Iterative phase estimation

C J O'Loan

School of Mathematics and Statistics, University of St Andrews, KY16 9SS, UK

E-mail: cjo2@st-and.ac.uk

Abstract

We give an iterative algorithm for phase estimation of a parameter θ, which is within a logarithmic factor of the Heisenberg limit. Unlike other methods, we do not need any entanglement or an extra rotation gate which can perform arbitrary rotations with almost perfect accuracy: only a single copy of the unitary channel and basic measurements are needed. Simulations show that the algorithm is successful. We also look at iterative phase estimation when depolarizing noise is present. It is seen that the algorithm is still successful provided the number of iterative stages is below a certain threshold.

PACS numbers: 03.65.Ta, 03.67.−a

1. Introduction

Phase estimation is of fundamental importance to quantum information and quantum computation. It is related to some very important problems such as estimating eigenvalues [1–4], the factoring and search algorithms [5, section 5.3], precision measurement of length and optical properties, and clock synchronization [6].

Suppose that we have a unitary matrix U_θ depending on an unknown parameter θ and that one of its eigenvectors $|u\rangle$ is completely known; furthermore U_θ acts on $|u\rangle$ in the following way: $U_\theta |u\rangle = e^{i2\pi\theta}|u\rangle$, where $\theta \in [0, 1)$. The task of phase estimation is to estimate the eigenvalue $e^{i2\pi\theta}$ and consequently θ, as accurately as possible. In this paper, we investigate phase estimation of a unitary matrix with known eigenvectors, which acts on a two-dimensional Hilbert space. In particular, we look at unitary matrices of the form

$$U_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i2\pi\theta} \end{pmatrix}, \tag{1}$$

where $\theta \in [0, 1)$. We will think of θ as being a point on a circle of unit circumference, and confidence intervals for θ as arcs on a circle of unit circumference, known as confidence arcs. We define the distance between an angle θ and an estimate $\hat{\theta}$ as

$$|\hat{\theta} - \theta|_1 = \min((\hat{\theta} - \theta)_{\mathrm{mod}\,1}, (\theta - \hat{\theta})_{\mathrm{mod}\,1}). \tag{2}$$

It is important to recognize that the angle θ is an angle on the circle, and so in this paper all arithmetic is modulo 1.

We now introduce the basic notions of quantum states and positive operator valued measures (POVMs). A quantum state is represented by a density matrix ρ and satisfies $\mathrm{tr}\{\rho\} = 1$, $\rho = \rho^\dagger$ and $\rho \geqslant 0$. In this paper, we will mainly be dealing with pure states. Any state which can be written as $\rho = |\psi\rangle\langle\psi|$ is said to be a pure state; we will often refer to a pure state by its state vector $|\psi\rangle$. Given a quantum state ρ, we can measure it using a POVM $M = \{M_m\}$, obtaining outcome m with probability given by the Born rule

$$p(m) = \mathrm{tr}\{\rho M_m\}. \tag{3}$$

A POVM $M = \{M_m\}$ satisfies $M_m = M_m^\dagger$, $M_m \geqslant 0$ and $\sum_m M_m = \mathbb{I}$.

We shall quantify performance of phase estimation schemes in terms of the expected fidelity $\langle F(U_{\hat\theta}, U_\theta)\rangle$ between $U_{\hat\theta}$ and U_θ. We use the cost function

$$1 - \langle F(U_{\hat\theta}, U_\theta)\rangle = 1 - \frac{\left\langle \left| \mathrm{tr}\left\{U_{\hat\theta}^{-1} U_\theta\right\}\right|^2\right\rangle}{4} \tag{4}$$

and look at its asymptotic scaling with n—the number of times that U_θ is used.

For a simple phase estimation approach where U_θ is used once on n identical copies of some input state (see section 2.1), we get $1 - \langle F\rangle = O(1/n)$. This is known as the *standard quantum limit* [6]. However, it has been shown that when n copies of the channel are available [7–9], we can get $1 - \langle F\rangle = O(1/n^2)$. This rate at which $1 - \langle F\rangle$ approaches zero is known as the *Heisenberg limit* [10] and cannot be beaten [8].

In this paper, we are interested in iterative phase estimation, when we have only a single copy of U_θ, similar to that of Kitaev [11], for which $1 - \langle F\rangle = O((\log n/n)^2)$. This is within a logarithmic factor of the Heisenberg limit. Note that there are other iterative phase estimation procedures with $1 - \langle F\rangle = O(1/n^2)$, but they require an extra rotation gate capable of doing arbitrary rotations to high precision (see section 2.5).

In section 2, a selection of different phase estimation schemes are described. In section 3, problems with some of the previous methods are explained. Section 4 contains our algorithm for phase estimation and a theoretical evaluation of its performance. In section 5, simulations are performed to check that our algorithm works. Section 6 looks at the performance of our phase estimation algorithm in the presence of depolarizing noise.

2. Phase estimation methods

2.1. Simple approach

A very simple method is to let U_θ act on the input state $|\psi_x\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$; the output state is $|\psi_\theta\rangle = 1/\sqrt{2}(|0\rangle + \mathrm{e}^{\mathrm{i}2\pi\theta}|1\rangle)$. If we measure in x that is we use the POVM $M_x = \{M_1 = |\psi_x\rangle\langle\psi_x|, M_0 = \mathbb{I} - M_1\}$, then we get outcome 1 with probability $p_x(1; \theta) = (1 + \cos(2\pi\theta))/2$ and outcome 0 with probability $p_x(0; \theta) = 1 - p_x(1; \theta)$. If we perform N measurements, we get an estimate $\cos(2\pi\hat\theta) = 2N_{x=1}/N - 1$ of $\cos(2\pi\theta)$, where $N_{x=1}$ is the number of times we obtain the outcome 1. If we measure in y, that is, use the POVM $M_y = \{M_1 = |\psi_y\rangle\langle\psi_y|, M_0 = \mathbb{I} - M_1\}$, where $|\psi_y\rangle = 1/\sqrt{2}(|0\rangle + \mathrm{i}|1\rangle)$, we observe outcome 1 with probability $p_y(1; \theta) = (1 + \sin(2\pi\theta))/2$ and outcome 0 with probability $p_y(0; \theta) = 1 - p_y(1; \theta)$. Performing N measurements, we get an estimate of $\sin(2\pi\theta)$; from estimates of $\cos(2\pi\theta)$ and $\sin(2\pi\theta)$ we are able to estimate θ.

2.2. Kitaev

As far as we are aware, Kitaev [11] gave the first l-stage iterative phase estimation procedure. (The number of stages l is chosen before hand and will depend on the precision desired and experimental limitations.) At the kth stage of Kitaev's procedure, U_θ acts 2^{k-1} times on a qubit, which is then measured. We perform some multiple of $\log(l/\epsilon)$ measurements of $(2^{k-1}\theta)_{\mathrm{mod}\,1}$. This ensures that we can 'localize each of the numbers $(2^{k-1}\theta)_{\mathrm{mod}\,1}$ in one of the 8 intervals $[(s-1)/8, (s+1)/8]$ $(s=0,\ldots,7)$ with error probability $\leqslant \epsilon/l$.' Using this information, an algorithm—which is not given—gives us an estimate $\hat\theta$ satisfying

$$\Pr\left(\left(\hat\theta - \frac{1}{2^{l+2}}, \hat\theta + \frac{1}{2^{l+2}}\right) \ni \theta\right) \geqslant 1 - \epsilon. \tag{5}$$

2.3. Rudolph and Grover

Rudolph and Grover [12] looked at the problem of transmitting a reference frame from Alice to Bob, which is linked to estimation of an unknown $U \in SU(2)$, parametrized by three parameters α, θ, ϕ. The scheme of Rudolph and Grover involves estimating the parameters α, θ, ϕ individually using the following l-stage iterative procedure. We set $\theta \in [0, 1)$ to have an infinite binary expansion $\theta = w_1 w_2 \ldots w_l \ldots$. At the kth stage a qubit is sent back and forth between Alice and Bob in such a way that, when Bob finally measures it, he obtains outcome 1 with probability $p_k(1; \theta) = (1 + \cos(2^k \pi \theta))/2$.

This is repeated a minimum of $N = 32 \log_2(2l/\epsilon)$ times [12], which ensures that Bob's estimate $\hat p_k(1; \theta)$ of $p_k(1; \theta)$ satisfies

$$\Pr\left(\left(\hat p_k - \frac{1}{4}, \hat p_k + \frac{1}{4}\right) \ni p_k\right) \geqslant 1 - \frac{\epsilon}{l}. \tag{6}$$

It is assumed that if $|\hat p_k - p_k| \leqslant 1/4$, then Bob can estimate the kth bit of θ correctly. If this is so, then from (6), the probability that Bob estimates the kth bit of θ correctly is at least $1 - \epsilon/l$, and the probability that he estimates all of the binary digits of θ correctly is at least $1 - \epsilon$. After l stages, we get an estimate $\hat\theta = \hat w_1 \hat w_2 \ldots \hat w_l$, satisfying

$$\Pr\left(\left(\hat\theta - \frac{1}{2^l}, \hat\theta + \frac{1}{2^l}\right) \ni \theta\right) \geqslant 1 - \epsilon. \tag{7}$$

A similar scheme is then used to estimate the parameters α and ϕ. The method of Rudolph and Grover has been used in [6] for the problem of clock synchronization.

2.4. Zhengfeng et al

Zhengfeng *et al* [13] highlighted two errors with the method of Rudolph and Grover: (i) knowing $|\hat\theta - \theta|_1 \leqslant 1/2^m$ does not imply that we know the first m bits of the binary expansion of θ—consider $\theta = 0.49$, $\hat\theta = 0.5$ and $m = 1$; (ii) the method is problematic (in the sense explained in section 3) for θ close to $1/2$.

Zhengfeng *et al* gave the following l-stage procedure. In the first stage, we let U_θ act on $|\psi_x\rangle$ and measure in x; we obtain outcome 1 with probability $p(1; \theta) = (1 + \cos(2\pi\theta))/2$. The state $U_\theta|\psi_x\rangle$ is measured some multiple of $\log(l/\epsilon)$ times (N), which gives an estimate $\hat\theta$ satisfying

$$\Pr\left(\left(\hat\theta - \frac{1}{12}, \hat\theta + \frac{1}{12}\right) \ni \theta\right) \geqslant 1 - \frac{\epsilon}{l}. \tag{8}$$

Having obtained an estimate $\hat{\theta}$

(1) if $\hat{\theta} \in [0, 5/12)$, define $r_1 = 2$ and $\nu_1 = 0$,
(2) if $\hat{\theta} \in [5/12, 7/12)$, define $r_1 = 3$ and $\nu_1 = 1$,
(3) if $\hat{\theta} \in [7/12, 1)$, define $r_1 = 2$ and $\nu_1 = 1$.

At the kth stage, we apply U_θ $r_1 r_2 \ldots r_{k-1}$ times. After measuring $U_\theta^{r_1 r_2 \ldots r_{k-1}} |\psi_x\rangle$ N times, we estimate $(r_1 r_2 \ldots r_{k-1}\theta)_{\mathrm{mod}\,1}$ and obtain r_k and ν_k in a similar way to that in which we obtained r_1 and ν_1. After l stages, we have $(r_1, \ldots, r_l, \nu_1, \ldots, \nu_l)$. Our final estimate of θ is

$$\hat{\theta} = \sum_{i=1}^{l} \frac{\nu_i}{\prod_{k=1}^{i} r_k}. \tag{9}$$

2.5. Dobšíček et al

A popular iterative estimation method is to take θ to have a binary expansion of length l plus some small remainder, that is, $\theta = w_1 w_2 \ldots w_l + \Delta$. The binary digits $w_1, \ldots, w_l$ are measured one at a time with a single measurement. This has been done in [14–17]. (Higgins *et al* [17] were the first to give, and carry out experimentally, a method of this form which achieves the Heisenberg limit.) We review the method as described by Dobšíček *et al* [15].

At the kth stage, we let $U_\theta^{2^{l-k+1}}$ act on one of two qubits; the other qubit is acted on by a Z-rotation gate $e^{i\alpha_k \sigma_z}$ before being measured—where $\alpha_0 = 0$ and α_k for $k = 2, \ldots, l$ depends on the results from the previous $k - 1$ stages. From this measurement, we get an estimate $\hat{w}_{l+1-k}$ of the $(l + 1 - k)$th binary digit. After l stages, we get an estimate $\hat{\theta} = \hat{w}_1 \hat{w}_2 \ldots \hat{w}_l$ of θ which satisfies

$$\Pr\left(\left(\hat{\theta} - \frac{1}{2^{l+1}}, \hat{\theta} + \frac{1}{2^{l+1}} \right) \ni \theta \right) \geqslant 0.81. \tag{10}$$

We can increase the probability that our final interval contains θ to $1 - \epsilon$ by either (a) increasing the number of rounds to $l' = l + \log(2 + 1/(2\epsilon))$ or (b) using $O(\log^2(1/\epsilon))$ extra measurements of the first few binary digits [15]. The method of Dobšíček *et al* has recently been carried out on experimental data in [18].

The method of Dobšíček *et al* [15] has been analysed when there are internal static imperfections and residual couplings between qubits [19]. It was shown that this type of noise is detrimental to the performance of Dobšíček's method; however, solutions were found to overcome this problem in [19].

3. Problems

There is nothing wrong with Kitaev's method of iterative estimation. However, he does not give an algorithm for (i) choosing which of the intervals contains $(2^{k-1}\theta)_{\mathrm{mod}\,1}$ with probability $1 - \epsilon/l$ and (ii) reconstructing θ given confidence intervals for $(2^{k-1}\theta)_{\mathrm{mod}\,1}$. As we will see in this section, there are gaps in the methods of Rudolph and Grover, and Zhengfeng *et al* for (i).

There are two main gaps in Rudolph and Grover's method, which we now explain. Firstly, $p_k(1; \theta) = (1 + \cos(2^k \pi \theta))/2$ is a multimodal function of θ. For example, $\theta = 3/4$ and $\theta' = 1/4$ give the same value of $p_1(1; \theta)$, even though they differ in the first binary digit. To overcome this, we need an estimate of $\sin(2\pi\theta)$ as well. This however is a trivial point and is easily overcome.

Secondly, if $\theta = 1/2 \pm \delta$, where δ is small, we require a large number of measurements to determine the first bit of θ correctly with high probability. If we do make a mistake then

for our final estimate $\hat{\theta}$, we will have $|\hat{\theta} - \theta|_1 \geqslant \delta$. This problem, which occurs for θ close to $1/2$, was pointed out by Zhengfeng *et al.*

A similar problem also occurs for $\theta = 0 \pm \delta$. Because of this, we will encounter difficulties in estimating the kth bit of θ whenever $(2^{k-1}\theta)_{\text{mod}\,1} \approx 0$, $(2^{k-1}\theta)_{\text{mod}\,1} \approx 1$ or $(2^{k-1}\theta)_{\text{mod}\,1} \approx 1/2$. However, it may also be possible to overcome this issue using extra rotation gates in these cases.

There are also gaps with the method of Zhengfeng *et al* [13]. Firstly, like Rudolph and Grover, they overlook the fact that $p_1(1; \theta) = (1 + \cos(2\pi\theta))/2$ is bimodal. Secondly, the accuracy of their final estimate relies on the assumption that if $|\hat{\theta} - \theta|_1 \leqslant 1/12$ and $\hat{\theta} \in [0, 5/12)$, then $\theta \in [0, 1/2)$. This is not true, as we could have $\theta = -1/12 \notin [0, 1/2)$. Similarly, they assume that if $|\hat{\theta} - \theta|_1 \leqslant 1/12$ and $\hat{\theta} \in [7/12, 1)$, then $\theta \in [1/2, 1)$, which again is not true, as we could have $\theta = 1/12 \notin [1/2, 1)$. Again we will get problems at the kth stage if $(r_1 \ldots r_{k-1}\theta)_{\text{mod}\,1} \approx 0$ or $(r_1 \ldots r_{k-1}\theta)_{\text{mod}\,1} \approx 1$.

4. Our approach

This section contains our method for phase estimation. Firstly, we describe the method for going from confidence arcs for θ, $(2\theta)_{\text{mod}\,1}$, $(4\theta)_{\text{mod}\,1}, \ldots, (2^{l-1}\theta)_{\text{mod}\,1}$, of length $1/3$ and coverage probability at least $1 - \epsilon/l$, to a confidence arc for θ of length $1/(2^{l-1} \times 3)$ and coverage probability at least $1 - \epsilon$. Secondly, we describe how to get a confidence arc for $(2^{k-1}\theta)_{\text{mod}\,1}$, of length $1/3$ and coverage probability $1 - \epsilon/l$. Thirdly, we use Hoeffding's inequality to calculate the number of measurements needed at each stage. Finally, we show that $1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle = O((\log n/n)^2)$.

4.1. The iterative phase estimation algorithm

In this section, we introduce our method for computing a confidence arc for θ of length $1/(2^{l-1} \times 3)$ and coverage probability $1 - \epsilon$. First we give an intuitive approach using examples. For computational simplicity, we look at confidence arcs of length 0.3 and coverage probability 1. L_k and J_k will denote confidence arcs for $(2^{k-1}\theta)_{\text{mod}\,1}$ and $2^{k-1}\theta$, respectively, of length 0.3 and coverage probability 1. (In our more general algorithm L_k and J_k will have length $1/3$ and coverage probability at least $1 - \epsilon/l$.) For the examples, we choose $l = 3$.

4.1.1. Example. Suppose that after doing some measurements of $U_\theta|\psi_x\rangle$, $U_\theta^2|\psi_x\rangle$ and $U_\theta^4|\psi_x\rangle$, we find

$$L_1 = [0.6, 0.9] \ni \theta \tag{11}$$

$$L_2 = [0.3, 0.6] \ni (2\theta)_{\text{mod}\,1} \tag{12}$$

$$L_3 = [0.8, 1.1] \ni (4\theta)_{\text{mod}\,1}. \tag{13}$$

It follows from (11) that

$$2L_1 = [1.2, 1.8] \ni 2\theta. \tag{14}$$

Using (12) and (14), it follows that

$$J_2 = [1.3, 1.6] \ni 2\theta. \tag{15}$$

From (15) we know that

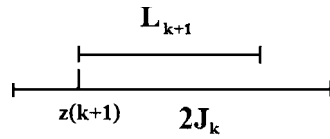$$2J_2 = [2.6, 3.2] \ni 4\theta. \tag{16}$$
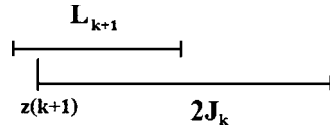
Figure 1. Situation 1.

Figure 2. Situation 2.

Using (13) and (16), we get

$$J_3 = [2.8, 3.1] \ni 4\theta. \tag{17}$$

Using confidence arcs (11), (12) and (13) for θ, $(2\theta)_{\text{mod}\,1}$ and $(4\theta)_{\text{mod}\,1}$, respectively, of length 0.3 and coverage probability 1, we have derived a confidence arc (17) for 4θ of length 0.3 and coverage probability 1. From this, we get a confidence arc for θ of length $0.3/2^{3-1} = 0.075$ and coverage probability 1, namely

$$(1/4)J_3 = [0.7, 0.775] \ni \theta.$$

Remember that we are looking at confidence arcs on a circle. On the circle the arc [1.2, 1.8] is equivalent to the arc [0.2, 0.8], as is [2.2, 2.8], [3.2, 3.8], ... Similarly [2.6, 3.2] is equivalent to [0.6, 1.2].

We define the symbol $\subset_1$ to mean that a confidence arc on the circle, is a subset of another confidence arc on the circle. Similarly, we define $\in_1$ to mean that a point is contained within an arc on the circle, e.g. $0.3 \in_1 [1.2, 1.8]$. The previous example was rather simple in that $[0.3, 0.6] \subset_1 [1.2, 1.8]$ and $[0.8, 1.1] \subset_1 [2.6, 3.2]$. In general, we cannot assume that $L_{k+1} \subset_1 2J_k$.

4.1.2. General algorithm. Our confidence arcs are now of length $1/3$ rather than 0.3. Let us put

$$L_k = [x(k), x(k) + 1/3], \qquad x(k) \in [0, 1),$$
$$J_k = [z(k), z(k) + 1/3].$$

As in section 4.1.1 we use $2J_k$ and L_{k+1} to find a confidence arc J_{k+1}. We insist that $J_{k+1} \subset 2J_k$. For this, we require that $z(k + 1) \in [2z(k), 2z(k) + 1/3]$. Assuming that $J_k \ni 2^{k-1}\theta$ and $L_{k+1} \ni (2^k\theta)_{\text{mod}\,1}$, then there are three possibilities. For each possibility we give a figure, with a small vertical line representing the choice of the lower bound for J_{k+1} (figures 1– 3). (Note that $J_1 = L_1$.)

(i) The simplest possibility is that $L_{k+1} \subset_1 2J_k$. This occurs when $(x(k + 1) - 2z(k))_{\text{mod}\,1} \in [0, 1/3)$. We choose J_{k+1} to have lower boundary $z(k+1) = 2z(k)+(x(k+1)-2z(k))_{\text{mod}\,1}$.
(ii) Another possibility is that $x(k + 1) \not\in_1 2J_k$ but $x(k + 1) + 1/3 \in_1 2J_k$. This occurs when $(x(k + 1) - 2z(k))_{\text{mod}\,1} \in [2/3, 1)$. In this case, we take the lower boundary of J_{k+1} to be $z(k + 1) = 2z(k)$.
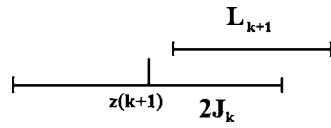
Figure 3. Situation 3.

(iii) The final possibility is that $x(k+1) \in_1 2J_k$ but $x(k+1) + 1/3 \notin_1 2J_k$. This occurs when $(x(k+1) - 2z(k))_{\text{mod }1} \in [1/3, 2/3)$. In this case, we take the lower boundary of J_{k+1} to be $z(k+1) = 2z(k) + 1/3$.

Using this iterative scheme, we end up with the confidence arc $J_l = [z(l), z(l) + 1/3]$ for $2^{l-1}\theta$ with coverage probability 1. This immediately gives us a confidence arc for θ of length $1/(2^{l-1} \times 3)$, namely $(1/2^{l-1})J_l = [z(l)/2^{l-1}, (z(l) + 1/3)/2^{l-1}]$. We take the middle of this interval modulo 1 as our estimate $\hat{\theta}$ of θ, i.e. $((z(l) + 1/6)/2^{l-1})_{\text{mod }1}$.

The final confidence arc for θ of length $1/(2^{l-1} \times 3)$ contains θ if $L_k \ni (2^{k-1}\theta)_{\text{mod }1}$, for every $k = 1, \ldots, l$. If each L_k has coverage probability at least $1 - \epsilon/l$, the probability that every L_k contains $(2^{k-1}\theta)_{\text{mod }1}$ is at least $1 - \epsilon$.

4.2. Finding L_k

Here we give the details of calculating the confidence arcs L_k for $(2^{k-1}\theta)_{\text{mod }1}$ of length $1/3$ and coverage probability at least $1 - \epsilon/l$. First we show how a confidence arc of length $1/3$ is computed, then we show how to make the coverage probability at least $1 - \epsilon/l$. We shall look at finding a confidence arc for θ. The analysis is exactly the same as for $(2^{k-1}\theta)_{\text{mod }1}$ except in the latter case where we let U_θ act 2^{k-1} times on the same $|\psi_x\rangle$.

We let U_θ act on $|\psi_x\rangle$ and measure in x. We obtain outcome 1 with probability $p_x(1; \theta) = (1 + \cos(2\pi\theta))/2$. We measure $U_\theta|\psi_x\rangle$ a total of N times and obtain outcome 1 $N_{x=1}$ times. We then have an estimate $2N_{x=1}/N - 1$ of $\cos(2\pi\theta)$.

We let U_θ act on $|\psi_x\rangle$ and measure in y. We obtain outcome 1 with probability $p_y(1; \theta) = (1 + \sin(2\pi\theta))/2$. We measure $U_\theta|\psi_x\rangle$ in y a total of N times and obtain outcome 1 $N_{y=1}$ times. We then have an estimate $2N_{y=1}/N - 1$ of $\sin(2\pi\theta)$. We get an estimate

$$\hat{\theta} = \frac{1}{2\pi}\left(\text{atan2}\left(\frac{2N_{y=1}}{N} - 1, \frac{2N_{x=1}}{N} - 1\right)\right)_{\text{mod }2\pi}$$

of θ. We can construct L_1 as

$$L_1 = \left(\left(\hat{\theta} - \frac{1}{6}\right)_{\text{mod }1}, \left(\hat{\theta} - \frac{1}{6}\right)_{\text{mod }1} + \frac{1}{3}\right).$$

More generally, given an estimate $(2^{k-1}\hat{\theta}_k)_{\text{mod }1}$ of $(2^{k-1}\theta)_{\text{mod }1}$, we get the confidence arc

$$L_k = \left(x(k), x(k) + \tfrac{1}{3}\right),$$
$$x(k) = \left((2^{k-1}\hat{\theta}_k)_{\text{mod }1} - \tfrac{1}{6}\right)_{\text{mod }1}.$$

It is shown in the appendix that if

$$\left|\frac{N_{x=1}}{N} - p_x(1; \theta)\right| \leqslant 0.306 \tag{18}$$

and

$$\left|\frac{N_{y=1}}{N} - p_y(1; \theta)\right| \leqslant 0.306, \tag{19}$$

then $L_1 \ni \theta$. It follows that if

$$\Pr\left(\left|\frac{N_{x=1}}{N} - p_x(1;\theta)\right| \leqslant 0.306\right) \geqslant \sqrt{1 - \frac{\epsilon}{l}} \tag{20}$$

and

$$\Pr\left(\left|\frac{N_{y=1}}{N} - p_y(1;\theta)\right| \leqslant 0.306\right) \geqslant \sqrt{1 - \frac{\epsilon}{l}}, \tag{21}$$

then

$$\Pr\left(L_1 \ni \theta\right) \geqslant 1 - \frac{\epsilon}{l}. \tag{22}$$

An analogous result holds for $L_k, k = 2, \ldots, l$. It is shown below that if $N = 5.34 \log(4l/\epsilon)$ then (20) and (21) hold.

4.3. Number of measurements needed

The Hoeffding inequality [20] will be used.

Theorem 1. *Given independent random variables $X_1, \ldots, X_n$ with $a_i \leqslant X_i \leqslant b_i$, then the following inequality holds for the sum $S_n = \sum_{i=1}^{n} X_i$:*

$$\Pr\left(|S_n - E[S_n]| \geqslant nt\right) \leqslant 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{23}$$

The observed measurement outcomes from a single measurement in x are independent random variables X_i with $0 \leqslant X_i \leqslant 1$, and for which $S_n = N_{x=1}$ and $E[S_n] = np_x(1;\theta)$. Using (23), it is straightforward to show that

$$\Pr\left(\left|\frac{N_{x=1}}{N} - p_x(1;\theta)\right| \geqslant t\right) \leqslant 2\exp(-2Nt^2). \tag{24}$$

From (24), it can be shown that (20) holds if

$$N = 5.34 \ln\left(\frac{4l}{\epsilon}\right) \tag{25}$$

measurements in x are performed at each stage. The analysis is exactly the same for measurements in y, and so a total number of

$$N_{\text{tot}} = 10.68 \ln\left(\frac{4l}{\epsilon}\right) \tag{26}$$

measurements are required at each stage. This ensures that (20) and (21) hold, and consequently (22) holds.

4.4. The behaviour of the fidelity

We now see how $1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle$ scales with the number of times U_θ is used. As in [12], we look at the worst-case value of $1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle$. That is, if the final confidence arc does not contain θ then $\hat{\theta} = (\theta + 1/2)_{\text{mod}\,1}$, and if it does then θ lies on the boundary of the confidence arc, i.e. $|\hat{\theta} - \theta|_1 = 1/(2^l \times 3)$. This gives

$$1 - \langle F(U_{\hat{\theta}}, U_\theta)\rangle \leqslant 1 - \left((1 - \epsilon)\frac{1 + \cos(2\pi/(2^l \times 3))}{2} + \epsilon \times 0\right)$$

$$\approx \epsilon + \frac{\pi^2}{2^{2l} \times 9} - \frac{\epsilon\pi^2}{2^{2l} \times 9}.$$

If we choose $\epsilon = 1/2^{2l}$, then $1 - \langle F(U_{\hat{\theta}}, U_\theta) \rangle = O(1/2^{2l})$. This requires a total of

$$N_{\text{tot}} = 10.68 \ln(4l \times 2^{2l}) \tag{27}$$

measurements at each stage. The number of times U_θ is used is $n = N_{\text{tot}}(2^l - 1)$, and so $1/2^l \approx N_{\text{tot}}/n$. The number of measurements, (27), made at each stage is $O(l)$; noting that $\log n$ is also $O(l)$, it follows that

$$1 - \langle F(U_{\hat{\theta}}, U_\theta) \rangle = O\left(\left(\frac{\log n}{n}\right)^2\right). \tag{28}$$

5. Simulations

The analysis in section 4.4 concentrated on optimizing the worst-case asymptotic scaling of $1 - \langle F \rangle$ with respect to n. The cost function $1 - \langle F \rangle$ is very sensitive to outliers. A large number of measurements, (27), were chosen so that the probability that the final interval did not cover θ was $1/2^{2l}$. This ensured that $1 - \langle F \rangle$ was within a logarithmic factor of the Heisenberg limit. Rather than choosing N_{tot} large to remove the large contribution of outliers, an experimenter may be happy enough if the probability that θ is contained by his final confidence arc is greater than some value. This approach does not sacrifice precision, but rather an unnecessarily large coverage probability.

In this section, we give a table of simulated results, for different numbers of iterative stages, and different numbers of measurements at each stage. From this table we show how to calculate a confidence interval for the coverage probability. An experimenter who wants a confidence arc for θ of certain length and coverage probability could look at the table and find the number of measurements needed to achieve this.

Simulations are performed with the computer package MAPLE. A value for the parameter $\theta \in [0, 1)$ is given by a random variable with a uniform distribution. Measurement results can be simulated, since the number of times outcome 1 is observed has a binomial distribution. For example, at the kth iterative stage, measuring in x, $N_{x=1} \sim \text{Bin}(N, (1 + \cos(2^k \pi \theta))/2)$. From the simulated results of measurements in x and y for stages $1, \dots, l$, an estimate of θ is obtained using the iterative algorithm given in section 4.1. We can then test whether our final confidence arc contains θ. This is done for 100 000 randomly chosen θ, and the number of times that θ is contained by the final confidence arc recorded.

For most recent iterative schemes the total number of iterations is reasonably small: 6 in [17] and 7 in [18]. We look at simulations with the number of iterations varying between 6 and 9. Table 1 gives the number of times the final confidence arc contains the true value of θ.

Using the above simulations the coverage probability can be estimated, i.e. the probability that, using the iterative algorithm, the known true value θ is contained in our final confidence interval.

Suppose the true (unknown) coverage probability is p. For the ith trial put

$$W_i = 1 \quad \text{if interval covers } \theta$$
$$= 0 \quad \text{if not.}$$

Then $W_1, \dots, W_M$ are independent identically distributed Bernoulli random variables, i.e. $W_i \sim \text{Bin}(1, p)$. Thus

$$W_1 + \cdots + W_M \sim \text{Bin}(M, p).$$

Table 1. Numbers of trials out of 100 000 with $|\hat{\theta} - \theta| \leqslant 1/(2^l \times 3)$.

N_{tot}	Number of iterative stages (l)			
	6	7	8	9
20	99 792	99 729	99 747	99 712
30	99 993	99 987	99 982	99 978
40	99 999	100 000	99 998	99 999
50	100 000	100 000	99 999	100 000

If m out of M intervals cover θ then p is estimated by m/M. An approximate 95% confidence interval for p is

$$\frac{m}{M} \pm 1.96\sqrt{\frac{m/M\,(1 - m/M)}{M}}.$$

The longest confidence interval (0.000 66) is that for using nine iterative stages and a total of 20 measurements at each stage. Using the half-length of this confidence interval, we can compute a confidence interval from the results given in table 1 with coverage probability at least 95%:

$$\frac{m}{100\,000} \pm 0.000\,33.$$

If an experimenter is content with a confidence arc of length no smaller than $1/(2^9 \times 3) = 1/1536$ and estimated coverage probability no greater than 99.6% then he need perform no more than 20 measurements at each stage. If the experimenter wanted to use even less measurements he could produce his own table of simulated results possibly even varying the number of measurements performed at each stage.

6. The noisy case

It is known that when even a small amount of noise is present, the performance of phase estimation schemes is greatly reduced [21, 22]. This section investigates the performance of the iterative estimation algorithm when depolarizing noise is present. The channel

$$\rho_0 \mapsto (1 - r)U_\theta \rho_0 U_\theta^\dagger + \frac{r}{2}\mathbb{I}_2, \qquad 0 < r < 1, \tag{29}$$

is considered, where U_θ is the same as before, (1), and $\rho_0 = |\psi_x\rangle\langle\psi_x|$. (The channel (29) is identical to $U_\theta \rho_0 U_\theta^\dagger$ undergoing phase damping with $\lambda = r(2 - r)$ [5, p 383].) Ji *et al* [13] gave the very interesting result that if $r > 0$, then the optimal asymptotic rate at which $1 - \langle F(U_{\hat{\theta}}, U_\theta) \rangle$ approaches zero is given by the standard quantum limit.

The whole point of using an iterative scheme is that the distinguishability of θ from $\cos(n2\pi\theta)$, with $n \gg 1$, is considerably greater than from $\cos(2\pi\theta)$. One measure of distinguishability is the Fisher information. Given a family of probability distributions with density functions $p(x; \theta)$, the Fisher information is defined as

$$F_\theta^M \equiv \int p(x; \theta)\left(\frac{\partial \ln p(x; \theta)}{\partial \theta}\right)^2 \mathrm{d}x \tag{30}$$

$$= \int \frac{1}{p(x; \theta)}\left(\frac{\partial p(x; \theta)}{\partial \theta}\right)^2 \mathrm{d}x. \tag{31}$$

Intuitively the Fisher information tells us the amount of 'information' about a parameter contained in a probability distribution. The Symmetric Logarithmic Derivative (SLD) quantum information H_θ tells us the maximal attainable Fisher information obtained from measuring a state depending on an unknown parameter [23], i.e.

$$F_\theta^M \leqslant H_\theta. \tag{32}$$

The SLD quantum information is defined in terms of the SLD quantum score λ_θ as

$$H_\theta = \text{tr}\{\lambda_\theta \rho_\theta \lambda_\theta\},$$

where λ_θ is any self-adjoint solution of the matrix equation

$$\frac{d\rho_\theta}{d\theta} = \frac{1}{2}(\rho_\theta \lambda_\theta + \lambda_\theta \rho_\theta).$$

To measure distinguishability, per use of the channel, the quantity F_θ^M/m will be used, where m is the number of times U_θ acts on the same input state.

If there is no noise, and the experimenter lets U_θ act m times on the input state and measures in x, then outcome 1 is observed with probability $p_x(1;\theta) = (1 + \cos(m2\pi\theta))/2$ and 0 with probability $p_x(0;\theta) = 1 - p_x(1;\theta)$. The Fisher information from this measurement is $F_\theta^{M_x} = 4\pi^2 m^2$, which is equal to the SLD quantum information. Measuring in y gives the same Fisher information. Thus, $F_\theta^{M_x}/m = F_\theta^{M_y}/m = 4\pi^2 m$. At the kth stage of the iterative procedure, we let U_θ act $m = 2^{k-1}$ times on the input state, and so $F_\theta^{M_x}/m = F_\theta^{M_y}/m = \pi^2 2^{k+1}$. Thus, F_θ^M/m (where M is an arbitrary measurement in x or y) increases exponentially with k.

In the noisy case, letting U_θ act m times on the output state and measuring in x, outcome 1 is observed with probability $p_x(1;\theta) = (1 + (1-r)^m \cos(m2\pi\theta))/2$ and 0 with probability $p_x(0;\theta) = 1 - p_x(1;\theta)$. Measuring in y, outcome 1 is observed with probability $p_y(1;\theta) = (1 + (1-r)^m \sin(m2\pi\theta))/2$ and 0 with probability $p_y(0;\theta) = 1 - p_y(1;\theta)$. This gives

$$F_\theta^{M_x} = \frac{4\pi^2 m^2 (1-r)^{2m} \sin^2(2m\pi\theta)}{1 - (1-r)^{2m} \cos^2(2m\pi\theta)}$$

$$F_\theta^{M_y} = \frac{4\pi^2 m^2 (1-r)^{2m} \cos^2(2m\pi\theta)}{1 - (1-r)^{2m} \sin^2(2m\pi\theta)}$$

$$H_\theta = 4\pi^2 m^2 (1-r)^{2m}.$$

Note that

$$F_\theta^{M_x} + F_\theta^{M_y} \approx H_\theta.$$

Thus, measuring both in x and y, the average Fisher information from a single measurement M is approximately $H_\theta/2$.

The maximal value of F_θ^M/m, taken over m, will occur close to the maximal value of H_θ/m. When $r > 0$, H_θ/m, and hence F_θ^M/m, do not increase indefinitely with m. Instead it reaches its maximum at

$$m = -\frac{1}{2\log(1-r)}, \tag{33}$$

after which it decreases. When r is small, this maximum is obtained at

$$m \approx \frac{1}{2r}. \tag{34}$$

The number of stages that can be performed, for small r, such that H_θ/m, and hence F_θ^M/m, increases at each stage is approximately $l \approx -\log_2 r$. A consequence of this is that estimation
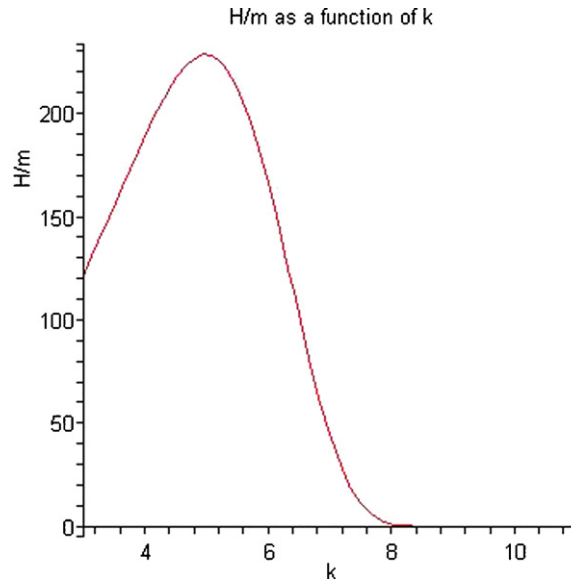
Figure 4. H_θ/m at the kth iterative stage, with $r = 2^{-5}$.

(This figure is in colour only in the electronic version)

Table 2. Numbers of trials out of 100 000 with $|\hat\theta - \theta|_1 \leqslant 1/(2^l \times 3)$, with $N_{tot} = 30$.

	Number of iterative stages (l)					
r	4	5	6	7	8	9
2^{-4}	98 290	88 340	60 423	32 445	16 059	8 042
2^{-5}	99 804	98 408	88 537	61 293	32 756	16 460
2^{-6}	99 967	99 807	98 430	88 708	61 148	32 595
2^{-7}	99 985	99 955	99 802	98 476	88 895	61 699
2^{-8}	99 988	99 977	99 962	99 812	98 467	88 864

close to the Heisenberg limit is not possible, asymptotically, when there is any depolarizing noise. This gives an alternative insight into the result of Ji *et al* [13].

Figure 4 gives H_θ/m at the kth iterative stage when $r = 2^{-5}$. It can be seen that H_θ/m increases up to $k = 5$, decreases slightly near $k = 6$ and falls rapidly for $k > 6$. Other figures not included here give similar information, showing H_θ/m increasing up to $k = -\log_2 r$, and decreasing rapidly for $k > -\log_2 r$.

Table 2 contains the results of simulations, for magnitudes of noise $r = 2^{-4}, 2^{-5}, \ldots, 2^{-8}$ and total number of iterative stages $l = 4, \ldots, 9$. Consider the diagonal of table 2, from $r = 2^{-4}, l = 4$ to $r = 2^{-8}, l = 8$. This corresponds to the experimenter performing $l = -\log_2 r$ iterative stages, which involves going up to the iterative stage at which F_θ^M/m is maximized. Similarly, the diagonal from $r = 2^{-4}, l = 5$ to $r = 2^{-8}, l = 9$ corresponds to the experimenter performing $l = -\log_2 r + 1$ iterative stages, etc. It is interesting to note that when $l > -\log_2 r$, there is a significant decrease in the number of confidence intervals containing θ. If the experimenter performs $l = -\log_2 r$ iterative stages then the final confidence interval contains θ approximately 98% of the time; for $l = -\log_2 r + 1$ iterative stages, the coverage

probability decreases to approximately 89%. For $l = -\log_2 r + 2$ iterative stages, the coverage probability is approximately 61%—a considerable drop in performance. Simulations using more measurements at each stage have given similar results.

It is interesting to see that the drop off in performance, in terms of the coverage probability, occurs at the same point as the drop in performance as measured by H_θ/m, and consequently F_θ/m, as seen in figure 4.

We suggest, more generally, that for the channel (29) the optimum number of iterative stages, where at the kth stage U_θ is used 2^{k-1} times, is $l = \lfloor -\log_2 r \rfloor$.

A related question was considered in [24], where the 'stopping point', was N the number of entangled photons to be included in the NOON input states. Rubin and Kaushik found that the optimal precision in measurement occurred for $N = 1.279/L$, where L is the magnitude of loss (analogous to the point, $n' = 1/(2r)$, at which F_θ^M/m is maximized).

7. Discussion

After completing this work, we were made aware of similar work already done by Higgins *et al* [25]. In [25], it was shown that the logarithmic factor can be removed, thus achieving the Heisenberg limit. Also, an experimental demonstration was given. However, an explicit algorithm which allows experimenters to implement this method was not given. Furthermore, the problem of noise was not dealt with.

8. Conclusion

In this paper, we have shown that there are gaps in the iterative phase estimation schemes of [12, 13].

We have shown how to compute confidence arcs for $\theta, (2\theta)_{\mathrm{mod}\,1}$, $(4\theta)_{\mathrm{mod}\,1}, \ldots, (2^{l-1}\theta)_{\mathrm{mod}\,1}$, of length $1/3$ and coverage probability at least $1 - \epsilon/l$. The main contribution of this paper has been to give an explicit algorithm which uses these confidence arcs to obtain a confidence arc for θ of length $1/(2^{l-1} \times 3)$ and coverage probability $1 - \epsilon$. Choosing $\epsilon = 1/2^{2l}$ gives $1 - \langle F(U_{\hat\theta}, U_\theta) \rangle = O((\log n/n)^2)$, i.e. within a logarithmic factor of the Heisenberg limit. The advantage of our scheme is that unlike other iterative phase estimation methods, such as [15, 17], it does not require an extra rotation gate capable of doing arbitrary rotations with almost perfect accuracy. Thus, our scheme has a simpler experimental setup and less potential for error.

Using computer simulations we have shown that the algorithm is successful. We have suggested the use of tables of simulated results to help choose the number of resources needed for desired levels of precision and coverage probability.

We have analysed our estimation scheme in the presence of depolarizing noise with magnitude r. We have shown that the iterative algorithm is still successful in this case provided that no more than $l = -\log_2 r$ iterative stages are performed.

Appendix

Put $x = \cos(2\pi\theta)$, $y = \sin(2\pi\theta)$, $x_0 = 2N_{x=1}/N - 1$, $y_0 = 2N_{y=1}/N - 1$, $\phi(x, y) = \text{atan2}(y, x)$ and $\hat{\phi}(x_0, y_0) = \text{atan2}(y_0, x_0)$. Define

$$\Delta\hat{\phi} = \min((\hat{\phi} - \phi)_{\text{mod}2\pi}, (\phi - \hat{\phi})_{\text{mod}2\pi}).$$

Given $\alpha \in [0, 1/\sqrt{2}]$ and

$$|x - x_0| \leqslant \alpha, \tag{A.1}$$

$$|y - y_0| \leqslant \alpha, \tag{A.2}$$

then (x_0, y_0) lies in a square with sides of length 2α centred around (x, y). From simple geometry, it is obvious that the maximum value of $\Delta\hat{\phi}$ occurs when (x_0, y_0) is one of the four corners of the square. For this case, consider the triangle given by the points $(0, 0)$, (x, y) and (x_0, y_0). The angle at point $(0, 0)$ is $\Delta\hat{\phi}$, and is opposite a side of length $\sqrt{2}\alpha$. The angle at point (x_0, y_0) is opposite to a side of length 1. It follows from the sine rule and monoticity of arcsin on $[0, 1]$ that

$$\Delta\hat{\phi} \leqslant \arcsin(\sqrt{2}\alpha). \tag{A.3}$$

For the iterative algorithm, it is required that $\Delta\hat{\phi} \leqslant \pi/3$, which holds if $\alpha = 0.612$. Then (A.1) and (A.2) are equivalent to (18) and (19).

References

[1] Wei L F and Nori F 2004 Quantum phase estimation algorithms with delays: effects of dynamical phases *J. Phys. A: Math. Gen.* **37** 4607–17
[2] Aspuru-Guzik A, Dutoi A D, Love P J and Head-Gordon M 2005 Simulated quantum computation of molecular energies *Science* **309** 1704–7
[3] Wang X B, You J Q and Nori F 2008 Simulated quantum computation of molecular energies *Phys. Rev.* A **77** 062339
[4] Wang H F, Wu L A, Liu Y X and Nori F 2009 Measurement-based quantum phase estimation algorithm for finding eigenvalues of non-Hermitian matrices *Phys. Rev.* A (at press) (arXiv:0906.2538)
[5] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[6] de Burgh M and Bartlett S D 2005 Quantum methods for clock synchronization: beating the standard quantum limit without entanglement *Phys. Rev.* A **72** 042301
[7] Hayashi M 2006 Parallel treatment of estimation of $SU(2)$ and phase estimation *Phys. Lett.* A **354** 183–9
[8] Kahn J 2007 Fast rate estimation of a unitary operation in $SU(d)$ *Phys. Rev.* A **75** 022326
[9] Imai H and Fujiwara A 2007 Geometry of optimal estimation scheme for $SU(d)$ channels *J. Phys. A: Math. Theor.* **40** 4391–400
[10] Giovannetti V, Lloyd S and Maccone L 2004 Quantum-enhanced measurements: beating the standard quantum limit *Science* **306** 1330–6
[11] Kitaev A Y 1996 Quantum measurements and the Abelian stabilizer problem *Electron. Colloquium Comput. Complexity* **3** (3) 1–20
[12] Rudolph T and Grover L 2003 Quantum communication complexity of establishing a shared reference frame *Phys. Rev. Lett.* **91** 217905
[13] Ji Z, Wang G, Duan R, Feng Y and Ying M 2008 Parameter estimation of quantum channels *IEEE Trans. Inf. Theory* **54** 5172–85
[14] Childs A M, Preskill J and Renes J 2000 Quantum information and precision measurement *J. Mod. Opt.* **47** (2) 155–76
[15] Dobšíček M, Johansson G, Shumeiko V and Wendin G 2007 Arbitrary accuracy iterative quantum phase estimation algorithm using a single ancillary qubit: a two-qubit benchmark *Phys. Rev.* A **76** 030306
[16] Knill E, Ortiz G and Somma R D 2007 Optimal quantum measurements of expectation values of observables *Phys. Rev.* A **75** 012328

[17] 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

[18] Liu X, Luo J and Sun X 2007 Experimental realization of arbitrary accuracy iterative phase estimation algorithms on ensemble quantum computers *Chin. Phys. Lett.* **24** 3316–9

[19] García-Mata I and Shepelyansky D L 2008 Quantum phase estimation algorithm in presence of static imperfections *Eur. J. Phys.* D **47** 151–6

[20] Hoeffding W 1963 Probability inequalities for sums of bounded random variables *J. Am. Stat. Assoc.* **58** 13–30

[21] Huelga S F, Macchiavello C, Pellizzari T, Ekert A K, Plenio M B and Cirac J I 1997 Improvement of frequency standards with quantum entanglement *Phys. Rev. Lett.* **79** 3865–8

[22] Shaji A and Caves C M 2007 Qubit metrology and decoherence *Phys. Rev.* A **76** 032111

[23] Braunstein S L and Caves C M 1994 Statistical distance and the geometry of quantum states *Phys. Rev. Lett.* **72** 3439–43

[24] Rubin M A and Kaushik S 2007 Loss-induced limits to phase measurement precision with maximally entangled states *Phys. Rev.* A **75** 053805

[25] Higgins B L, Berry D W, Bartlett S D, Mitchell M W, Wiseman H M and Pryde G J 2009 Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements *New J. Phys.* **11** 073023