor. **40** 4391–400
[12] Higgins B L, Berry D W, Bartlett S D, Wiseman H M and Pryde G J 2007 Entanglement-free Heisenberg-limited phase estimation Nature **450** 393–6
[13] Nagata T, Okamoto R, O’Brien J L, Sasaki K and Takeuchi S 2007 Beating the standard quantum limit with four-entangled photons Science **316** 726
[14] Kitaev A Y 1997 Quantum computations: algorithms and error correction Russ. Math. Surv. **52** 1191–249
[15] Guta M and Kahn J 2006 Local asymptotic normality for qubit states Phys. Rev. A **73** 052108
[16] Guta M, Janssens B and Kahn J 2008 Optimal estimation of qubit states with continuous time measurements Commun. Math. Phys. **277** 127–60
[17] Guta M and Jencova A 2007 Local asymptotic normality in quantum statistics Commun. Math. Phys. **276** 341–79
[18] Pope D, Wiseman H M and Langford N K 2004 Adaptive phase estimation is more accurate than nonadaptive phase estimation for continuous beams of light Phys. Rev. A **70** 043812
[19] Larsen R J and Marx M L 2005 *An Introduction to Mathematical Statistics and its Applications* (Boston, MA: Pearson Education)
[20] Slepian D and Pollak H O 1961 Prolate spheroidal wave functions, Fourier analysis and uncertainty-I Bell Syst. Tech. J. **40** 43–63
[21] Slepian D 1965 Some asymptotic expansions for prolate spheroidal functions J. Math. Phys. **44** 99–140
[22] Li L, Leong M, Ye P and Gan Y 2002 Electromagnetic radiation from a prolate spheroidal antenna enclosed in a confocal spheroidal radome IEEE. Trans. Antennas Propag. **50** 1525–33
[23] Nazmi P, Kapadia P and Dowden J 1993 A mathematical model of heat conduction in a prolate spheroidal coordinate system with applications to the theory of welding J. Phys. D: Appl. Phys. **26** 563–73
[24] HOLEVO A S 1979 Covariant measurements and uncertainty relations Rep. Math. Phys. **16** 385–400
[25] van Dam W, D’Ariano G M, Ekert A, Macchiavello C and Mosca M 2007 Optimal quantum circuits for general phase estimation Phys. Rev. Lett. **98** 090501
[26] Chiribella G, D’Ariano G M and Perinotti P 2008 Memory effects in quantum channel discrimination Phys. Rev. Lett. **101** 180501
[27] Coddington E A and Levinson N 1955 *Theory of Differential Equations* (New York: McGraw-Hill)
[28] Kelly F P (ed) 1994 *Probability, Statistics and Optimization: A Tribute to Peter Whittle (Wiley Series in Probability and Statistics)* (New York: Wiley)
[29] Giovannetti V, Lloyd S and Maccone L 2006 Quantum Metrology Phys. Rev. Lett. **96** 010401
[30] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland) (Originally published in Russian 1980)