2

Quantum parameter estimation

2.1 Quantum limits to parameter estimation

2.1.1 Introduction

Many experiments can be thought of as comprising two steps: (i) a preparation procedure in which the system to be measured is isolated and prepared, and the apparatus is initialized; and (ii) a measurement step in which the system is coupled to an apparatus and the measurement result recorded. The preparation procedure can be specified by a set of classical parameters, or settings of a physical device. The measurement results are random classical variables that will be correlated with the preparation procedure. In this chapter we are concerned with the case in which the classical parameters specifying the preparation of the state are imperfectly known. Then, assuming that the physical system is well understood, these correlations allow the unknown parameters to be estimated from the measurement results.

As we saw in the last chapter, in quantum mechanics the results of measurements are generally statistical, even when one has complete knowledge of the preparation procedure. A single preparation step and measurement step might not be sufficient to estimate a parameter well. Thus it is common to repeat the two steps of preparation and measurement on a large number of systems, either all at one time or sequentially. Whether measuring one quantum system or many, one is faced with a number of questions. How should one prepare the system state? What sort of measurement should one make on the system? What is the optimum way to extract the parameter from the measurement results? These questions generally fall under the heading of quantum parameter estimation, the topic of this chapter.

From the above, it should be clear that the 'quantum' in 'quantum parameter estimation' is there simply because it is a quantum system that mediates the transfer of information from the classical parameters of the preparation procedure to the classical measurement results. If the system were classical then in principle the parameters could be estimated with perfect accuracy from a single measurement. The role of quantum mechanics is therefore to set *limits* on the accuracy of the estimation of these parameters. Since all of Nature is ultimately quantum mechanical, these are of course fundamental limits. In the following section, we show how a simple expression for quantum limits to parameter estimation can be easily derived.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

2.1.2 The Helstrom–Holevo lower bound

The classic studies of Holevo [Hol82] and Helstrom [Hel76] laid the foundations for a study of quantum parameter estimation. They did so in the context of quantum communication theory. This fits within the above paradigm of parameter estimation as follows. The message or signal is the set of parameters used by the sender to prepare the state of a quantum system. This system is the physical medium for the message. The receiver of the system must make a measurement upon it, to try to recover the message. In this context we need to answer questions such as the following. What are the physical limitations on encoding the message (i.e. in preparing a sequence of quantum systems)? Given these, what is the optimal way to encode the information? What are the physical limitations on decoding the message (i.e. in measuring a sequence of quantum systems)? Given these, what is the optimal way to decode the information?

The work of Helstrom and Holevo established an ultimate quantum limit for the error in parameter estimation. We refer to this as the Helstrom–Holevo *lower bound*. A special case can be derived simply as follows. We assume that the sender has access to a quantum system in a *fiducial state* ρ_0. The only further preparation the sender can do is to transform the state:

$$\rho_0 \to \rho_X = \mathrm{e}^{-\mathrm{i}X\hat{G}}\rho\mathrm{e}^{\mathrm{i}X\hat{G}}. \tag{2.1}$$

Here $\hat{G}$ is an Hermitian operator known as the *generator*, and X is a real parameter. The aim of the receiver is to estimate X. The receiver does this by measuring a quantity X_{est}, which is the receiver's best estimate for X. We assume that X_{est} is represented by an Hermitian operator $\hat{X}_{\mathrm{est}}$.

The simplest way to characterize the quality of the estimate is the mean-square deviation

$$\langle (X_{\mathrm{est}} - X)^2 \rangle_X = \langle (\Delta X_{\mathrm{est}})^2 \rangle_X + [b(X)]^2. \tag{2.2}$$

Here this is decomposed into the variance of the estimator (in the transformed state),

$$\langle (\Delta X_{\mathrm{est}})^2 \rangle_X = \mathrm{Tr}\big[(\hat{X}_{\mathrm{est}} - \langle X_{\mathrm{est}} \rangle_X)^2 \rho_X\big], \tag{2.3}$$

plus the square of the *bias* of the estimator X_{est},

$$b(X) = \langle X_{\mathrm{est}} \rangle_X - X, \tag{2.4}$$

that is, how different the mean of the estimator $\langle X_{\mathrm{est}} \rangle_X = \mathrm{Tr}\big[\hat{X}_{\mathrm{est}}\rho_X\big]$ is from the true value of X.

We now derive an inequality for the mean-square deviation of the estimate. First we note that, from Eq. (2.1),

$$\frac{\mathrm{d}\langle X_{\mathrm{est}} \rangle_X}{\mathrm{d}X} = -\mathrm{i}\,\mathrm{Tr}\big[[\hat{X}_{\mathrm{est}}, \hat{G}]\rho_X\big]. \tag{2.5}$$

Using the general Heisenberg uncertainty principle (A.9),

$$\langle (\Delta X_{\mathrm{est}})^2 \rangle_X \langle (\Delta G)^2 \rangle_X \geq \frac{1}{4}\big|\mathrm{Tr}\big[[\hat{X}_{\mathrm{est}}, \hat{G}]\rho_X\big]\big|^2, \tag{2.6}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

we find that

$$\langle(\Delta X_{\rm est})^2\rangle_X \langle(\Delta G)^2\rangle_X \geq \frac{1}{4}\left|\frac{{\rm d}\langle X_{\rm est}\rangle_X}{{\rm d}X}\right|^2. \tag{2.7}$$

This inequality then sets a lower bound for the mean-square deviation,

$$\langle(X_{\rm est}-X)^2\rangle_X \geq \frac{[1+b'(X)]^2}{4\langle(\Delta G)^2\rangle_X} + b^2(X). \tag{2.8}$$

If there is no systematic error in the estimator, then $\langle X_{\rm est}\rangle_X = X$ and the bias is zero, $b(X) = 0$. In this case, the lower bound to the mean-square deviation of the estimate is

$$\langle(\Delta X_{\rm est})^2\rangle_X \geq \frac{1}{4\langle(\Delta G)^2\rangle_0}. \tag{2.9}$$

Here we have set $X = 0$ on the right-hand side using the fact that $\hat{G}$ commutes with the unitary parameter transformation, so that moments of the generator do not depend on the shift parameter.

Consider the case in which $\hat{X}_{\rm est}$ and $\hat{G}$ are conjugate operators, by which we mean

$$[\hat{G}, \hat{X}_{\rm est}] = -{\rm i}. \tag{2.10}$$

Assume also that ${\rm Tr}\big[\hat{X}_{\rm est}\rho_0\big] = 0$, so that $\langle X_{\rm est}\rangle_X = X$.

Exercise 2.1 *Show this, by considering the Taylor-series expansion for* $e^{{\rm i}X\hat{G}}\hat{X}_{\rm est}e^{-{\rm i}X\hat{G}}$ *and using Eq. (2.10).*

In this case the parameter-estimation uncertainty principle, given in Eq. (2.9), follows directly from the general Heisenberg uncertainty relation (2.6) on using the commutation relations (2.10).

The obvious example of canonically conjugate operators is position and momentum. Let the unitary parameter transformation be $\exp(-{\rm i}X\hat{P})$, where $\hat{P}$ is the momentum operator. Let $\hat{Q}$ be the canonically conjugate position operator, defined by the inner product $\langle q|p\rangle = \exp({\rm i}pq)/\sqrt{2\pi}$. Then $[\hat{Q}, \hat{P}] = {\rm i}$ (see Appendix A). Choosing $\hat{X}_{\rm est} = \hat{Q} - {\rm Tr}\big[\hat{Q}\rho_0\big]$ gives

$$\langle(X_{\rm est}-X)^2\rangle \geq \frac{1}{4\langle(\Delta P)^2\rangle_0}. \tag{2.11}$$

In general, $\hat{X}_{\rm est}$ need not be canonically conjugate to the generator $\hat{G}$. Indeed, in general $\hat{X}_{\rm est}$ need not be an operator in the Hilbert space of the system at all. This means that the Holevo–Helstrom lower bound on the mean-square deviation in the estimate applies not only to projective measurements on the system. It also applies to generalized measurements described by effects. This is because, as we have seen in Section 1.3.2, a generalized measurement on the system is equivalent to a projective measurement of a joint observable on system and meter, namely the unitarily evolved meter readout observable $\hat{R}_{\rm A}(t+T)$. In fact, it turns out that in many cases the *optimal measurement* is just such a generalized measurement. This is one of the most important results to come out of the work by Helstrom and Holevo.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

In the above we have talked of optimal measurement without defining what we mean by optimal. Typically we assume that the receiver knows the fiducial state ρ_0 and the generator $\hat{G}$, but has no information about the parameter X. The aim of the receiver is then to minimize some *cost function* associated with the error in X_{est}. The Helstrom–Holevo lower bound relates to a particular cost function, the mean-square error $\langle (X_{\text{est}} - X)^2 \rangle$. An alternative cost function is $-\delta(X_{\text{est}} - X)$, which when minimized yields the *maximum-likelihood estimate*. However, the singular nature of this function makes working with it difficult. We will consider other alternatives later in this chapter, and in the next section we treat in detail optimality defined in terms of Fisher information.

2.2 Optimality using Fisher information

In some contexts, it makes sense to define optimality in terms other than minimizing the mean-square error $\langle (X_{\text{est}} - X)^2 \rangle$. In particular, for measurements repeated an asymptotically large number of times, what matters asymptotically is not the mean-square error, but the *distinguishability*. That is, how well two slightly different values of X can be distinguished on the basis of a set of M measurement results derived from an ensemble of M systems, each prepared in state ρ_X. This notion of distinguishability is quantified by the *Fisher information* [KJ93], defined as

$$F(X) = \int \mathrm{d}\xi \, \wp(\xi|X) \left(\frac{\mathrm{d} \ln \wp(\xi|X)}{\mathrm{d}X} \right)^2 . \tag{2.12}$$

Here ξ is the result of a measurement of X_{est} on a single copy of the system.

As we will show, not only can this quantity be used instead of the mean-square error, but also it actually provides a stronger lower bound for it than does the Holevo–Helstrom bound:

$$\langle (\delta X_{\text{est}})^2 \rangle_X \geq \frac{1}{MF(X)} \geq \frac{1}{M 4 \langle (\Delta G)^2 \rangle_0} . \tag{2.13}$$

Here M is the number of copies of the system used to obtain the estimate X_{est}. The deviation δX_{est} is not $X_{\text{est}} - X$, but rather

$$\delta X_{\text{est}} = \frac{X_{\text{est}}}{|\mathrm{d}\langle X_{\text{est}} \rangle_X / \mathrm{d}X|} - X. \tag{2.14}$$

This is necessary to compensate for any bias in the estimate, and, for an unbiased estimate, $\langle (\delta X_{\text{est}})^2 \rangle_X$ reduces to the mean-square error. Equation (2.13) will be derived later in this section, but first we show how the Fisher information arises from consideration of distinguishability.

2.2.1 Distinguishability and Fisher information

As noted in Box 1.4, Ramsey interferometry is a way to measure the passage of time using measurements on a large ensemble of two-level atoms. The probability for an atom to be

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

found in the ground state by the final measurement is $\wp_g = \cos^2\theta$, where $\theta = \delta T/2$. Here δ is an adjustable detuning and T is the time interval to be measured.

Clearly a measurement on a single atom would not tell us very much about the parameter θ. If we could actually measure the probability $\wp_g$ then we could easily determine θ. However, the best we can do is to measure whether the atom is in the ground state on a large number M of atoms prepared in the same way. For a finite sample we will then obtain an estimate f_g of $\wp_g$, equal to the observed frequency of the ground-state outcome. Owing to statistical fluctuations, this estimate will not be exactly the same as the actual probability.

In a sample of size M, the probability of obtaining m_g outcomes in the ground state is given by the binomial distribution

$$\wp^{(M)}(m_g) = \binom{M}{m_g} \wp_g(\theta)^{m_g}(1-\wp_g(\theta))^{M-m_g}. \tag{2.15}$$

The mean and variance for the fraction $f_g = m_g/M$ are

$$\langle f_g \rangle_\theta = \wp_g(\theta), \tag{2.16}$$

$$\langle (\Delta f_g)^2 \rangle_\theta = \wp_g(\theta)(1-\wp_g(\theta))/M. \tag{2.17}$$

It is then easy to see that the error in estimating θ by estimating the probability $\wp_g(\theta)$ in a finite sample is

$$\delta\theta = \left|\frac{\mathrm{d}\wp_g}{\mathrm{d}\theta}\right|^{-1} \delta\wp_g = \left|\frac{\mathrm{d}\wp_g}{\mathrm{d}\theta}\right|^{-1} \left[\frac{\wp_g(1-\wp_g)}{M}\right]^{1/2}. \tag{2.18}$$

In order to be able to measure a small shift, $\Delta\theta = \theta' - \theta$, in the parameter from some fiducial setting, θ, the shift must be larger than this error: $\Delta\theta \geq \delta\theta$.

Since $\delta\theta$ is the minimum distance in parameter space between two distinguishable distributions, we can characterize the *statistical distance* between two distributions as the number of distinguishable distributions that can fit between them, along a line joining them in parameter space. This idea was first applied to quantum measurement by Wootters [Woo81]. Because $\delta\theta$ varies inversely with the square root of the the sample size M, we define the statistical distance between two distributions with close parameters θ and θ' as

$$\Delta s = \lim_{M\to\infty} \frac{1}{\sqrt{M}} \frac{\Delta\theta}{\delta\theta}. \tag{2.19}$$

Strictly, for any finite difference $\Delta\theta$ we should use the integral form

$$\Delta s(\theta,\theta') = \lim_{M\to\infty} \frac{1}{\sqrt{M}} \int_\theta^{\theta'} \frac{\mathrm{d}\theta}{\delta\theta}. \tag{2.20}$$

Exercise 2.2 *Show that, for this case of Ramsey interferometry, $\delta\theta = 1/2\sqrt{M}$, independently of θ, so that $\Delta s(\theta,\theta') = 2|\theta' - \theta|$.*

The result in Eq. (2.20) is a special case of a more general result for a probability distribution for a measurement with K outcomes. Let $\wp_k$ be the probability for the outcome

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

k. It can be shown that the infinitesimal statistical distance ds between two distributions, $\wp_k$ and $\wp_k + \mathrm{d}\wp_k$, is best defined by [CT06]

$$(\mathrm{d}s)^2 = \sum_{k=1}^{K} \frac{(\mathrm{d}\wp_k)^2}{\wp_k} = \sum_{k=1}^{K} \wp_k (\mathrm{d} \ln \wp_k)^2. \tag{2.21}$$

If we assume that the distribution depends on a single parameter X, then we have

$$\left(\frac{\mathrm{d}s}{\mathrm{d}X}\right)^2 = \sum_k \wp_k \left(\frac{\mathrm{d} \ln \wp_k(X)}{\mathrm{d}X}\right)^2 \equiv F(X), \tag{2.22}$$

where this quantity is known as the *Fisher information*. The generalization for continuous readout results was already given in Eq. (2.12).

Clearly the Fisher information has the same dimensions as X^{-2}. From the above arguments we see that the reciprocal square root of $MF(X)$ is a measure of the change ΔX in the parameter X that can be detected reliably by M trials. It can be proven that $(\Delta X)^2$ is a lower bound to the mean of the square of the 'debiased' error (2.14) in the estimate X_{est} of X from the set of M measurement results. That is,

$$\langle (\delta X_{\mathrm{est}})^2 \rangle_X \geq \frac{1}{MF(X)}. \tag{2.23}$$

This, the first half of Eq. (2.13), is known as the *Cramér–Rao lower bound*.

Exercise 2.3 *Show that, for the Ramsey-interferometry example, $F(\theta) = 4$, independently of θ. If, for $M = 1$, one estimates θ as 0 if the atom is found in the ground state and $\pi/2$ if it is found in the excited state, show that*

$$\left\langle (\delta\theta_{\mathrm{est}})^2 \right\rangle_\theta = \theta^2 \cos^2\theta + (\theta - |\csc(2\theta)|)^2 \sin^2\theta. \tag{2.24}$$

Verify numerically that the inequality Eq. (2.23) is always satisifed, and is saturated at discrete points, $\theta \approx 0, 1.1656, 1.8366, \ldots$.

Estimators that saturate the Cramér–Rao lower bound at *all* parameter values are known in the statistical literature as *efficient*. We will not use that term, because we use it with a very different meaning for quantum measurements. Instead we will call such estimators Cramér–Rao optimal (CR optimal).

Exercise 2.4 *Show that, if $\wp(\xi|X)$ is a Gaussian of mean X, then $X_{\mathrm{est}} = \xi$ is a Cramér–Rao-optimal estimate of X.*

2.2.2 Quantum statistical distance

The Cramér–Rao lower bound involves properties of the probability distribution $\wp(\xi|X)$ for a single measurement on a system parameterized by X. Quantum measurement theory so far has only entered in that the system is taken to be a quantum system, so that $\wp(\xi|X)$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

is generated by some POM:

$$\wp(\xi|X) = \mathrm{Tr}\big[\rho_X \hat{E}_\xi\big]. \tag{2.25}$$

By contrast, the third expression in Eq. (2.13) involves only the properties of $\rho(X) = \exp(-\mathrm{i}\hat{G}X)\rho(0)\exp(\mathrm{i}\hat{G}X)$. Thus, to prove the second inequality in Eq. (2.13), we must seek an upper bound on the Fisher information over the set of all possible POMs $\{\hat{E}_\xi\colon \xi\}$. Recall that the Fisher information is related to the squared statistical distance between two distributions $\wp(\xi|X)$ and $\wp(\xi|X + \mathrm{d}X)$:

$$(\mathrm{d}s)^2 = (\mathrm{d}X)^2 F(X). \tag{2.26}$$

What we want is a measure of the squared distance between two states ρ_X and $\rho_{X+\mathrm{d}X}$ that generate the distributions:

$$(\mathrm{d}s_Q)^2 = (\mathrm{d}X)^2 \max_{\{\hat{E}_\xi:\xi\}} F(X). \tag{2.27}$$

Here we use the notation $\mathrm{d}s_Q$ to denote the infinitesimal *quantum statistical distance* between two states. Clearly $(\mathrm{d}s_Q)^2/(\mathrm{d}X)^2$ will be the sought upper bound on $F(X)$.

We now present a heuristic (rather than rigorous) derivation of an explicit expression for $(\mathrm{d}s_Q)^2/(\mathrm{d}X)^2$. The classical statistical distance in Eq. (2.21) can be rewritten appealingly as

$$(\mathrm{d}s)^2 = 4\sum_{k=1}^{K}(\mathrm{d}a_k)^2, \tag{2.28}$$

where $a_k = \sqrt{\wp_k}$.

Exercise 2.5 *Verify this.*

That is, the statistical distance between probability distributions is the Euclidean distance between the *probability amplitude* distributions. Now consider two quantum states, $|\psi\rangle$ and $|\psi\rangle + \mathrm{d}|\psi\rangle$. In terms of distinguishing these states, any change in the global phase is irrelevant; the only relevant change is the part $|\mathrm{d}\psi_\perp\rangle = (1 - |\psi\rangle\langle\psi|)\mathrm{d}|\psi\rangle$ which is orthogonal to the first state. That is, without loss of generality, we can take the second state to be $|\psi\rangle + |\mathrm{d}\psi_\perp\rangle$.

From the above it is apparent that, regardless of the dimensionality of the system, there are only two relevant basis states for the problem, $|1\rangle \equiv |\psi\rangle$ and $|2\rangle \equiv |\mathrm{d}\psi_\perp\rangle/||\mathrm{d}\psi_\perp\rangle|$. That is, the system effectively reduces to a two-dimensional system. It is intuitively clear that the relevant eigenstates for the optimal measurement observable will be linear combinations of $|1\rangle$ and $|2\rangle$. If we choose to measure in the basis $|1\rangle$ and $|2\rangle$, then the result $k = 2$ will imply that the state is definitely $|\psi\rangle + \mathrm{d}|\psi\rangle$, whereas the result 1 leaves us uncertain as to the state. In this case, we have $(\mathrm{d}a_1)^2 = 0$, and $(\mathrm{d}a_2)^2 = \langle\mathrm{d}\psi_\perp|\mathrm{d}\psi_\perp\rangle$. Thus, Eq. (2.27) evaluates to

$$(\mathrm{d}s)_Q^2 = 4\langle\mathrm{d}\psi_\perp|\mathrm{d}\psi_\perp\rangle. \tag{2.29}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

Exercise 2.6 *Show that this expression holds for* any *projective measurement in the two-dimensional Hilbert space spanned by* $|1\rangle$ *and* $|2\rangle$. *(This follows from the result of Exercise 2.2.)*

For the case of single-parameter estimation with $|\psi_X\rangle = \exp(-\mathrm{i}\hat{G}X)|\psi_0\rangle$, it is clear that

$$|\mathrm{d}\psi_\perp\rangle = -\mathrm{i}\big(\hat{G} - \langle G\rangle_X\big)|\psi_X\rangle \mathrm{d}X. \tag{2.30}$$

Thus we get

$$\left(\frac{\mathrm{d}s_Q}{\mathrm{d}X}\right)^2 = 4\big\langle(\Delta G)^2\big\rangle_0, \tag{2.31}$$

proving the second lower bound in Eq. (2.13) for the pure-state case.

The case of mixed states is considerably more difficult. The explicit form for the quantum statistical distance turns out to be

$$(\mathrm{d}s)_Q^2 = \mathrm{Tr}[\mathrm{d}\rho\, \mathcal{L}[\rho]\mathrm{d}\rho]\,. \tag{2.32}$$

Here $\mathcal{L}[\rho]$ is a superoperator taking ρ as its argument. If ρ has the diagonal representation $\rho = \sum_j p_j|j\rangle\langle j|$, then

$$\mathcal{L}[\rho]\hat{A} = {\sum_{j,k}}' \frac{2}{p_j + p_k} A_{jk}|j\rangle\langle k|, \tag{2.33}$$

where the prime on the sum means that it excludes the terms for which $p_j + p_k = 0$. If ρ has all non-zero eigenvalues, then $\mathcal{L}[\rho]$ can be defined more elegantly as

$$\mathcal{L}[\rho] = \mathcal{R}^{-1}[\rho], \tag{2.34}$$

where the action of the superoperator $\mathcal{R}[\rho]$ on an arbibtrary operator $\hat{A}$ is defined as

$$\mathcal{R}[\rho]\hat{A} = (\rho\hat{A} + \hat{A}\rho)/2. \tag{2.35}$$

It is clear that $\mathcal{L}[\rho]$ is a superoperator version of the reciprocal of ρ. With this understanding, Eq. (2.32) also looks like a quantum version of the classical statistical distance (2.21).

Exercise 2.7 *Show that, for the pure-state case, Eq. (2.32) reduces to Eq. (2.29), by using the basis* $|1\rangle$, $|2\rangle$ *defined above.*

Now consider again the case of a unitary transformation as X varies, so that

$$\mathrm{d}\rho = -\mathrm{i}[\hat{G}, \rho]\mathrm{d}X. \tag{2.36}$$

To find $(\mathrm{d}s)_Q^2$ from Eq. (2.32), we first need to find the operator $\hat{A} = \mathcal{R}^{-1}[\rho]\mathrm{d}\rho$. From Eq. (2.35), this must satisfy

$$(\rho\hat{A} + \hat{A}\rho) = -2\mathrm{i}[\hat{G}, \rho]\mathrm{d}X. \tag{2.37}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

If $\rho = \hat{\pi}$ (a pure state satisfying $\hat{\pi}^2 = \hat{\pi}$), then $\hat{A} = -2\mathrm{i}[\hat{G}, \hat{\pi}]\mathrm{d}X = 2\,\mathrm{d}\hat{\pi}$ is a solution of this equation. That gives

$$\left(\frac{\mathrm{d}s_Q}{\mathrm{d}X}\right)^2 = -2\,\mathrm{Tr}\big[[\hat{G}, \hat{\pi}]^2\big] = 4\big\langle(\Delta G)^2\big\rangle_0\,, \tag{2.38}$$

as found above. If ρ is not pure then it can be shown that

$$\left(\frac{\mathrm{d}s_Q}{\mathrm{d}X}\right)^2 \leq 4\big\langle(\Delta G)^2\big\rangle_0. \tag{2.39}$$

Putting all of the above results together, we have now three inequalities:

$$M\langle(\delta X_{\mathrm{est}})^2\rangle_X \geq \frac{1}{F(X)} \geq \left(\frac{\mathrm{d}X}{\mathrm{d}s_Q}\right)^2 \geq \frac{1}{4\langle(\Delta G)^2\rangle_0}. \tag{2.40}$$

The first of these is the classical Cramér–Rao inequality. The second we will call the Braunstein–Caves inequality. It applies even if the transformation of $\rho(X)$ as X varies is non-unitary. That is, even if there is no $\hat{G}$ that generates the transformation. The final inequality obviously applies only if there is such a generator. In the case of pure states, it can be replaced by an equality.

Omitting the second term (the classical Fisher information) gives what we will call the Helstrom–Holevo inequality. Like the Cramér–Rao inequality, this cannot always be saturated for a given set $\{\rho_X\}_X$. The advantage of the Braunstein–Caves inequality is that it can always be saturated, as is clear from the definition of the quantum statistical distance in Eq. (2.27). If there is a unitary transformation generated by $\hat{G}$, omitting both the second and the third term gives the inequality (2.9) for the special case of unbiased estimates with $M = 1$. As is apparent, the inequality (2.9) was derived much more easily than those in Eq. (2.40), but the advantage of generality and saturability offered by Eq. (2.40) should also now be apparent.

2.2.3 Achieving Braunstein–Caves optimality

For the case of pure states $|\psi_X\rangle = \exp(-\mathrm{i}\hat{G}X)|\psi_0\rangle$, the inequality

$$F(X) = 4\langle(\Delta G)^2\rangle_0 \tag{2.41}$$

can always be saturated, and the POM $\left\{E_\xi\right\}_\xi$ that achieves this for all X we will call Braunstein–Caves optimal (BC optimal). Clearly this optimality is for a given fiducial state $|\psi_0\rangle$. We will follow the reasoning of Ref. [BCM96] in seeking this optimality.

Requiring the bound to be achieved for all X suggests considering distributions $\wp(\xi|X)$ that are functions of $\xi - X$ only. Since

$$\wp(\xi|X) = \langle\psi_0|\mathrm{e}^{\mathrm{i}\hat{G}X}\hat{E}(\xi)\mathrm{e}^{-\mathrm{i}\hat{G}X}|\psi_0\rangle, \tag{2.42}$$

this implies that

$$\mathrm{e}^{\mathrm{i}X\hat{G}}\hat{E}(\xi)\mathrm{e}^{-\mathrm{i}X\hat{G}} = \hat{E}(\xi - X). \tag{2.43}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

Such measurements are called *covariant* by Holevo [Hol82]. In addition we will posit that the optimal POM is a multiple of a projection operator:

$$\hat{E}(\xi)\mathrm{d}\xi = \mu|\xi\rangle\langle\xi|\mathrm{d}\xi \tag{2.44}$$

for μ a real constant. It is important to note that we do *not* require the states $\{|\xi\rangle\}$ to be orthogonal.

Since the POM is independent of a change of phase for the states $|\xi\rangle$, we can choose with no loss of generality

$$\mathrm{e}^{-\mathrm{i}X\hat{G}}|\xi\rangle = |\xi + X\rangle. \tag{2.45}$$

This means that

$$\langle\xi|\mathrm{e}^{-\mathrm{i}X\hat{G}}|\psi\rangle = \langle\xi - X|\psi\rangle = \exp\left(-X\frac{\partial}{\partial\xi}\right)\langle\xi|\psi\rangle. \tag{2.46}$$

In other words the generator $\hat{G}$ is a displacement operator in the $|\xi\rangle$ representation:

$$\hat{G} \equiv -\mathrm{i}\,\frac{\partial}{\partial\xi}. \tag{2.47}$$

Moreover, the probability distribution $\wp(\xi|X)$ is simply expressed in the ξ representation as

$$\wp(\xi|X) = |\psi_X(\xi)|^2 = |\psi_0(\xi - X)|^2 \equiv \wp_0(\xi - X), \tag{2.48}$$

where $\psi(\xi) \equiv \langle\xi|\psi\rangle/\sqrt{\mu}$.

For the POM to be optimal, it must maximize the Fisher information at $F = 4\langle(\Delta G)^2\rangle_0$. For a covariant measurement, the Fisher information takes the form

$$F = \int \mathrm{d}\xi\,\frac{[\wp_0'(\xi)]^2}{\wp_0(\xi)}, \tag{2.49}$$

where the prime here denotes differentiation with respect to the argument. Note that the conditioning on the true value X has been dropped, because for a covariant measurement F is independent of X. Braunstein and Caves have shown [BC94] that F is maximized if and only if the wavefunction of the fiducial state is, up to an overall phase, given by

$$\psi_0(\xi) = \sqrt{\wp_0(\xi)}\mathrm{e}^{\mathrm{i}\langle G\rangle_0\xi}. \tag{2.50}$$

To see this we can calculate the mean and variance of $\hat{G}$ in the $|\xi\rangle$ representation for the state $\sqrt{\wp_0(\xi)}\mathrm{e}^{\mathrm{i}\Theta(\xi)}$,

$$\langle G\rangle_0 = \int \mathrm{d}\xi\,\psi_0^*(\xi)\left(-\mathrm{i}\,\frac{\partial}{\partial\xi}\right)\psi_0(\xi) \tag{2.51}$$

$$= \int \mathrm{d}\xi\,\wp_0(\xi)\Theta'(\xi), \tag{2.52}$$

$$\langle(\Delta G)^2\rangle_0 = \int \mathrm{d}\xi\,\psi_0^*(\xi)\left(-\mathrm{i}\,\frac{\partial}{\partial\xi} - \langle\hat{G}\rangle_0\right)^2\psi_0(\xi) \tag{2.53}$$

$$= \frac{1}{4}\int \mathrm{d}\xi\,\frac{[\wp_0'(\xi)]^2}{\wp_0(\xi)} + \int \mathrm{d}\xi\,\wp_0(\xi)[\Theta'(\xi) - \langle G\rangle_0]^2. \tag{2.54}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

Exercise 2.8 *Verify these results.*

Clearly, for the Fisher information to attain its maximum value, the phase $\Theta(\xi)$ must be linear in ξ with slope $\langle G\rangle_0$.

We now address the conditions under which a BC-optimal measurement is also a CR-optimal measurement. That is, when it also saturates the first inequality in Eq. (2.40). The mean and variance of the measurement result Ξ are

$$\langle \Xi\rangle_X = \int \mathrm{d}\xi\, \wp_0(\xi - X)\xi = X + \langle \Xi\rangle_0, \tag{2.55}$$

$$\langle (\Delta\Xi)^2\rangle_X = \int \mathrm{d}\xi\, \wp_0(\xi - X)(\xi - \langle \Xi\rangle_X)^2 = \langle (\Delta\Xi)^2\rangle_0. \tag{2.56}$$

Thus there may be a global bias in the mean of Ξ, but the variance of Ξ is independent of X. Suppose now we make M measurements and form the unbiased estimator

$$X_{\mathrm{est}} = \frac{1}{M}\sum_{j=1}^{M}(\Xi_j - \langle \Xi\rangle_0). \tag{2.57}$$

The deviation is $\delta X_{\mathrm{est}} = X_{\mathrm{est}} - X$ and

$$\left\langle (\delta X_{\mathrm{est}})^2\right\rangle = \langle (\Delta\Xi)^2\rangle_0/M. \tag{2.58}$$

Thus, to be a CR-optimal measurement for any M, the POM must saturate the Cramér–Rao bound for $M = 1$. It can be shown that this requires that $\wp_0(\xi)$ be a Gaussian. (Recall Exercise 2.4.) It is very important to remember, however, that, for a given generator $\hat{G}$, physical restrictions on the form of the wavefunctions may make Gaussian states impossible. Thus there may be no states that achieve the Cramér–Rao lower bound for a BC-optimal measurement. Moreover, if we choose estimators other than the sample mean, the fiducial wavefunction that achieves the lower bound need not be Gaussian. In particular, for $M \to \infty$, a maximum-likelihood estimate of X will be CR optimal for any wavefunction of the form (2.50).

The relation

$$\langle (\Delta\Xi)^2\rangle\langle (\Delta G)^2\rangle \geq 1/4 \tag{2.59}$$

looks like the Heisenberg uncertainty relation of the usual form, since $\hat{G} = -\mathrm{i}\,\partial/\partial\xi$ in the ξ representation. However, nothing in our derivation assumed that the states $|\xi\rangle$ were the eigenstates of an Hermitian operator. Indeed, as we shall see, there are many examples for which the BC-optimal measurement is described by a POM with non-orthogonal elements. This is an important reason for introducing generalized measurements, as we did in Chapter 1. One further technical point should be made. In order to find the BC-optimal measurement, we must carefully consider the states for which the generator $\hat{G}$ is a displacement operator. If $\hat{G}$ has a degenerate spectrum then it is not possible to find a BC-optimal measurement in terms of a POM described by a single real number ξ. Further details may be found in [BCM96].

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

2.3 Examples of BC-optimal parameter estimation

2.3.1 Spatial displacement

The generator of spatial displacements is the momentum, $\hat{P}$, so we consider families of states defined by

$$|\psi_X\rangle = \mathrm{e}^{-\mathrm{i}X\hat{P}}|\psi_0\rangle. \tag{2.60}$$

The uncertainty relation Eq. (2.59) then becomes

$$\langle(\delta X_{\mathrm{est}})^2\rangle_X \langle(\Delta P)^2\rangle \geq \frac{1}{4M}. \tag{2.61}$$

The BC-optimal POM $\{\hat{E}(\xi)\mathrm{d}\xi\}_\xi$ is of the form $|\xi\rangle\langle\xi|$, with $\hat{P} = -\mathrm{i}\,\partial/\partial\xi$. This is satisfied for

$$|\xi\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{d}p|p\rangle\mathrm{e}^{-\mathrm{i}\xi p}\mathrm{e}^{\mathrm{i}f(p)}, \tag{2.62}$$

where $|p\rangle$ are the canonical delta-function-normalized eigenstates of $\hat{P}$ (see Appendix A) and $f(p)$ is an arbitrary real function.

Exercise 2.9 *Show that* $\exp(-\mathrm{i}\hat{P}X)|\xi\rangle = |\xi + X\rangle$ *regardless of* $f(p)$.

This illustrates an important point: the conjugate basis to momentum is not unique. The fiducial states $|\xi\rangle$ can be written as

$$|\xi\rangle = \exp[\mathrm{i}f(\hat{P})]|q := \xi\rangle, \tag{2.63}$$

where $|q\rangle$ is a *canonical* position state, defined by Eq. (2.62) with $\xi = q$ and $f(p) \equiv 0$. (See Appendix A.)

In this case the states $|\xi\rangle$ *are* eigenstates of an Hermitian operator, namely

$$\hat{\Xi} = \hat{Q} + f'(\hat{P}), \tag{2.64}$$

where $\hat{Q}$ is the *canonical* position operator with $|q\rangle$ as its eigenstates.

Exercise 2.10 *Show this, using the fact that* $\hat{Q} = \mathrm{i}\,\partial/\partial p$ *in the momentum basis.*

The condition for BC optimality is that the position wavefunction of the fiducial state have the form

$$\langle\xi|\psi_0\rangle = \psi_0(\xi) = r(\xi)\mathrm{e}^{\mathrm{i}\langle P\rangle_0\xi}, \tag{2.65}$$

for $r(\xi)$ real. In the momentum basis, this becomes

$$\langle p|\psi_0\rangle = \mathrm{e}^{\mathrm{i}f(p)}\tilde{r}(p - \langle P\rangle_0), \tag{2.66}$$

where $\tilde{r}$, the Fourier transform of r, is a skew-symmetric function (that is, $\tilde{r}(k) = \tilde{r}^*(-k)$). Thus, if any $f(p)$ is allowed, the condition on $|\psi_0\rangle$ for achieving BC optimality is just that $|\langle p|\psi_0\rangle|^2$ be symmetric in p about $p_0 \equiv \langle P\rangle_0$. If we allow only *canonical* position

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

measurements, with $f(p) = 0$, then Eq. (2.66) implies that

$$\forall k,\ \langle p := p_0 + k|\psi_0\rangle = \langle\psi_0|p := p_0 - k\rangle. \tag{2.67}$$

2.3.2 Spatial displacement of a squeezed state

It is instructive to illustrate these ideas by considering a fiducial state that does not satisfy Eq. (2.67), but that does have a symmetric momentum distribution and does achieve the Cramér–Rao lower bound. We use the special class of fiducial states, the *squeezed vacuum state* [Sch86], defined by

$$|\psi_0\rangle = |r,\phi\rangle = \exp\left[r\left(\mathrm{e}^{-2\mathrm{i}\phi}\hat{a}^2 - \mathrm{e}^{2\mathrm{i}\phi}\hat{a}^{\dagger 2}\right)/2\right]|0\rangle. \tag{2.68}$$

As discussed in Appendix A, this squeezed state is in fact a zero-amplitude coherent state for rotated and rescaled canonical coordinates, $\hat{Q}'$ and $\hat{P}'$, defined by

$$\hat{Q} + \mathrm{i}\hat{P} = (\hat{Q}'\mathrm{e}^{r} + \mathrm{i}\hat{P}'\mathrm{e}^{-r})\mathrm{e}^{\mathrm{i}\phi}. \tag{2.69}$$

If we graphically represent a vacuum state as a circle in phase space with the parametric equation

$$Q^2 + P^2 = \langle 0|(\hat{Q}^2 + \hat{P}^2)|0\rangle = 1, \tag{2.70}$$

then the squeezed vacuum state can be represented by an ellipse in phase space with the parametric equation

$$Q'^2 + P'^2 = \langle\psi_0|(\hat{Q}'^2 + \hat{P}'^2)|\psi_0\rangle = 1. \tag{2.71}$$

This ellipse, oriented at angle ϕ, is shown in Fig. 2.1. These curves can also be thought of as contours for the Wigner or Q function – see Section A.5.

The momentum wavefunction for this fiducial state can be shown to be

$$\langle p|\psi_0\rangle \propto \exp\left(-\frac{p^2}{2\gamma}\right), \tag{2.72}$$

where γ is a complex parameter

$$\gamma = \frac{\cosh r + \mathrm{e}^{2\mathrm{i}\phi}\sinh r}{\cosh r - \mathrm{e}^{2\mathrm{i}\phi}\sinh r}. \tag{2.73}$$

The condition for BC optimality is Eq. (2.66). In this case, since $\langle P\rangle_0 = 0$, it reduces to

$$\langle p|\psi_0\rangle\mathrm{e}^{-\mathrm{i}f(p)} = \langle -p|\psi_0\rangle^*\mathrm{e}^{+\mathrm{i}f(-p)}. \tag{2.74}$$

From Eq. (2.72), this will be the case if

$$f(p) = \frac{p^2\,\mathrm{Im}(\gamma)}{2|\gamma|^2}. \tag{2.75}$$

That is, the CR-optimal measurement is a measurement of

$$\hat{\Xi} = \hat{Q} - \mathrm{Im}(\gamma^{-1})\hat{P}. \tag{2.76}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

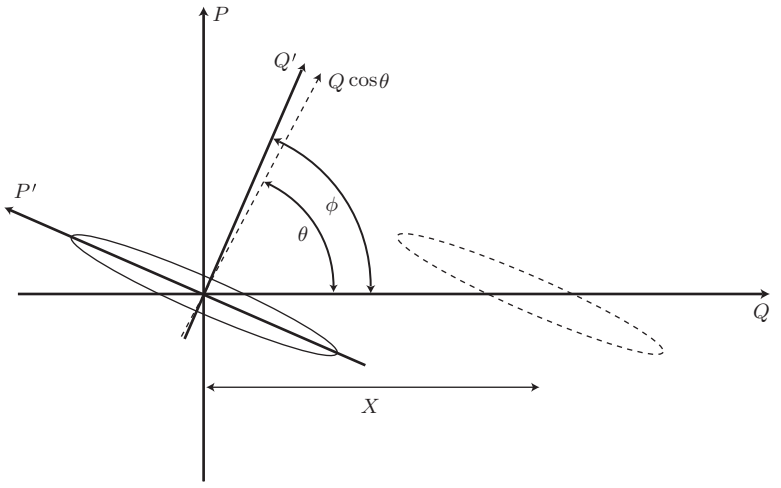

Fig. 2.1 Phase-space representation of optimal measurement of displacement of a squeezed vacuum state. The solid ellipse represents (see the text) the squeezed vacuum state and the dashed-line ellipse that of the state displaced by X in the Q direction. The principal axes of the ellipse are labelled Q' and P', rotated by ϕ relative to Q and P. The BC-optimal measurement for estimating X is to measure the observable $\hat{\Xi} = \left(\hat{Q}\cos\theta + \hat{P}\sin\theta\right)/\cos\theta$. Its eigenstates are displaced along the dashed line labelled $Q\cos\theta$, oriented at an angle θ (determined by ϕ and the degree of squeezing). This measurement can be understood as a compromise between the maximal 'signal', which would be obtained by measuring Q, and the minimal 'noise', which would be obtained by measuring Q'. Adapted from *Annals of Physics* **247**, S. L. Braunstein *et al.*, *Generalized uncertainty relations: Theory, examples, and Lorentz invariance*, 135, Copyright (1996), with permission from Elsevier.

In other words, if we use the squeezed coherent state as a fiducial state, then the optimal measurement is a *linear combination* of $\hat{Q}$ and $\hat{P}$. In this case the $\hat{\Xi}$ representation of the fiducial state is a Gaussian state as expected, with the probability density

$$\wp_0(\xi) = \left(\pi\,\mathrm{Re}[\gamma^{-1}]\right)^{-1/2}\exp\left(-\frac{\xi^2}{\mathrm{Re}[\gamma^{-1}]}\right). \tag{2.77}$$

This has a mean of zero (indicating that Ξ is an unbiased estimator) and a variance of

$$\langle\psi_0|(\Delta\Xi)^2|\psi_0\rangle = \frac{1}{2}\,\mathrm{Re}(\gamma^{-1}) = \frac{1}{4\langle\psi_0|(\Delta P)^2|\psi_0\rangle}. \tag{2.78}$$

Since the probability density is Gaussian, we do not need to appeal to the large-M limit to achieve the Cramér–Rao lower bound. The sample mean of Ξ provides an efficient unbiased estimator of X for all values of M.

The optimal observable $\hat{\Xi}$ can be written as

$$\hat{\Xi} = \frac{\hat{Q}\cos\theta + \hat{P}\sin\theta}{\cos\theta}, \tag{2.79}$$

where $\theta = \mathrm{artan}[-\mathrm{Im}(\gamma^{-1})]$. This operator can be thought of as a modified position operator arising from the rotation in the phase plane by an angle θ followed by a rescaling by $1/\cos\theta$.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

The rescaling means that displacement by X produces the same change in $\hat{\Xi}$ as it does in the canonical position operator $\hat{Q}$. Note that the optimal rotation angle θ is not the same as ϕ, the rotation angle that defines the major and minor axes of the squeezed state. Figure 2.1 and its caption give an intuitive explanation for this optimal measurement.

2.3.3 Harmonic oscillator phase

The question of how to give a quantum description for the phase coordinate of a simple harmonic oscillator goes back to the beginning of quantum mechanics [Dir27] and has given rise to a very large number of attempted answers. Many of these answers involve using an appropriate operator to represent the phase (see Ref. [BP86] for what is probably the most successful approach). Quantum parameter estimation avoids the need to find an operator to represent the parameter.

In appropriately scaled units, the energy of a simple harmonic oscillator of angular frequency ω is given by

$$\hat{H} = \frac{\omega}{2}(\hat{Q}^2 + \hat{P}^2) = \omega(\hat{N} + 1/2). \tag{2.80}$$

Here $\hat{N}$ is the number operator (see Section A.4). In a time τ the phase of a local oscillator changes by $\Theta = \omega\tau$. Thus the unitary operator for a phase shift Θ is

$$\exp(-\mathrm{i}\hat{H}\tau) = \exp(-\mathrm{i}\hat{N}\Theta). \tag{2.81}$$

Here we have removed the constant vacuum energy by redefining $\hat{H}$ as $\omega\hat{N}$.

To find a BC-optimal measurement we seek a POM of the form

$$\hat{E}(\phi)\mathrm{d}\phi = \mu|\phi\rangle\langle\phi|\mathrm{d}\phi, \tag{2.82}$$

such that, following Eq. (2.45), $|\phi\rangle$ is a state for which $\hat{N}$ generates displacement:

$$\exp(-\mathrm{i}\hat{N}\Theta)|\phi\rangle = |\phi + \Theta\rangle. \tag{2.83}$$

This implies that $|\phi\rangle$ is of the form

$$|\phi\rangle = \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}\phi n + \mathrm{i}f(n)}|n\rangle. \tag{2.84}$$

The canonical choice is $f(n) \equiv 0$. This will be appropriate if the fiducial state is of the form

$$|\psi_0\rangle = \sum \sqrt{\wp_n}\mathrm{e}^{\mathrm{i}\theta_0 n}|n\rangle. \tag{2.85}$$

This is the case for many commonly produced states, such as the coherent states (see Section A.4).

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

Since $\hat{E}(\phi)$ is periodic with period 2π, we have to restrict the range of results, for example to the interval $-\pi \leq \phi < \pi$. Normalizing the canonical phase POM then gives

$$\hat{E}(\phi)\mathrm{d}\phi = \frac{1}{2\pi}|\phi\rangle\langle\phi|\mathrm{d}\phi \tag{2.86}$$

with

$$|\phi\rangle = \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}\phi n}|n\rangle. \tag{2.87}$$

These are the Susskind–Glogower phase states [SG64], which are not orthogonal:

$$\langle\phi|\phi'\rangle = \pi\delta(\phi-\phi') - \frac{\mathrm{i}}{2}\cot\left(\frac{\phi-\phi'}{2}\right) + \frac{1}{2}. \tag{2.88}$$

They are overcomplete and are not the eigenstates of any Hermitian operator.

As mooted earlier, this example illustrates the important feature of quantum parameter estimation, namely that it does not restrict us to measuring system observables, but rather allows general POMs. The phase states are in fact eigenstates of a *non-unitary* operator

$$\widehat{\mathrm{e}^{\mathrm{i}\Phi}} = (\hat{N}+1)^{-1/2}\hat{a} = \hat{a}\hat{N}^{-1/2} = \sum_{n=1}^{\infty}|n-1\rangle\langle n|, \tag{2.89}$$

such that

$$\widehat{\mathrm{e}^{\mathrm{i}\Phi}}|\phi\rangle = \mathrm{e}^{\mathrm{i}\phi}|\phi\rangle. \tag{2.90}$$

The ϕ and n representations of the state $|\psi\rangle$ are related by

$$\langle n|\psi\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{d}\phi\,\mathrm{e}^{-\mathrm{i}n\phi}\langle\phi|\psi\rangle. \tag{2.91}$$

The condition on the fiducial state for the measurement to be BC optimal is

$$\langle n := \langle N\rangle + u|\psi_0\rangle = \langle n := \langle N\rangle - u|\psi_0\rangle^*. \tag{2.92}$$

This can be satisfied only for a limited class of states because n is discrete and bounded below by zero. Specifically, choosing $\theta_0 = 0$, it is satisfied by states of the form

$$|\psi_0\rangle = \sum_{n=0}^{2\mu}\sqrt{\wp_n}|n\rangle:\ \wp_n = \wp_{2\mu-n}, \tag{2.93}$$

where μ (integer or half-integer) is the mean photon number. States for which this is satisfied achieve BC optimality:

$$F(\Theta) = 4\big\langle(N-\mu)^2\big\rangle. \tag{2.94}$$

However, because $\wp_n$ has finite support (that is, it is zero outside a finite range of ns), $\wp_0(\phi) = |\langle\phi|\psi_0\rangle|^2$ cannot be a Gaussian. Thus, even assuming that Θ is restricted to

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

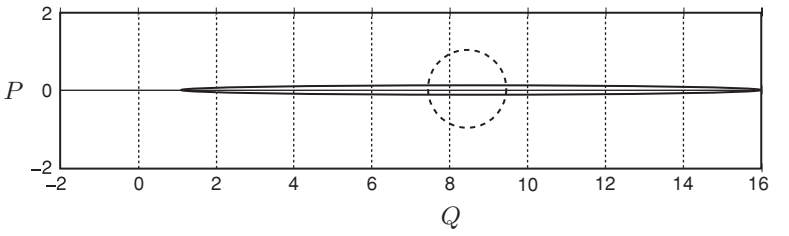

Fig. 2.2 A heuristic phase-space representation of a coherent state (dashed) with amplitude $\alpha = 6$ and a phase-squeezed state (solid) with amplitude $\alpha = 6$ and squeezing parameters $r = 2$ and $\phi = \pi/2$. The contours are defined by parametric equations like Eq. (2.70) and Eq. (2.71).

$[-\pi, \pi)$, the CR lower bound cannot be achieved and we have a strict inequality

$$\left\langle (\delta\Theta_{\rm est})^2 \right\rangle \left\langle (\Delta N)^2 \right\rangle > 1/(4M), \tag{2.95}$$

where M is the number of ϕ measurements contributing to the estimate $\Theta_{\rm est}$, as usual.

It is difficult to produce states with an exact upper bound on their photon number. However, there are many states that can satisfy the BC-optimality condition approximately. For example a state with large mean $\hat{N}$, so that the lower bound at $n = 0$ is not important, and a broad spread in the number distribution that is approximately symmetric about the mean will do. The easiest to produce is a coherent state $|\alpha\rangle$ (see Section A.4). These have a Poissonian number distribution, for which the mean and variance are both equal to $|\alpha|^2$. Thus for coherent states (see Fig. 2.2) we have

$$\left\langle (\delta\Theta_{\rm est})^2 \right\rangle > \frac{1}{MF(\Theta)} > \frac{1}{4M\mu}, \tag{2.96}$$

where the inequalities can be approximately satisfied for large μ. The μ^{-1} scaling is known as the *standard quantum limit*. Here 'standard' arises simply because coherent states are the easiest suitable states to produce.

One could beat the standard quantum limit with another fiducial state, such as a phase-squeezed state of a simple harmonic oscillator. We have already met a class of squeezed state, defined in Eq. (2.68). The phase-squeezed state is this state displaced in phase space in the direction orthogonal to the direction of squeezing (see Fig. 2.2).

The *ultimate quantum limit* arises from choosing a BC-optimal state with the largest number variance for a fixed mean number μ. Clearly this is a state of the form

$$\sqrt{2}|\psi_0\rangle = |0\rangle + |2\mu\rangle, \tag{2.97}$$

which has a variance of μ^2. Thus, for a fixed mean photon number, the ultimate limit is

$$F(\Theta) = 4\mu^2. \tag{2.98}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

The probability distribution for the measurement result ϕ is

$$\wp(\phi|\Theta) = \frac{1}{\pi}\cos^2[\mu(\phi+\Theta)]. \tag{2.99}$$

Although this satisfies Eq. (2.98), for $\mu \geq 1$ it is clear that ϕ is useless for finding an estimate for $\Theta \in [-\pi, \pi)$, because Eq. (2.99) has a periodicity of π/μ. The explanation is that the Fisher information quantifies how well small changes in Θ can be detected, not how well an unknown Θ can be estimated.

Exercise 2.11 *Show Eq. (2.99), and verify Eq. (2.98) by calculating the Fisher information directly from Eq. (2.99).*

2.4 Interferometry – other optimality conditions

As the immediately preceding discussion shows, although BC optimality captures the optimal states and measurements for detecting small shifts in parameters from multiple measurements, it does not necessarily guarantee states or measurements that are good for estimating a completely unknown parameter from one measurement. For this we need some other conditions for optimality. As discussed in Section 2.1.2, Helstrom and Holevo consider optimality in terms of minimizing a cost function. It turns out that, for many problems, any reasonable cost function is minimized for a measurement constructed in the manner described above. That is, for a generator $\hat{G}$, the effects are proportional to rank-1 projectors (Eq. (2.44)), such that $\hat{G}$ generates displacements in the effect basis (Eq. (2.45)). However, the states that minimize the cost may be very different from the states that maximize the Fisher information. (To obtain a finite optimal state in both cases it may be necessary to apply some constraint, such as a fixed mean energy, as we considered above.) In this section we investigate this difference in the context of phase-difference estimation.

2.4.1 The standard quantum limit for interferometry

We have already met interferometry via the Ramsey technique in Section 2.2.1. Interferometry is the basis of many high-precision measurements of time, distance and other physical quantities. The ultimate limit to the precision is, of course, set by quantum mechanics. It is easiest to investigate this limit using a device known as a Mach–Zehnder interferometer (MZI) (see Fig. 2.3). This is built from two optical beam-splitters and two mirrors. The input of the device consists of two modes of a bosonic field, such as the electromagnetic field. The beam-splitter mixes them coherently and turns them into a new pair of modes that form the two arms of the interferometer. In one arm there is an element that introduces a *phase difference* Θ, which is to be estimated. The second beam-splitter coherently mixes the modes of the two arms and gives the two output modes. The outputs of this device are then measured to yield an estimate $\check{\Theta}$ of the phase difference Θ between the two arms of

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

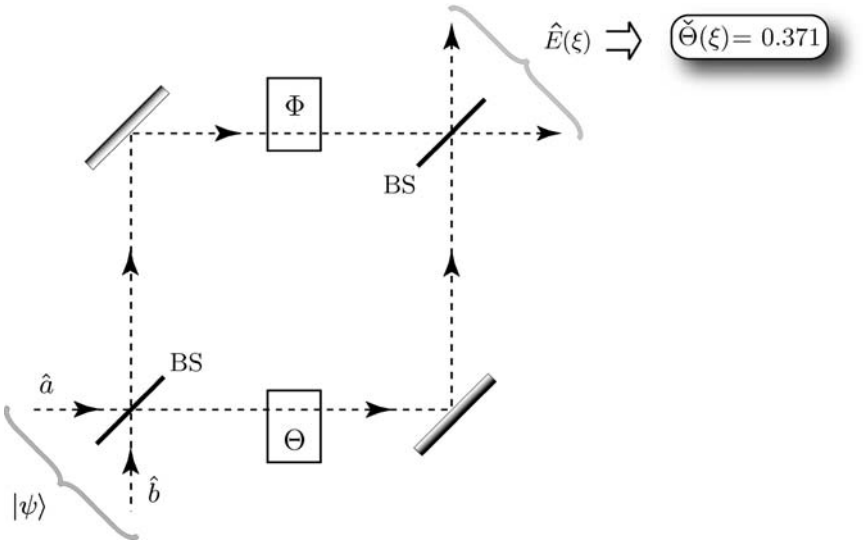

Fig. 2.3 The Mach–Zehnder interferometer. The unknown phase to be estimated is Θ. Both beam-splitters (BS) are 50 : 50. The final measurement is described by a POM $\hat{E}(\xi)$ whose outcome ξ is used to obtain a phase estimate $\check{\Theta}$. (The value shown for this was chosen arbitrarily.) Figure 1 adapted with permission from D. W. Berry *et al.*, *Phys. Rev.* A **63**, 053804, (2001). Copyrighted by the American Physical Society.

the interferometer. We use $\check{\Theta}$ rather than Θ_{est} because in this case the estimate $\check{\Theta}$ is made from a *single* measurement.

The quantum description of the MZI requires a two-mode Hilbert space, with annihilation operators $\hat{a}$ and $\hat{b}$ obeying $[\hat{a}, \hat{a}^\dagger] = [\hat{b}, \hat{b}^\dagger] = 1$, with all other commutators in $\hat{a}$, $\hat{a}^\dagger$, $\hat{b}$ and $\hat{b}^\dagger$ being zero. For convenience we will use nomenclature appropriate for the electromagnetic field and call the eigenstates of $\hat{a}^\dagger\hat{a}$ and $\hat{b}^\dagger\hat{b}$ photon number states, since they have integer eigenvalues (see Section A.4). It is useful to define the following operators:

$$\hat{J}_x = (\hat{a}^\dagger\hat{b} + \hat{a}\hat{b}^\dagger)/2, \tag{2.100}$$

$$\hat{J}_y = (\hat{a}^\dagger\hat{b} - \hat{a}\hat{b}^\dagger)/(2\mathrm{i}), \tag{2.101}$$

$$\hat{J}_z = (\hat{a}^\dagger\hat{a} - \hat{b}^\dagger\hat{b})/2, \tag{2.102}$$

$$\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2 = \hat{j}(\hat{j}+1), \tag{2.103}$$

where

$$\hat{j} = (\hat{a}^\dagger\hat{a} + \hat{b}^\dagger\hat{b})/2 \tag{2.104}$$

has integer and half-integer eigenvalues. This is known as the Schwinger representation of angular momentum, because the operators obey the usual angular-momentum *operator algebra*. A set of operators is said to form an operator algebra if all the commutators are

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

members of that set, as in this case:

$$[\hat{J}_x, \hat{J}_y] = \mathrm{i}\hat{J}_z, \tag{2.105}$$

with cyclic permutations of x, y, z.

Exercise 2.12 *Show this from the commutation relations for the mode operators.*

For simplicity we consider states that are eigenstates of $\hat{j}$. That is, states with an exact total number of photons $\hat{a}^\dagger\hat{a} + \hat{b}^\dagger\hat{b} = 2j$. The MZI elements preserve photon number, so we can always work using the angular-momentum algebra appropriate to a particle of spin j. A balanced (50/50) beam-splitter can be described by the unitary operator

$$\hat{B}_\pm = \exp(\pm\mathrm{i}\pi\hat{J}_x/2), \tag{2.106}$$

where the $\pm$ represents two choices for a phase convention. For convenience we will take the first beam-splitter to be described by $\hat{B}_+$ and the second by $\hat{B}_-$. Thus, in the absence of a phase shift in one of the arms, the nett effect of the MZI is nothing: $\hat{B}_-\hat{B}_+ = \hat{1}$ and the beams a and b come out in the same state as that in which they entered. The choice of $\hat{B}_\pm$ is a convention, rather than a physically determinable fact, because in optics the distances in the interferometer are not usually measured to wavelength scale except by using interferometry. Thus an experimenter would set up an interferometer with the unknown phase Θ set to zero, and then adjust the arms until the desired output (no change) is achieved.

The effect of the unknown phase shift in the lower arm of the interferometer is described by the unitary operator $\hat{U}(\Theta) = \exp(\mathrm{i}\Theta\hat{a}^\dagger\hat{a})$. The operator $\hat{a}$ (rather than $\hat{b}$) appears here because the input beam $\hat{a}$ is identified with the *transmitted* (i.e. straight through) beam. Because $\hat{j}$ is a constant with value j, we can add $\exp(-\mathrm{i}\Theta j)$ to this unitary operator with no physical effect, and rewrite it as $\hat{U}(\Theta) = \exp(\mathrm{i}\Theta\hat{J}_z)$. If we also include a *known* phase shift Φ in the other arm of the MZI, as shown in Fig. 2.3 (this will be motivated later), then we have between the beam-splitters

$$\hat{U}(\Theta - \Phi) = \exp[\mathrm{i}(\Theta - \Phi)\hat{J}_z]. \tag{2.107}$$

The total unitary operator for the MZI is thus

$$\hat{I}(\Theta - \Phi) = \hat{B}_-\hat{U}(\Theta - \Phi)\hat{B}_+ = \exp[-\mathrm{i}(\Theta - \Phi)\hat{J}_y]. \tag{2.108}$$

Exercise 2.13 *Show this. First show the following theorem for arbitrary operators $\hat{R}$ and $\hat{S}$:*

$$\mathrm{e}^{\xi\hat{R}}\hat{S}\mathrm{e}^{-\xi\hat{R}} = \hat{S} + \xi[\hat{R}, \hat{S}] + \frac{\xi^2}{2!}[\hat{R}, [\hat{R}, \hat{S}]] + \cdots. \tag{2.109}$$

Then use the commutation relations for the $\hat{J}$s to show that $\hat{B}_-\hat{J}_z\hat{B}_+ = -\hat{J}_y$. Use this to show that $\hat{B}_- f(\hat{J}_z)\hat{B}_+ = f(-\hat{J}_y)$ for an arbitrary function f.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

The MZI unitary operator $\hat{I}(\Theta - \Phi)$ transforms the photon-number difference operator from the input $2\hat{J}_z$ to the output

$$(2\hat{J}_z)_{\text{out}} = \cos(\Theta - \Phi)2\hat{J}_z + \sin(\Theta - \Phi)2\hat{J}_x. \tag{2.110}$$

Here the subscript 'out' refers to an output operator (that is, a Heisenberg-picture operator for a time after the pulse has traversed the MZI). An output operator is related to the corresponding input operator (that is, the Heisenberg-picture operator for a time before the pulse has met the MZI) by

$$\hat{O}_{\text{out}} = \hat{I}(\Theta - \Phi)\hat{O}\hat{I}(\Theta - \Phi)^{\dagger}. \tag{2.111}$$

Exercise 2.14 *Show Eq. (2.110) using similar techniques to those in Exercise 2.13 above.*

We can use this expression to derive the standard quantum limit (SQL) to interferometry. As defined in Section 2.3.3 above, the SQL is smply the limit that can be obtained using an easily prepared state and a simple measurement scheme. The easily prepared state is a state with all photons in one input, say the a mode.[1] That is, the input state is a $\hat{J}_z$ eigenstate with eigenvalue j. If Θ is approximately known already, we can choose $\Phi \approx \Theta + \pi/2$. Then the SQL is achieved simply by measuring the output photon-number difference operator

$$(2\hat{J}_z)_{\text{out}} = \sin(\Theta + \pi/2 - \Phi)2\hat{J}_z - \cos(\Theta + \pi/2 - \Phi)2\hat{J}_x \tag{2.112}$$

$$\simeq (\Theta + \pi/2 - \Phi)2\hat{J}_z - 2\hat{J}_x \tag{2.113}$$

$$= (\Theta + \pi/2 - \Phi)2j - 2\hat{J}_x. \tag{2.114}$$

This operator can be measured simply by counting the numbers of photons in the two output modes and subtracting one number from the other. We can use this to obtain an estimate via

$$\check{\Theta} = (2J_z)_{\text{out}}/(2j) - (\pi/2 - \Phi), \tag{2.115}$$

where $(2J_z)_{\text{out}}$ is the result of the measurement. It is easy to verify that for a $\hat{J}_z$ eigenstate $\langle J_x \rangle = 0$, so that the mean of the estimate is approximately Θ, as desired. From Eq. (2.114) the variance is

$$\left\langle \check{\Theta}^2 - \left\langle \check{\Theta} \right\rangle^2 \right\rangle \approx \left\langle J_x^2 \right\rangle / j^2. \tag{2.116}$$

For the state $J_z = j$, we have $\left\langle J_z^2 \right\rangle = j^2$, while by symmetry $\left\langle J_x^2 \right\rangle = \left\langle J_y^2 \right\rangle$. Since the sum of these three squared operators is $j(j+1)$, it follows that $\left\langle J_x^2 \right\rangle = j/2$. Thus we get

$$\left\langle \check{\Theta}^2 - \left\langle \check{\Theta} \right\rangle^2 \right\rangle \approx 1/(2j). \tag{2.117}$$

[1] Actually it is not easy experimentally to prepare a state with a definite number of photons in one mode. However, it is easy to prepare a state with an indefinite number of photons in one mode, and then to measure the photon number in each output beam (as discussed below). Since the total number of photons is preserved by the MZI, the experimental results are exactly the same as if a photon-number state, containing the measured number of photons, had been prepared.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

That is, provided that the unknown phase is approximately known already, the SQL for the variance in the estimate is equal to the reciprocal of the photon number.

In fact, this SQL can be obtained without the restriction that $\Phi \approx \Theta + \pi/2$, provided that one uses a more sophisticated technique for estimating the phase from the data. This can be understood from the fact that, with all $2j$ photons entering one port, the action of the MZI is equivalent to that of the Ramsey interferometer (Section 2.2.1) repeated $2j$ times.

Exercise 2.15 *Convince yourself of this fact. Note that the parameter θ in the Ramsey-interferometry example is analogous to $\Theta/2$ in the MZI example.*

As shown in Exercise 2.2, the Fisher information implies that the minimum detectable phase shift is independent of the true phase. Indeed, with $M = 2j$ repetitions we get $(2\,\delta\theta)^2 = 1(2j)$, which is exactly the same as the SQL found above for the MZI.

Note, however, that using the Fisher information to define the SQL has problems, as discussed previously. In the current situation, it is apparent from Eq. (2.112) that the same measurement statistics will result if $\Theta + \pi/2 - \Phi$ is replaced by $\Phi + \pi/2 - \Theta$.

Exercise 2.16 *Convince yourself of this. Remember that $\hat{J}_x$ is pure noise.*

That is, the results make it impossible to distinguish Θ from $2\Phi - \Theta$. Thus, it is still necessary to have prior knowledge, restricting Θ to half of its range, say $[0, \pi)$.

More importantly, if one tries to go beyond the SQL by using states entangled across both input ports (as will be considered in Section 2.4.3) then the equivalence between the MZI and Ramsey interferometry breaks down. In such cases, the simple measurement scheme of counting photons in the output ports will *not* enable an estimate of Θ with accuracy independent of Θ. Rather, one finds that one does need to be able to set $\Phi \approx \Theta + \pi/2$ in order to obtain a good estimate of Θ. To get around the restriction (of having to know Θ before one tries to estimate it), it is necessary to consider measurement schemes that go beyond simply counting photons in the output ports. It is to this topic that we now turn.

2.4.2 Canonical phase-difference measurements

The optimal measurement scheme is of course the BC-optimal measurement, as defined in Section 2.2.3. It follows from Eq. (2.108) that we seek a continuum of states for which $\hat{J}_y$ generates displacements. First we introduce the $\hat{J}_y$ eigenstates $|j, \mu\rangle^y$, satisfying $\hat{J}_y|j, \mu\rangle^y = \mu|j, \mu\rangle^y$, with $-j \leq \mu \leq j$. Then we define unnormalized phase states

$$|j\xi\rangle = \sum_{\mu=-j}^{j} \mathrm{e}^{-\mathrm{i}\mu\xi}|j, \mu\rangle^y, \tag{2.118}$$

with ξ an angle variable. We could have included an additional exponential term $\mathrm{e}^{\mathrm{i}f(\mu)}$ for an arbitrary function f, analogously to Eq. (2.62). By choosing $f \equiv 0$ we are defining *canonical* phase states.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

The canonical POM using these phase states is

$$\hat{E}(\xi)\mathrm{d}\xi = |j\xi\rangle\langle j\xi|\mathrm{d}\xi/(2\pi). \tag{2.119}$$

In terms of the $\hat{J}_y$ eigenstates,

$$\hat{E}(\xi)\mathrm{d}\xi = \frac{1}{2\pi}\sum_{\mu,\nu=-j}^{j} \mathrm{e}^{-\mathrm{i}(\mu-\nu)\xi}\,|j,\mu\rangle^{y}\langle j,\nu|\mathrm{d}\xi. \tag{2.120}$$

A canonical phase-difference measurement is appropriate for a fiducial state $|\psi_0\rangle$ that satisfies

$${}^{y}\langle j,\mu|\psi_0\rangle = {}^{y}\langle j, 2j-\mu|\psi_0\rangle^*. \tag{2.121}$$

We also want Ξ to be an unbiased estimate of Θ. For cyclic variables, the appropriate sense of unbiasedness is that

$$\arg\langle \mathrm{e}^{\mathrm{i}\Xi}\rangle = \arg\int \langle\psi_0|\hat{I}(\Theta)^\dagger\hat{E}(\xi)\hat{I}(\Theta)|\psi_0\rangle \mathrm{e}^{\mathrm{i}\xi}\,\mathrm{d}\xi = \Theta. \tag{2.122}$$

Exercise 2.17 *Show that this will be the case, if we make the coefficients* ${}^{y}\langle j,\mu|\psi_0\rangle$ *real and positive.*

Since we are going to optimize over the input states, we can impose these restrictions without loss of generality. Similarly, there is no need to consider the auxiliary phase shift Φ.

The fiducial state in the $|j,\mu\rangle^y$ basis is not easily physically interpretable. We would prefer to have it in the $|j,\mu\rangle^z$ basis, which is equivalent to the photon-number basis for the two input modes:

$$|j,\mu\rangle^z = |n_a := j+\mu\rangle|n_b := j-\mu\rangle. \tag{2.123}$$

It can be shown [SM95] that the two angular-momentum bases are related by

$${}^{y}\langle j\mu|j\nu\rangle^z = \mathrm{e}^{\mathrm{i}(\pi/2)(\nu-\mu)} I^j_{\mu\nu}(\pi/2), \tag{2.124}$$

where $I^j_{\mu\nu}(\pi/2)$ are the interferometer matrix elements in the $|j,\mu\rangle^z$ basis given by

$$I^j_{\mu\nu}(\pi/2) = 2^{-\mu}\left[\frac{(j-\mu)!}{(j-\nu)!}\frac{(j+\mu)!}{(j+\nu)!}\right]^{1/2} P^{(\mu-\nu,\mu+\nu)}_{j-\mu}(0),$$
$$\text{for } \mu-\nu>-1, \quad \mu+\nu>-1, \tag{2.125}$$

where $P^{(\alpha,\beta)}_n(x)$ are the Jacobi polynomials, and the other matrix elements are obtained using the symmetry relations

$$I^j_{\mu\nu}(\Theta) = (-1)^{\mu-\nu} I^j_{\nu\mu}(\Theta) = I^{\,j}_{-\nu,-\mu}(\Theta). \tag{2.126}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

2.4.3 Optimal states for interferometry

We expect that, with a canonical measurement and optimized input states, the interferometer should perform quadratically better than the standard quantum limit of $(\Delta\check{\Theta})^2 \simeq 1/(2j)$. This expectation is based on an analogy with measurement of the phase of a single mode, treated in Section 2.3.3. There the SQL was achieved with a canonical measurement and coherent states, which gave a Fisher information equal to 4μ, where μ was half the maximum photon number. By contrast, the ultimate quantum limit was a Fisher information scaling as $4\mu^2$.

In order to prove rigorously that there is a quadratic improvement, we need to use a better measure for spread than the variance, because this is strictly infinite for cyclic variables, and depends upon θ_0 if the range is restricted to $[\theta_0, \theta_0 + 2\pi)$. We could use the Fisher information, but, as discussed above, this is not necessarily appropriate if Θ is completely unknown in a range of 2π. Instead we choose the natural measure of spread for a cyclic variable [Hol84], which we will call the Holevo variance:

$$\mathrm{HV} \equiv S^{-2} - 1, \tag{2.127}$$

where $S \in [0, 1]$ we call the *sharpness* of the phase distribution, defined as

$$S \equiv \left|\left\langle \mathrm{e}^{\mathrm{i}\Phi}\right\rangle\right| \equiv \int_0^{2\pi} \mathrm{d}\phi\, \wp(\phi)\mathrm{e}^{\mathrm{i}(\phi-\bar{\phi})}, \tag{2.128}$$

where the 'mean phase' $\bar{\phi}$ is here defined by the requirement that S is real and non-negative. If the Holevo variance is small then it can be shown that

$$\mathrm{HV} \simeq \int_{-\pi}^{\pi} 4\sin^2\left(\frac{\phi-\bar{\phi}}{2}\right)\wp(\phi)\mathrm{d}\phi. \tag{2.129}$$

Exercise 2.18 *Verify this.*

From this it is apparent that, provided that there is no significant contribution to the variance from $\wp(\phi)$ far from $\bar{\phi}$, this definition of the variance is equivalent to the usual definition. Note that (unlike the usual variance) the Holevo phase variance approaches infinity in the limit of a phase distribution that is flat on $[\theta_0, \theta_0 + 2\pi)$.

We now assume as above that the $c_\mu = \langle\psi_0|j, \mu\rangle^y$ are positive. Then, from Eq. (2.120), the sharpness of the distribution $\wp(\xi)$ for the fiducial state $|\psi_0\rangle$ is

$$S = \sum_{\mu=-j}^{j-1} c_\mu c_{\mu+1}. \tag{2.130}$$

We wish to maximize this subject to the constraint $\sum_\mu |c_\mu|^2 = 1$. From linear algebra the solution can be shown to be [BWB01]

$$S_{\max} = \cos\left(\frac{\pi}{2j+2}\right) \tag{2.131}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

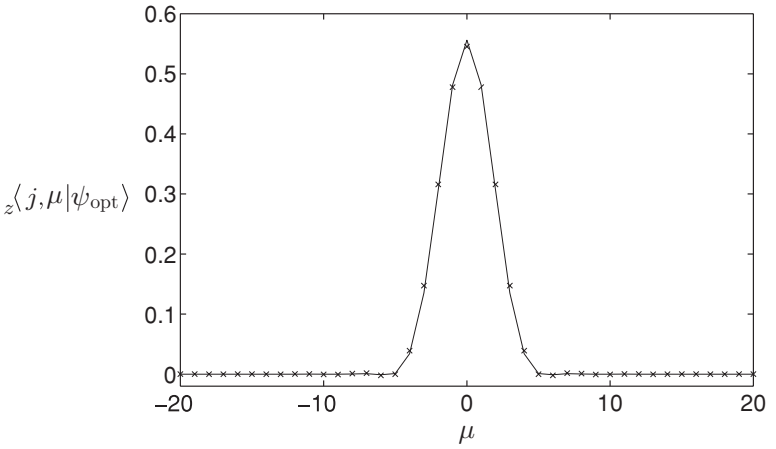

Fig. 2.4 The coefficients $_z\langle j,\mu|\psi_{\rm opt}\rangle$ for the state optimized for minimum phase variance under ideal measurements. All coefficients for a photon number of $2j = 40$ are shown as the continuous line, and those near $\mu = 0$ for a photon number of $2j = 1200$ as crosses. Figure 2 adapted with permission from D. W. Berry *et al.*, *Phys. Rev.* A **63**, 053804, (2001). Copyrighted by the American Physical Society.

for

$$c_\mu = \frac{1}{\sqrt{j+1}} \sin\left[\frac{(\mu+j+1)\pi}{2j+2}\right]. \tag{2.132}$$

The minimum Holevo variance is thus

$$\mathrm{HV} = \tan^2\left(\frac{\pi}{2j+2}\right) = \frac{\pi^2}{(2j)^2} + O(j^{-3}). \tag{2.133}$$

This is known as the Heisenberg limit and is indeed quadratically improved over the SQL. Note that the coefficients (2.132) are symmetric about the mean, so these states are also BC optimal. However, they are very different from the states that maximize the Fisher information, which, following the argument in Section 2.3.3, would have only two non-zero coefficients, $c_{\pm j} = 1/\sqrt{2}$.

Using Eq. (2.124), the state in terms of the eigenstates of $\hat{J}_z$ is

$$|\psi_{\rm opt}\rangle = \frac{1}{\sqrt{j+1}} \sum_{\mu,\nu=-j}^{j} \sin\left[\frac{(\mu+j+1)\,\pi}{2j+2}\right] \mathrm{e}^{\mathrm{i}(\pi/2)(\mu-\nu)} I^j_{\mu\nu}(\pi/2)|j\,\nu\rangle_z. \tag{2.134}$$

An example of this state for 40 photons is plotted in Fig. 2.4. This state contains contributions from all the $\hat{J}_z$ eigenstates, but the only significant contributions are from 9 or 10 states near $\mu = 0$. The distribution near the centre is fairly independent of photon number. To demonstrate this, the distribution near the centre for 1200 photons is also shown in Fig. 2.4. In Ref. [YMK86] a practical scheme for generating a combination of two states near $\mu = 0$ was proposed. Since the optimum states described here have significant contributions

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

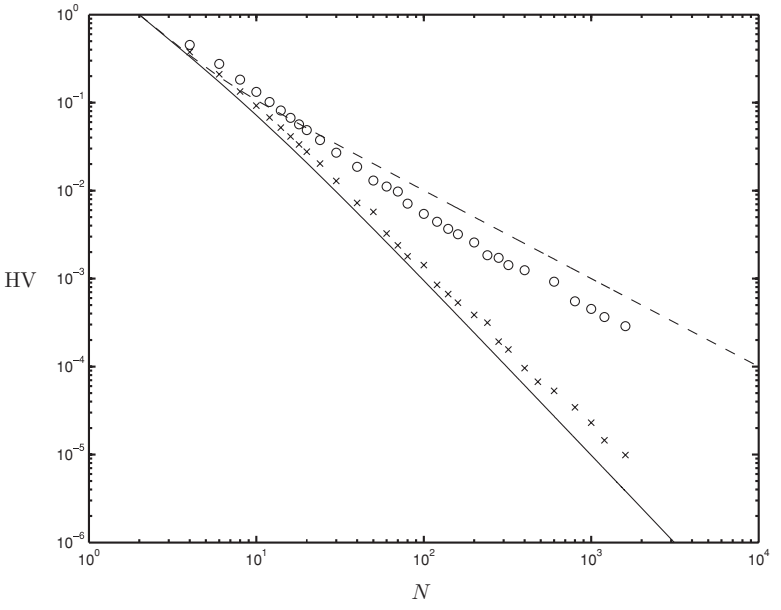

Fig. 2.5 Variances in the phase estimate versus input photon number $N = 2j$. The lines are exact results for canonical measurements on optimized states $|\psi_{\text{opt}}\rangle$ (continuous line) and on states with all photons incident on one input port $|jj\rangle_z$ (dashed line). The crosses are the numerical results for the adaptive phase-measurement scheme on $|\psi_{\text{opt}}\rangle$. The circles are numerical results for a non-adaptive phase-measurement scheme on $|\psi_{\text{opt}}\rangle$. Figure 3 adapted with permission from D. W. Berry *et al.*, *Phys. Rev.* A **63**, 053804, (2001). Copyrighted by the American Physical Society.

from a small number of states near $\mu = 0$, it should be possible to produce a reasonable approximation of these states using a similar method to that in Ref. [YMK86].

The Holevo phase variance (2.133) for the optimal state is plotted in Fig. 2.5. The exact Holevo phase variance of the state for which all the photons are incident on one port, $|jj\rangle_z$, is also shown for comparison. This is the state used in Section 2.4.1 to obtain the SQL $\left\langle \check{\Theta}^2 - \langle \check{\Theta} \rangle^2 \right\rangle \approx 1/(2j)$ under a simple measurement scheme for an approximately known phase shift. As Fig. 2.5 shows, exactly the same result is achieved here asymptotically for the Holevo variance (that is, HV $\sim 1/(2j)$). The difference is that the canonical measurement, as used here, does not require Θ to be approximately known before the measurement begins. Other results are also shown in this figure, which are to be discussed later.

2.5 Interferometry – adaptive parameter estimation

2.5.1 Constrained measurements

The theory of parameter estimation we have presented above is guaranteed to find the BC-optimal measurement scheme (at least for the simple case of a single parameter with generator $\hat{G}$ having a non-degenerate spectrum). In practice, this may be of limited use, because

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

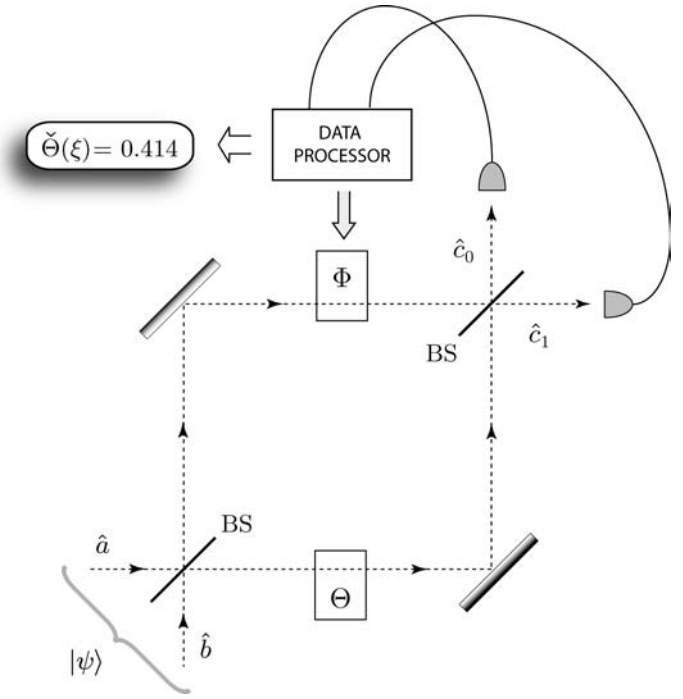

Fig. 2.6 The adaptive Mach–Zehnder interferometer, allowing for feedback to control the phase Φ. Figure 1 adapted with permission from D. W. Berry *et al.*, *Phys. Rev*. A, **63**, 053804, (2001). Copyrighted by the American Physical Society.

it might not be possible to perform the optimal measurement with available experimental techniques. For this reason, it is often necessary to consider measurements constrained by some practical consideration. It is easiest to understand this idea in a specific context, so in this section we again consider the case of interferometric phase measurements.

As we explained in Section 2.4.1, the standard way to do quantum-limited interferometry is simply to count the total number of photons exiting at each output port. For a $2j$-photon input state, the total number of output photons is fixed also, so all of the information from the measurement is contained in the operator

$$(\hat{J}_z)_{\text{out}} = (\hat{a}^\dagger_{\text{out}}\hat{a}_{\text{out}} - \hat{b}^\dagger_{\text{out}}\hat{b}_{\text{out}})/2. \tag{2.135}$$

In Fig. 2.6 we use the new notation

$$\hat{c}_0 = \hat{a}_{\text{out}}, \qquad \hat{c}_1 = \hat{b}_{\text{out}}. \tag{2.136}$$

These output annihilation operators are given by

$$\hat{c}_u(\Theta, \Phi) = \hat{b}\sin\left(\frac{\Theta - \Phi + u\pi}{2}\right) + \hat{a}\cos\left(\frac{\Theta - \Phi + u\pi}{2}\right). \tag{2.137}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

Exercise 2.19 *Show this, again using the technique of Exercise 2.13.*

As noted above, for an arbitrary input state, measuring $(\hat{J}_z)_{\text{out}}$ gives a good estimate of Θ only if $\Phi \simeq \Theta + \pi/2$. This of course requires prior knowledge of Θ, which is not always true. However, it is possible to perform a measurement, still constrained to be realized by photon counting, whose accuracy is close to that of the canonical measurement and independent of Θ. In this respect it is like the canonical measurement, but its accuracy will typically be worse than that of the canonical measurement. In order to realize the measurement we are referring to, it is necessary to make the auxiliary phase Φ *time-varying*. That is, it will be adjusted *over the course of a single measurement*. For example, it could be changed after each detection. This requires breaking down the measurement into individual detections, rather than counting only the total number of detections at one detector minus the total at the other. The measurement operators which describe individual detections are in fact just proportional to the output annihilation operators defined above, as we will now show.

Let us denote the result u from the mth detection as u_m (which is 0 or 1 according to whether the photon is detected in mode c_0 or c_1, respectively) and the measurement record up to and including the mth detection as the binary string $r_m \equiv u_m \dots u_2 u_1$. The state of the two-mode field after m detections will be a function of the measurement record and we denote it as $|\psi(r_m)\rangle$. Denoting the null string by r_0, the state before any detections is $|\psi\rangle = |\psi(r_0)\rangle$. Since we are considering demolition detection, the state after the $(m-1)$th detection will be a two-mode state containing exactly $2j+1-m$ photons.

Define measurement operators corresponding to the two outcomes resulting from the mth photodetection:

$$\hat{M}^{(m)}_{u_m} = \frac{\hat{c}_{u_m}}{\sqrt{2j+1-m}}. \tag{2.138}$$

From Eq. (2.137), the effects

$$\hat{E}^{(m)}_{u_1} = \frac{\hat{c}^\dagger_{u_m}\hat{c}_{u_m}}{2j+1-m} \tag{2.139}$$

satisfy

$$\hat{E}^{(m)}_1 + \hat{E}^{(m)}_0 = \frac{\hat{a}^\dagger\hat{a} + \hat{b}^\dagger\hat{b}}{2j+1-m}. \tag{2.140}$$

On the two-mode subspace of states having exactly $2j+1-m$ photons, this is an identity operator. Thus, these effects do satisfy the completeness condition (1.78) for all states on which they act. Moreover, it is clear that the action of either measurement operator is to reduce the number of photons in the state by one (see Section A.4), as stated above.

The probability for a complete measurement record r_{2j} is

$$\Pr[R_{2j} = u_{2j}u_{2j-1}\dots u_2u_1] = \langle\psi|(\hat{M}^{(1)}_{u_1})^\dagger \dots (\hat{M}^{(2j)}_{u_{2j}})^\dagger \hat{M}^{(2j)}_{u_{2j}} \dots \hat{M}^{(1)}_{u_1}|\psi\rangle. \tag{2.141}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

Now, if Φ is fixed, the $\hat{M}_u^{(m)}$ are independent of m (apart from a constant). Moreover, $\hat{M}_1^m$ and $\hat{M}_0^{m'}$ commute, because $\hat{a}_{\rm out}$ and $\hat{b}_{\rm out}$ commute for Φ fixed. Thus we obtain

$$\Pr[R_{2j} = u_{2j}u_{2j-1}\ldots u_2u_1] = \frac{1}{(2j)!}\langle\psi|(\hat{a}^\dagger)_{\rm out}^{n_a}(\hat{b}^\dagger)_{\rm out}^{n_b}\hat{b}_{\rm out}^{n_b}\hat{a}_{\rm out}^{n_a}|\psi\rangle, \tag{2.142}$$

where

$$n_a = 2j - n_b = \sum_{m=1}^{2j} u_m. \tag{2.143}$$

Exercise 2.20 *Show this.*

That is to say, the probability for the record does not depend at all upon the order of the results, only upon the total number of detections n_a in mode $a_{\rm out}$. Thus, for Φ fixed we recover the result that photon counting measures $(\hat{J}_z)_{\rm out}$, with result $(n_a - n_b)/2$.

If Φ is not fixed, but is made to change during the course of the measurement, then more general measurements can be made. The only relevant values of Φ will be those pertaining to the times at which detections occur, which we will denote Φ_m. Obviously, if Φ_m depends upon m, the results $u_m = 0, 1$ have different significance for different m. Thus the order of the bits in r_{2j} will be important. In general the measurement operators will not commute, and it will not be possible to collapse the probability as in Eq. (2.142).

In the next subsection we will consider the case of adaptive measurements, for which Φ_m depends upon previous results r_{m-1}. However, before considering that, we note that an adjustable second phase Φ is of use even without feedback [HMP+96]. By setting

$$\Phi_m = \Phi_0 + \frac{m\pi}{2j}, \tag{2.144}$$

where Φ_0 is chosen randomly, we vary the total phase shift $\Theta - \Phi$ by a half-cycle over the course of the measurement. (A full cycle is not necessary because an additional phase shift of π merely swaps the operators $\hat{M}_1$ and $\hat{M}_0$.) This means that an estimate $\check{\Theta}$ of Θ can be made with an accuracy independent of Θ. However, as we will show, this phase estimate always has a variance scaling as $O(j^{-1})$, which is much worse than the optimal limit of $O(j^{-2})$ from a canonical measurement.

2.5.2 Adaptive measurements

Before turning to adaptive interferometric measurements, it is worth making a few remarks about what constitutes an adaptive measurement in general.

In Section 1.4.2 we introduced the idea of a *complete* measurement as one for which the conditioned state of the system ρ_r' after the measurement depended only on the result r, not upon the initial state. Clearly no further measurements on this system will yield any more information about its initial state. The complementary class of measurements, *incomplete* measurements, consists of ones for which further measurement of ρ may yield more information about the initial state.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

If an incomplete measurement is followed by another measurement, then the results of the two measurements can be taken together, so as to constitute a greater measurement. Say the set of operations for the first measurement is $\{\mathcal{O}_p\colon\ p\}$ and that for the second is $\{\mathcal{O}'_q\colon\ q\}$. Then the operation for the greater measurement is simply the second operation acting after the first:

$$\mathcal{O}_r\rho = \mathcal{O}'_q(\mathcal{O}_p\rho), \tag{2.145}$$

where $r = (q,\ p)$. Depending upon what sort of information one wishes to obtain, it may be advantageous to choose a different second measurement depending on the result of the first measurement. That is, the measurement $\{\mathcal{O}'_q\colon\ q\}$ will depend upon p. This is the idea of an *adaptive measurement*.

By making a measurement adaptive, the greater measurement may more closely approach the ideal measurement one would like to make. As long as the greater measurement remains incomplete, one may continue to add to it by making more adaptive measurements. Obviously it only makes sense to consider adaptive measurements in the context of measurements that are constrained in some way. For unconstrained measurements, one would simply make the ideal measurement one wishes to make.

It is worth emphasizing again that when we say adaptive measurements we mean measurements on a single system. Another concept of adaptive measurement is to make a (perhaps complete) measurement on the system, and use the result to determine what sort of measurement to make on a second identical copy of the system, and so on. This could be incorporated into our definition of adaptive measurements by considering the system to consist of the original system plus the set of all copies.

The earliest example of using adaptive measurements to make a better constrained measurement is due to Dolinar [Dol73] (see also Ref. [Hel76], p. 163). The *Dolinar receiver* was proposed in the context of trying to discriminate between two non-orthogonal (coherent) states by photodetection, and has recently been realized experimentally [CMG07]. Adaptive measurements have also been found to be useful in estimating the phase (relative to a phase reference called a local oscillator) of a single-mode field, with the measurement again constrained to be realized by photodetection [Wis95]. An experimental demonstration of this will be discussed in the following section. Meanwhile we will illustrate adaptive detection by a similar application: estimating the phase difference in an interferometer as introduced in Ref. [BW00] and studied in more detail in Ref. [BWB01].

Unconstrained interferometric measurements were considered in Section 2.4.2, and constrained interferometric measurements in Section 2.5.1. Here we consider again constrained measurements, where all one can do is detect photons in the output ports, but we allow the measurement to be adaptive, by making the auxiliary phase Φ depend upon the counts so far. Using the notation of Section 2.5.1, the phase Φ_m, before the detection of the mth photon, depends upon the record $r_{m-1} = u_{m-1}\cdots u_1$ of detections (where $u_k = 0$ or 1 denotes a detection in detector 0 or 1, respectively). The question is, how should Φ_m depend upon r_{m-1}?

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

We will assume that the two-mode $2j$-photon input state $|\psi\rangle$ is known by the experimenter – only the phase Θ is unknown. The state after m detections will be a function of the measurement record r_m and Θ, and we denote it as $|\tilde{\psi}(r_m, \Theta)\rangle$. It is determined by the initial condition $|\tilde{\psi}(r_0, \Theta)\rangle = |\psi\rangle$ and the recurrence relation

$$|\tilde{\psi}(u_m r_{m-1}, \Theta)\rangle = \hat{M}^{(m)}_{u_m}(\Theta, \Phi_m)|\tilde{\psi}(r_{m-1}, \Theta)\rangle. \tag{2.146}$$

These states are unnormalized, and the norm of the state matrix represents the probability for the record r_m, given Θ:

$$\wp(r_m|\Theta) = \langle\tilde{\psi}(r_m, \Theta)|\tilde{\psi}(r_m, \Theta)\rangle. \tag{2.147}$$

Thus the probability of obtaining the result u_m at the mth measurement, given the previous results r_{m-1}, is

$$\wp(u_m|\Theta, r_{m-1}) = \frac{\langle\tilde{\psi}(u_m r_{m-1}, \Theta)|\tilde{\psi}(u_m r_{m-1}, \Theta)\rangle}{\langle\tilde{\psi}(r_{m-1}, \Theta)|\tilde{\psi}(r_{m-1}, \Theta)\rangle}. \tag{2.148}$$

Also, the posterior probability distribution for Θ is

$$\wp(\Theta|r_m) = N_m(r_m)\langle\tilde{\psi}(r_m, \Theta)|\tilde{\psi}(r_m, \Theta)\rangle, \tag{2.149}$$

where $N(r_m)$ is a normalization factor. To obtain this we have used Bayes' theorem assuming a flat prior distribution for Θ (that is, an initially unknown phase). A Bayesian approach to interferometry was realized experimentally in Ref. [HMP+96], but only with non-adaptive measurements.

With this background, we can now specify the adaptive algorithm for Φ_m. The sharpness of the distribution after the mth detection is given by

$$S(u_m r_{m-1}) = \left|\int_0^{2\pi} \wp(\theta|u_m r_{m-1})\mathrm{e}^{\mathrm{i}\theta}\,\mathrm{d}\theta\right|. \tag{2.150}$$

A reasonable (not necessarily optimal) choice for the feedback phase before the mth detection, Φ_m, is the one that will maximize the sharpness after the mth detection. Since we do not know u_m beforehand, we weight the sharpnesses for the two alternative results by their probabilities of occurring on the basis of the previous measurement record. Therefore the expression we wish to maximize is

$$M(\Phi_m|r_m) = \sum_{u_m=0,1} \wp(u_m|r_{m-1})S(u_m r_{m-1}). \tag{2.151}$$

Using Eqs. (2.148), (2.149) and (2.150), and ignoring the constant $N_m(r_m)$, the maximand can be rewritten as

$$\sum_{u_m=0,1} \left|\int_0^{2\pi} \langle\tilde{\psi}(u_m r_{m-1}, \theta)|\tilde{\psi}(u_m r_{m-1}, \theta)\rangle\mathrm{e}^{\mathrm{i}\theta}\,\mathrm{d}\theta\right|. \tag{2.152}$$

The controlled phase Φ_m appears implicitly in Eq. (2.152) through the recurrence relation (2.146), since the measurement operator $\hat{M}^{(m)}_{u_m}$ in Eq. (2.138) is defined in terms

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

of $\hat{c}_{u_m}(\Theta, \Phi_m)$ in Eq. (2.137). The maximizing solution Φ_m can be found analytically [BWB01], but we will not exhibit it here.

The final part of the adaptive scheme is choosing the phase estimate $\check{\Theta}$ of Θ from the complete data set r_{2j}. For cyclic variables, the analogue to minimizing the mean-square error is to maximize

$$\left\langle \cos(\check{\Theta} - \Theta) \right\rangle . \tag{2.153}$$

To achieve this, $\check{\Theta}$ is chosen to be the appropriate mean of the posterior distribution $\wp(\theta|r_{2j})$, which from Eq. (2.149) is

$$\check{\Theta} = \arg \int_0^{2\pi} \langle \tilde{\psi}(r_{2j}, \theta) | \tilde{\psi}(r_{2j}, \Theta) \rangle \mathrm{e}^{\mathrm{i}\theta} \, \mathrm{d}\theta . \tag{2.154}$$

This completes the formal description of the algorithm. Its effectiveness can be determined numerically, by generating the measurement results randomly with probabilities determined using $\Theta = 0$, and the final estimate $\check{\Theta}$ determined as above. From Eq. (2.127), an ensemble $\{\check{\Theta}_\mu\}_{\mu=1}^M$ of M final estimates allows the Holevo phase variance to be approximated by

$$\mathrm{HV} \simeq -1 + \left| M^{-1} \sum_{\mu=1}^{M} \mathrm{e}^{\mathrm{i}\check{\Theta}_\mu} \right|^{-2} . \tag{2.155}$$

It is also possible to determine the phase variance exactly by systematically going through all the possible measurement records and averaging over Φ_1 (the auxiliary phase before the first detection). However, this method is feasible only for photon numbers up to about 30.

The results of using this adaptive phase-measurement scheme on the optimal input states determined above are shown in Fig. 2.5. The phase variance is very close to the phase variance for ideal measurements, with scaling very close to j^{-2}. The phase variances do differ relatively more from the ideal values for larger photon numbers, however, indicating a scaling slightly worse than j^{-2}. For comparison, we also show the variance from the non-adaptive phase measurement defined by Eq. (2.144). As is apparent, this has a variance scaling as j^{-1}. Evidently, an adaptive measurement has an enormous advantage over a non-adaptive measurement, at least for the optimal input state.

We can sum up the results of this section as follows. Constrained non-adaptive measurements are often far inferior to constrained adaptive measurements, which are often almost as good as unconstrained measurements. That is, a measurement constrained by some requirement of experimental feasibility typically reaches only the standard quantum limit of parameter estimation. This may be much worse than the Heisenberg limit, which can be achieved by the optimal unconstrained measurement. However, if the experiment is made just a little more complex, by allowing adaptive measurements, then most of the difference can be made up. Note, however, that achieving the Heisenberg limit, whether by adaptive or unconstrained measurements, typically requires preparation of an optimal (i.e. non-standard) input state.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

2.6 Experimental results for adaptive phase estimation

The above adaptive interferometric phase-estimation scheme has recently been achieved experimentally [HBB+07] for the case of $|jj\rangle_z$ input, although there is a twist in the tale (see the final paragraph of Section 7.10). However, it was preceded some years earlier by the closely related single-mode adaptive phase estimation referred to above, in work done by the group of Mabuchi [AAS+02]. In the single-mode case, an unknown phase shift is imprinted upon a single mode, and this mode is made to interfere with an optical local oscillator (that is, an effectively classical mode) before detection. We call this form of measurement *dyne* detection, for reasons that will become obvious. It allows one to estimate the phase of the system relative to that of the local oscillator. The level of mathematics required to analyse this single-mode case is considerably higher than that for the interferometric case (although it yields asymptotic analytical solutions more easily). We will therefore not present the theory, which is contained in Refs. [WK97, WK98]. However, we will present a particularly simple case [Wis95] in Section 7.9.

In a single-shot adaptive phase measurement, the aim is to make a good estimate of the phase of a single pulse of light relative to the optical local oscillator. In the experiment of Armen *et al.* [AAS+02], each pulse was prepared (approximately) in a coherent state of mean photon number $\bar{n}$, with a randomly assigned phase. The best possible phase measurement would be a canonical phase measurement, as described in Section 2.3.3. From Eq. (2.96) with $M = 1$ (for a single-shot measurement), the canonical phase variance is, for $\bar{n}$ large, close to the Helstrom–Holevo lower bound:

$$\left\langle (\delta\check{\Theta}_{\rm can})^2 \right\rangle \simeq 1/(4\bar{n}). \tag{2.156}$$

As in the interferometric case, if the phase to be estimated was known approximately before the measurement, then a simple scheme would allow the phase to be estimated with an uncertainty close to the canonical limit. This is the technique of *homodyne* detection, so called because the local oscillator frequency is the same as that of the signal. But, in a communication context, the phase would be completely unknown. Since canonical measurements are not feasible, the usual alternative is *heterodyne* detection. This involves a local oscillator, which is detuned (i.e. at slightly different frequency from the system) so that it cycles over all possible relative phases with the system. That is, it is analogous to the non-adaptive interferometric phase measurement introduced in Section 2.5.1. Again, as in the interferometric case, this technique introduces noise scaling as $1/\bar{n}$. Specifically, the heterodyne limit to phase measurements on a coherent state is twice the canonical limit [WK97]:

$$\left\langle (\delta\check{\Theta}_{\rm het})^2 \right\rangle \simeq 1/(2\bar{n}). \tag{2.157}$$

The aim of the experiments by Armen *et al.* was to realize an adaptive measurement that can beat the standard limit of heterodyne detection. As in the interferometric case, this involves real-time feedback to control an auxiliary phase Φ, here that of the local oscillator. Since each optical pulse has some temporal extent, the measurement signal generated by the leading edge of a given pulse can be used to form a preliminary estimate of its phase.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

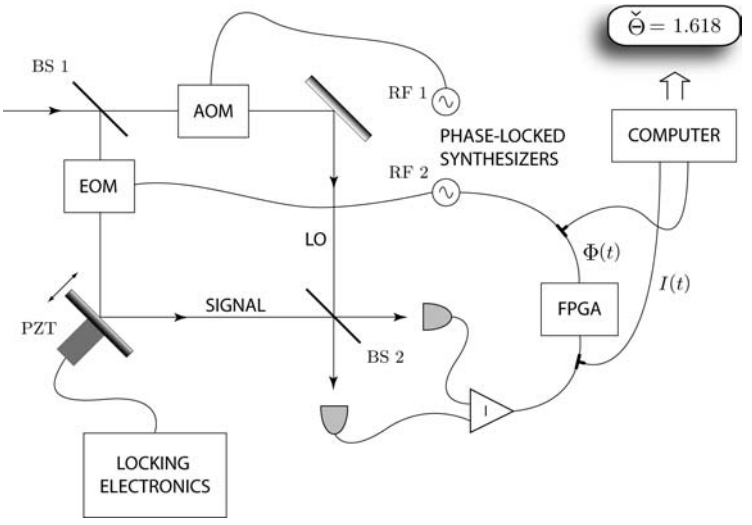

Fig. 2.7 Apparatus used to perform both adaptive homodyne and heterodyne measurements (see the text) in the experiment of Armen *et al.* Solid lines denote optical paths, and dashed lines denote electrical paths. 'PZT' indicates a piezoelectric transducer. Figure 2(a) adapted with permission from M. A. Armen *et al.*, *Phys. Rev. Lett.*, **89**, 133602, (2002). Copyrighted by the American Physical Society.

This can then be used to adjust the local oscillator phase in a sensible way before the next part of the pulse is detected, and so on. Detailed theoretical analyses of such adaptive 'dyne' schemes [WK97, WK98] show that they are very close to canonical measurements. Specifically, for coherent state inputs the difference is negligible even for mean photon numbers of order 10.

Accurately assessing the performance of a single-shot measurement requires many repetitions of the measurement under controlled conditions. Figure 2.7 shows a schematic diagram of the experimental apparatus [AAS+02]. Light from a single-mode (continuous-wave) laser enters the Mach–Zehnder interferometer at beam-splitter 1 (BS 1), thereby creating two beams with well-defined relative phase. The local oscillator (LO) is generated using an acousto-optic modulator (AOM) driven by a radio-frequency (RF) synthesizer (RF 1 in Fig. 2.7). The signal whose phase is to be measured is a weak *sideband* to the carrier (local oscillator). That is, it is created from the local oscillator by an electro-optic modulator (EOM) driven by a RF synthesizer (RF 2) that is phase-locked to RF 1. A pair of photodetectors is used to collect the light emerging from the two output ports of the final 50 : 50 beam-splitter (BS 2). Balanced detection is used: the difference of their photocurrents provides the basic signal used for either heterodyne or adaptive phase estimation. The measurements were performed on optical pulses of duration 50 μs.

In this experimental configuration, the adaptive measurement was performed by feedback control of the phase of RF 2, which sets the relative phase between the signal and the LO. The real-time electronic signal processing required in order to implement the feedback

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

algorithm was performed by a field-programmable gate array (FPGA) that can execute complex computations with high bandwidth and short delays. The feedback and phase-estimation procedure corresponded to the 'Mark II' scheme of Ref. [WK97], in which the photocurrent is integrated with time-dependent gain to determine the instantaneous feedback signal. When performing heterodyne measurements, RF 2 was simply detuned from RF 1 by 1.8 MHz. For both types of measurement, both the photocurrent, $I(t)$, and the feedback signal, $\Phi(t)$, were stored on a computer for post-processing. This is required because the final phase estimate in the 'Mark II' scheme of Ref. [WK97] is not simply the estimate used in the feedback loop, but rather depends upon the full history of the photocurrent and feedback signal. (This estimate is also the optimal one for heterodyne detection.)

The data plotted in Fig. 2.8(a) demonstrate the superiority of an adaptive homodyne measurement procedure over the standard heterodyne measurement procedure. Also plotted is the theoretical prediction for the variance of ideal heterodyne measurement (2.157), both with (thin solid line) and without (dotted line) correction for a small amount of excess electronic noise in the balanced photocurrent. The excellent agreement between the heterodyne data and theory indicates that there is no excess phase noise in the coherent signal states. In the range of 10–300 photons per pulse, most of the adaptive data lie below the absolute theoretical limit for heterodyne measurement (dotted line), and all of them lie below the curve that has been corrected for excess electronic noise (which also has a detrimental effect on the adaptive data).

For signals with large mean photon number, the adaptive estimation scheme used in the experiment was inferior to heterodyne detection, because of technical noise in the feedback loop. At the other end of the scale (very low photon numbers), the intrinsic phase uncertainty of coherent states becomes large and the relative differences among the expected variances for adaptive, heterodyne and ideal estimation become small. Accordingly, Armen *et al.* were unable to beat the heterodyne limit for the mean-square error in the phase estimates for mean photon numbers less than about 8.

However, Armen *et al.* were able to show that the estimator distribution for adaptive homodyne detection remains narrower than that for heterodyne detection even for pulses with mean photon number down to $\bar{n} \approx 0.8$. This is shown in Fig. 2.8(b), which plots the adaptive and heterodyne phase-estimator distributions for $\bar{n} \approx 2.5$. Note that the distributions are plotted on a logarithmic scale. The adaptive phase distribution has a narrower peak than the heterodyne distribution, but exhibits rather high tails. These features agree qualitatively with the numerical and analytical predictions of Ref. [WK98]. It can be partly explained by the fact that the feedback loop occasionally locks on to a phase that is wrong by π.

2.7 Quantum state discrimination

So far in this chapter we have considered the parameter to be estimated as having a continuous spectrum. However, it is quite natural, especially in the context of communication, to

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

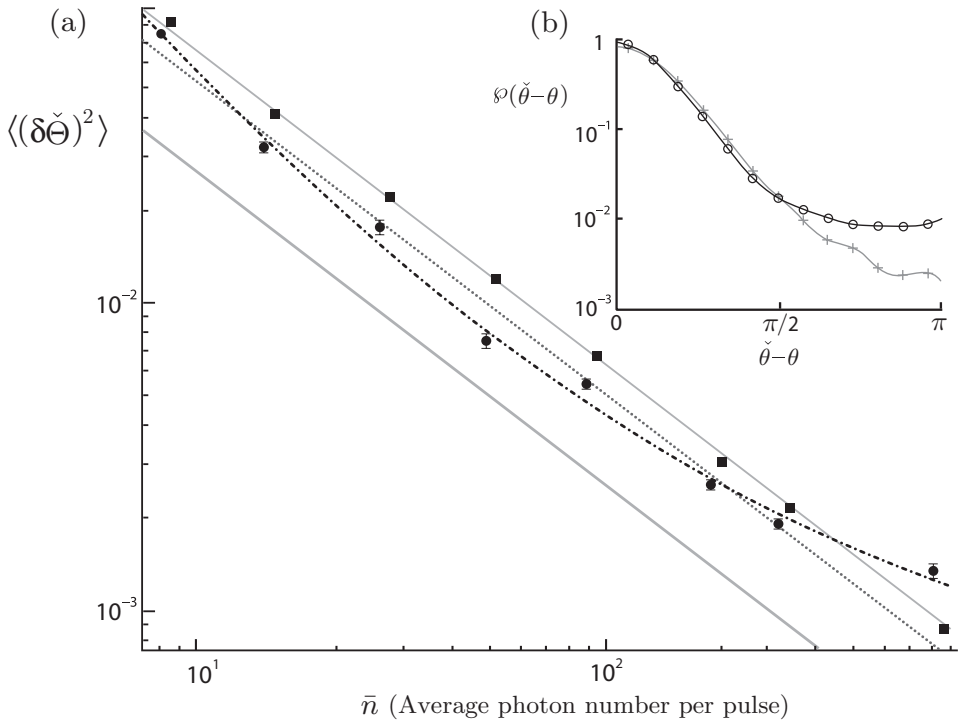

Fig. 2.8 Experimental results from the adaptive and heterodyne measurements. (a) Adaptive (circles) and heterodyne (crosses) phase-estimate variance versus mean photon number per pulse. The dash–dotted line is a second-order curve through the adaptive data, to guide the eye. The thin lines are the theoretical curves for heterodyne detection with (solid) and without (dotted) corrections for detector electronic noise. The thick solid line denotes the fundamental quantum uncertainty limit, given the overall photodetection efficiency. (b) Probability distributions for the error in the phase estimate for adaptive (circles) and heterodyne (crosses) measurements, for pulses with mean photon number of about 2.5. Figure 1 adapted with permission from M. A. Armen *et al.*, *Phys. Rev. Lett.* **89**, 133602, (2002). Copyrighted by the American Physical Society.

consider the case in which the parameter can take values in a finite discrete set. In this case the best estimate should be one of the values in this set, and the problem is really that of deciding which one. This is known as a *quantum decision* or *quantum state-discrimination* problem.

2.7.1 Minimizing the error probability

The simplest example is a communication system that transmits binary information ($s = 0$ or 1) encoded as two states of a physical system ($|\psi_1\rangle$ or $|\psi_0\rangle$). For example, they could be states of light with different intensities. A sequence of systems is directed to a receiver where a measurement is made on each element in the sequence to try to determine which state,

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

and thus which symbol, was transmitted in each case. This is an interesting problem for the case of non-orthogonal states, $\langle\psi_1|\psi_0\rangle = \mu \neq 0$. We will suppose that the prior probability distributions for the source are π_0 and π_1 for the states $|\psi_0\rangle$ and $|\psi_1\rangle$, respectively. The measurements are described by some POM $\{\hat{E}_a\colon a\}$. If the states were orthogonal, they could be discriminated without error, but otherwise there is some finite probability that an error will be made. An error is when the receiver assigns a value 1 when the state transmitted was in fact 0, or conversely. We want to find the optimal POM, which will minimize the effect of errors in a suitable sense to be discussed below. Pioneering work on this question was done by Helstrom [Hel76]. We will follow the presentation of Fuchs [Fuc96].

For a given POM, the above problem also arises in classical decision problems. We need consider only the two probability distributions $\wp(a|s)$ for $s = 0$ and $s = 1$. Classically, such probability distributions arise from noisy transmission and noisy measurement. In order to distinguish $s = 0$ and $s = 1$, the receiver must try to discriminate between the two distributions $\wp(a|s)$ on the basis of a single measurement. If the distributions overlap then this will result in errors, and one should minimize the probability of making an error.[2] Another approach is to relax the requirement that the decision is conclusive; that is, to allow for the possibility of three decisions, yes, no and inconclusive. It may then be possible to make a decision without error, at the expense of a finite probability of not being able to make a decision at all. We will return to this approach in Section 2.7.3.

In each trial of the measurement there are n possible results $\{a\}$, but there are only two possible decision outcomes, 0 and 1. A decision function δ must then take one of the n results of the measurement and give a binary number; $\delta\colon \{1, \ldots, n\} \to \{0, 1\}$. The probability that this decision is wrong is then

$$\wp_e(\delta) = \pi_0\wp(\delta := 1|s := 0) + \pi_1\wp(\delta := 0|s := 1). \tag{2.158}$$

The optimal strategy (for minimizing the error probability) is a *Bayesian decision function*, defined as follows. The posterior conditional probability distributions for the two states, given a particular outcome a, are

$$\wp(s|a) = \frac{\wp(a|s)\pi_s}{\wp(a)}, \tag{2.159}$$

where $\wp(a) = \pi_0\wp(a|s := 0) + \pi_1\wp(a|s := 1)$ is the total probability for outcome a in the measurement. Then the optimal decision function is

$$\delta_B(a) = \begin{cases} 0 & \text{if } \wp(s := 0|a) > \wp(s := 1|a), \\ 1 & \text{if } \wp(s := 1|a) > \wp(s := 0|a), \\ \text{either} & \text{if } \wp(s := 0|a) = \wp(s := 1|a). \end{cases} \tag{2.160}$$

[2] A more sophisticated strategy is to minimize some cost associated with making the wrong decision. The simplest cost function is one that is the same for any wrong decision. That is, we care as much about wrongly guessing $s = 1$ as we do about wrongly guessing $s = 0$. This leads back simply to minimizing the probability of error. However, there are many situations for which other cost functions may be more appropriate, such as in weather prediction, where $s = 1$ indicates a cyclone and $s = 0$ indicates none.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

The (minimal) probability of error under this strategy is

$$\wp_e = \sum_{a=1}^{n} \wp(a)(1 - \max\{\wp(s := 0|a), \wp(s := 1|a)\}) \tag{2.161}$$

$$= \sum_{a=1}^{n} \min\{\pi_0 \wp(a|s := 0), \pi_1 \wp(a|s := 1)\}. \tag{2.162}$$

Exercise 2.21 *Show this.*
Hint: *For a given outcome a, the probability of a correct decision is* $\max\{\wp(s := 0|a), \wp(s := 1|a)\}$.

In the quantum case, the probability of measurement outcome a depends both on the states of the system and on the particular measurement we make on the system: $\wp(a|s) = \mathrm{Tr}[\rho_s \hat{E}_a]$. Thus the probability of error is

$$\wp_e = \sum_a \min\left\{\pi_0 \,\mathrm{Tr}[\rho_0 \hat{E}_a], \pi_1 \,\mathrm{Tr}[\rho_1 \hat{E}_a]\right\}. \tag{2.163}$$

In this case we wish to minimize the error over all POMs. Because we sum over all the results a when making our decision, we really need consider only a binary POM with outcomes $\{0, 1\}$ corresponding to the decision δ introduced above. Then the probability of error becomes

$$\wp_e = \pi_0 \,\mathrm{Tr}[\rho_0 \hat{E}_1] + \pi_1 \,\mathrm{Tr}[\rho_1 \hat{E}_0]. \tag{2.164}$$

Exercise 2.22 *Show this.*

In this case the optimum (i.e. minimum) probability of error is achieved by minimization over all POMs. Because $\hat{E}_0 + \hat{E}_1 = \hat{1}$, we can rewrite Eq. (2.164) as

$$\wp_e = \pi_0 + \mathrm{Tr}\left[(\pi_1\rho_1 - \pi_0\rho_0)\hat{E}_0\right] = \pi_0 + \mathrm{Tr}\left[\hat{\Gamma}\,\hat{E}_0\right], \tag{2.165}$$

where $\hat{\Gamma} = \pi_1\rho_1 - \pi_0\rho_0$. The optimization problem is thus one of finding the minimum of $\mathrm{Tr}[\hat{\Gamma}\,\hat{E}_0]$ over all Hermitian operators $0 \leq \hat{E}_0 \leq \hat{1}$.

Let us write $\hat{\Gamma}$ in terms of its eigenvalues, which may be positive or negative:

$$\hat{\Gamma} = \sum_j \gamma_j |j\rangle\langle j|. \tag{2.166}$$

It is not difficult to see that $\mathrm{Tr}\left[\hat{\Gamma}\,\hat{E}_0\right]$ will be minimized if we choose

$$\hat{E}_0^{\mathrm{opt}} = \sum_{j:\gamma_j<0} |j\rangle\langle j|. \tag{2.167}$$

Exercise 2.23 *Convince yourself of this.*

The minimum error probability is thus

$$\wp_e = \pi_0 + \sum_{j:\gamma_j<0} \gamma_j, \tag{2.168}$$

and this is known as the *Helstrom lower bound*.

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

The optimal measurement $\{\hat{E}_0^{\rm opt}, \hat{E}_1^{\rm opt}\}$ with $\hat{E}_1 = \hat{1} - \hat{E}_0^{\rm opt}$ can be performed by making a measurement of the operator $\hat{\Gamma}$, and sorting the results into the outcomes $s = 0$, $s = 1$ or either, according to whether they correspond to positive, negative or zero eigenvalues, respectively. This is exactly as given in Eq. (2.163), where a plays the role of γ. This shows that the Helstrom lower bound can be achieved using a projective measurement.

We now restrict the discussion to pure states $\rho_s = |\psi_s\rangle\langle\psi_s|$, in which case $\hat{\Gamma} = \pi_1|\psi_1\rangle\langle\psi_1| - \pi_o|\psi_0\rangle\langle\psi_0|$. The eigenvalues are given by

$$\gamma_\pm = -\frac{\pi_0 - \pi_1}{2} \pm \frac{1}{2}\sqrt{1 - 4\pi_0\pi_1|\mu|^2}. \tag{2.169}$$

Exercise 2.24 *Show this.*
Hint: *Only two basis states are needed in order to express* $\hat{\Gamma}$ *as a matrix.*

Thus we find the well-known Helstrom lower bound for the error probability for discriminating two pure states,

$$\wp_e^{\rm min} = \frac{1}{2}\left(1 - \sqrt{1 - 4\pi_0\pi_1|\mu|^2}\right), \tag{2.170}$$

2.7.2 Experimental demonstration of the Helstrom bound

Barnett and Riis [BR97] performed an experiment that realizes the Helstrom lower bound when trying to discriminate between non-orthogonal polarization states of a single photon. The polarization state of a single photon is described in a two-dimensional Hilbert space with basis states corresponding to horizontal (H) and vertical (V) polarization. Barnett and Riis set up the experiment to prepare either of the two states,

$$|\psi_0\rangle = \cos\theta|H\rangle + \sin\theta|V\rangle, \tag{2.171}$$

$$|\psi_1\rangle = \cos\theta|H\rangle - \sin\theta|V\rangle, \tag{2.172}$$

with prior probabilities π_0 and π_1, respectively, and a measurement of the projector

$$\hat{E}_\phi = |\phi\rangle\langle\phi|, \tag{2.173}$$

with $|\phi\rangle = \cos\phi|H\rangle + \sin\phi|V\rangle$, for which an outcome of 1 indicates a polarization in the direction ϕ with respect to the horizontal, while an outcome of 0 indicates a polarization in the direction $\pi/2 + \phi$ with respect to the horizontal. The Helstrom lower bound is attained for

$$\tan(2\phi^{\rm opt}) = \frac{\tan(2\theta)}{\pi_0 - \pi_1}. \tag{2.174}$$

Exercise 2.25 *Show this.*

If the prepared states are equally likely then we have $\phi^{\rm opt} = \pi/4$ and the Helstrom lower bound for the error probability is

$$\wp_e = \frac{1}{2}[1 - \sin(2\theta)]. \tag{2.175}$$

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

In order to perform this experiment as described above we would need a reliable source of single-photon states, suitably polarized. However, deterministic single-photon sources do not yet exist, though they are in active development. Instead Barnett and Riis used an attenuated coherent state from a pulsed source. A coherent-state pulse has a non-determinate photon number, with a Poissonian distribution (see Section A.4). In the experiment light from a mode-locked laser produced a sequence of pulses, which were heavily attenuated (with a neutral-density filter) so that on average each pulse contained about 0.1 photons. For such a low-intensity field, only one in 200 pulses will have more than one photon and most will have none. The laser was operated at a wavelength of 790 nm and had a pulse-repetition rate of 80.3 MHz. The output was linearly polarized in the horizontal plane. Each pulse was then passed through a Glan–Thompson polarizer set at θ or $-\theta$ to produce either of the two prescribed input states.

The polarization measurement was accomplished by passing the pulse through a polarizing beam-splitter, set at an angle $\pi/4$ to the horizontal. This device transmits light polarized in this direction while reflecting light polarized in the orthogonal direction. If the pulse was transmitted, the measurement was said to give an outcome of 1, whereas if it was reflected, the outcome was 0. A 'right' output is when the outcome $a = 0$ or 1 agreed with the prepared state $|\psi_0\rangle$, or $|\psi_1\rangle$, respectively, and a 'wrong' output occurs when it did not. The pulses were directed to photodiodes and the photocurrent integrated, this being simply proportional to the probability of detecting a single photon.

In the experiment the probability of error was determined by repeating the experiment for many photons; that is, simply by running it continually. Call the integrated output from the 'wrong' output I_W, and that from the 'right' output I_R, in arbitrary units. If the Glan–Thompson polarizer is set to θ then the error probability is given by the quantity

$$\wp_e^0 = \frac{I_W}{I_R + I_W} = \left(2 + \frac{I_R - I_W}{I_W}\right)^{-1}. \tag{2.176}$$

If this polarizer is rotated from θ to $-\theta$, the corresponding error probability $\wp_e^1$ is determined similarly. The mean probability of these errors is then taken as an experimental determination of the error probability

$$\wp_e = \frac{1}{2}(\wp_e^0 + \wp_e^1). \tag{2.177}$$

Barnett and Riis determined this quantity as a function of θ over the range 0 to $\pi/4$. The results are shown in Fig. 2.9. Good agreement with the Helstrom lower bound was found.

2.7.3 Inconclusive state discrimination

Thus far, we have discussed an optimal protocol for discriminating two non-orthogonal states that is conclusive, but likely to result in an error. A different protocol, first introduced by Ivanovic [Iva87], requires that the discrimination be *unambiguous* (that is, the probability

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

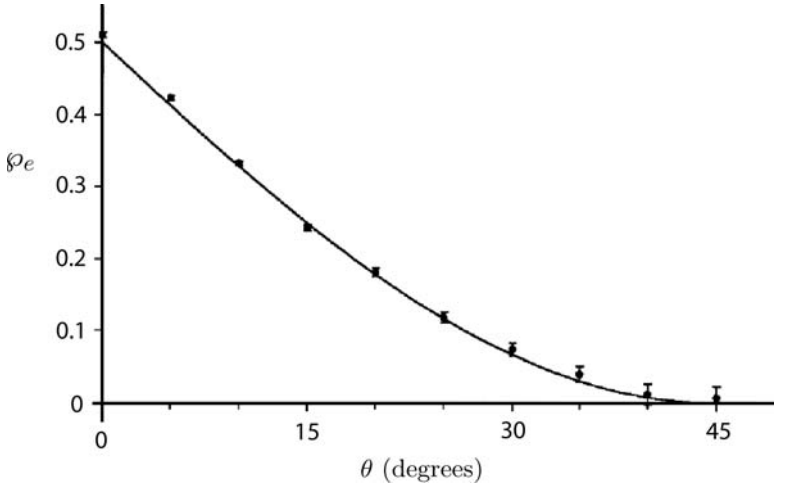

Fig. 2.9 The experimental results of Barnett and Riis. The measured error probability is plotted as a function of the half-angle between the two linear polarization states. The solid curve is the Helstrom bound. Figure 2 adapted with permission from S. M. Barnett and E. Riis, *Experimental demonstration of polarization discrimination at the Helstrom bound*, *Journal of Modern Optics* **44**, 1061, (1997), Taylor & Francis Ltd, http://www.informaworld.com, reprinted by permission of the publisher.

of making an error must be zero), but at the expense of a non-zero probability of an *inconclusive* outcome (that is, an outcome for which no decision can be made). The idea was subsequently elaborated by Dieks [Die88] and Peres [Per88]. We shall refer to the optimal protocol, which minimizes the probability of an inconclusive result, while never making an error, as realizing the IDP lower bound. Unlike the case of the Helstrom lower bound, the IDP bound cannot be achieved using projective measurements on the system. Instead, an ancilla is introduced, so as to make a generalized measurement on the system.

Consider the simple case in which we seek to discriminate between two pure states $|\psi_\pm\rangle$ that have equal prior probability. We can always choose a two-dimensional Hilbert space to represent these states as

$$|\psi_\pm\rangle = \cos\alpha|1\rangle \pm \sin\alpha|0\rangle, \tag{2.178}$$

where $|0\rangle$ and $|1\rangle$ constitute an orthonormal basis, and without loss of generality we can take $0 \leq \alpha \leq \pi/4$. The first step in the protocol is to couple the system to an ancilla two-level system in an appropriate way, so that in the full four-dimensional tensor-product space we can have at least two mutually exclusive outcomes (and hence at most two inconclusive outcomes). Let the initial state of the ancilla be $|0\rangle$. The coupling is the 'exchange coupling' and performs a rotation in the two-dimensional subspace of the tensor-product space spanned by $\{|0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle\}$. The states $|0\rangle \otimes |0\rangle$ and $|1\rangle \otimes |1\rangle$ are invariant. On writing the unitary operator for the exchange as $\hat{U}$, and parameterizing it by

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

θ, the total initial state $|\psi_\pm\rangle \otimes |0\rangle$ transforms to

$$|\Psi_\pm\rangle = \hat{U}(\theta)[(\cos\alpha|1\rangle \pm \sin\alpha|0\rangle)] \otimes |0\rangle \tag{2.179}$$
$$= \cos\alpha[\cos\theta|1\rangle \otimes |0\rangle + \sin\theta|0\rangle \otimes |1\rangle] \pm \sin\alpha|0\rangle \otimes |0\rangle.$$

If we choose $\cos\theta = \tan\alpha$ this state may be written as

$$|\Psi_\pm\rangle = (1-\nu)^{1/2}(|1\rangle \pm |0\rangle) \otimes |0\rangle + (2\nu-1)^{1/2}|0\rangle \otimes |1\rangle, \tag{2.180}$$

where $\nu = \cos^2\alpha$.

Exercise 2.26 *Verify this.*

Since the amplitudes in the first term are orthogonal for the two different input states, they may be discriminated by a projective readout. If we measure the operators $\hat{\sigma}_x \otimes \hat{\sigma}_z$, where $\hat{\sigma}_z = |1\rangle\langle 1| - |0\rangle\langle 0|$ and $\hat{\sigma}_x = |1\rangle\langle 0| + |0\rangle\langle 1|$, the results will be $(1,-1)$ only for the state $|\psi_+\rangle$ and $(-1,-1)$ only for the state $|\psi_-\rangle$. The other two results could arise from either state and indicate an inconclusive result. The probability of an inconclusive result is easily seen to be

$$\wp_i = |\langle\psi_-|\psi_+\rangle| = \cos(2\alpha). \tag{2.181}$$

This is the optimal result, the IDP bound. Of course, the unitary interaction between ancilla and system followed by a projective measurement is equivalent to a generalized measurement on the system alone, as explained in Section 1.2.3.

Exercise 2.27 *Determine the effects $\hat{E}_-$, $\hat{E}_+$ and $\hat{E}_i$ (operators in the two-dimensional system Hilbert space) corresponding to the three measurement outcomes.*

2.7.4 Experimental demonstration of the IDP bound

Experiments achieving the IDP bound have been performed by Huttner *et al.* [HMG+96] and Clarke *et al.* [CCBR01]. Here we present results from the latter. A simplified schematic diagram of the experiment is shown in Fig. 2.10. The two-dimensional state space for the system is the polarization degree of freedom for a single photon in a fixed momentum mode, initially taken as the a direction. Thus

$$|1\rangle = |H\rangle_a, \tag{2.182}$$
$$|0\rangle = |V\rangle_a, \tag{2.183}$$

where $|H\rangle_a$ means a single-photon state in the horizontal polarization of the a momentum mode, and $|V\rangle_a$ similarly for the vertical polarization. The first step is to separate out the horizontal and vertical polarization so that we may conditionally couple to an ancilla mode, labelled c, which is initially in the vacuum state. This is easily achieved by changing the momentum mode of the H-polarized photon so that it is now travelling in a different direction. This may be done using the polarizing beam-splitter, PBS 1 in Fig. 2.10, to

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

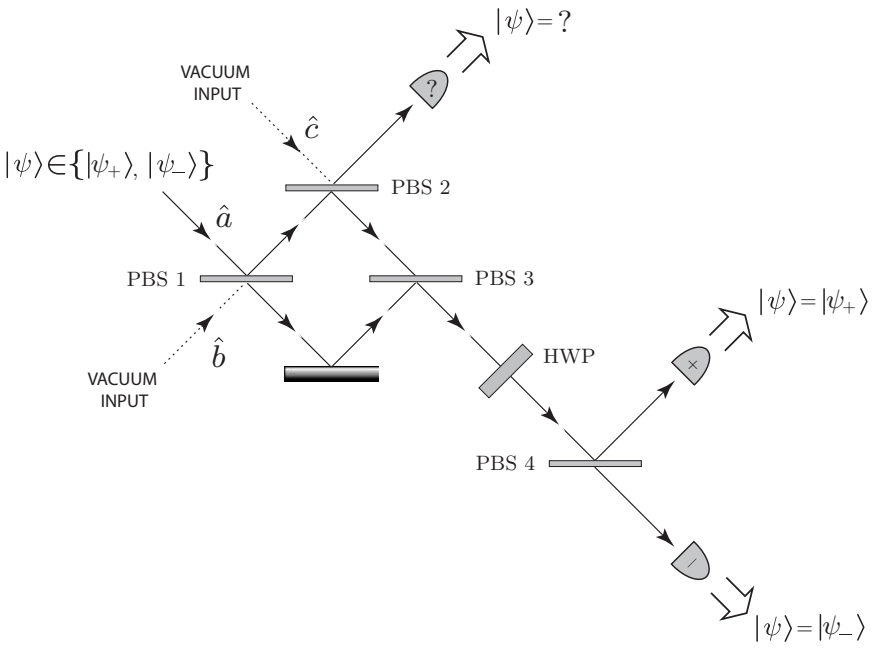

Fig. 2.10 A schematic representation of the optimal state-discrimination experiment of Clarke *et al.* [CCBR01]. PBS indicates a polarizing beam-splitter and PD indicates a photon detector. Note that the device is essentially a polarization Mach–Zehnder interferometer.

transmit vertically polarized light and reflect horizontally polarized light. The total state of the system after PBS 1 is

$$|\psi_{\pm}^{(1)}\rangle = (\cos\alpha|0_a\rangle|H\rangle_b \pm \sin\alpha|V\rangle_a|0_b\rangle)|0_c\rangle, \tag{2.184}$$

where we have dispensed with tensor-product symbols.

The next step is to couple modes b and c using a beam-splitter with a variable transmittivity, PBS 2, with amplitude transmittivity given by $\cos\theta$. For single-photon states, a beam-splitter is equivalent to the exchange interaction. Thus at PBS 2 we have the transformation

$$|H\rangle_b|0\rangle_c \rightarrow \cos\theta|H\rangle_b|0\rangle_c + \sin\theta|0\rangle_b|H\rangle_c, \tag{2.185}$$

where we assume that the transmitted photon does not change its polarization. Thus, just after PBS 2 we can write the total state as

$$|\psi_{\pm}^{(2)}\rangle = (1-\nu)^{1/2}\left[|0_a\rangle|H\rangle_b \pm |V\rangle_a|0_b\rangle\right]|0\rangle_c + (2\nu-1)^{1/2}|0_a\rangle|0\rangle_b|H\rangle_c, \tag{2.186}$$

where $\nu = \cos^2\alpha$ and we have taken $\cos\theta = \tan\alpha$ as in Section 2.7.3.

A photon detector at the output of PBS 2 will now determine whether there is a photon in mode c, which is simply the output for an inconclusive result, labelled '$|\psi\rangle$ =?' in Fig. 2.10. The final step is to make projective measurements to distinguish the two possible

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

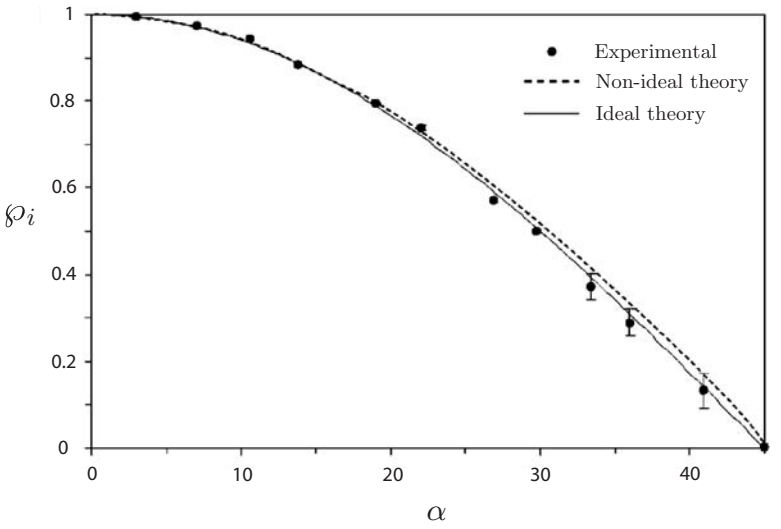

Fig. 2.11 Experiment results for an IDP state-discrimination measurement for the states $\cos\alpha|H\rangle_b \pm \sin\alpha|V\rangle_a$. The probability of an inconclusive result is plotted as a function of α. Figure 3 adapted with permission from R. B. M. Clarke *et al.*, *Phys. Rev.* A **63**, 040305 (R), (2001). Copyrighted by the American Physical Society.

orthogonal states that occur in the first term. This may be done by using a polarizing beam-splitter to put both photons back into the same momentum mode. Thus, after PBS 3 the total state is

$$|\psi_{\pm}^{(3)}\rangle = (1-\nu)^{1/2}\left[|H\rangle_a \pm |V\rangle_a\right]|0\rangle_b|0\rangle_c + (2\nu-1)^{1/2}|0\rangle_a|0\rangle_b|H\rangle_c. \tag{2.187}$$

It now suffices to measure the polarization of mode a in the diagonal basis $|H\rangle_a \pm |V\rangle_a$. This is done by using a sequence of polarizer and polarizing beam-splitter followed by two photon detectors, as indicated in Fig. 2.10.

In Fig. 2.11 we show the results of Clarke *et al.* for the probability of inconclusive results, at the optimal configuration, as the angle between the two states is varied, as a function of α. Within experimental error the agreement between the IDP bound and the experiment is very good.

2.8 Further reading

2.8.1 Quantum tomography

In this chapter we have been concerned mainly with the problem in which the state ρ being measured depends (smoothly) upon a single unknown parameter λ. Of course, it is possible to consider the case in which it depends upon more than one parameter. For a D-dimensional system, if there are D^2-1 parameters upon which ρ depends, then it is possible that the complete space of all possible ρs can be generated by varying the parameter values (an

https://doi.org/10.1017/CBO9780511813948.003 Published online by Cambridge University Press

Hermitian matrix has D^2 independent elements, but the condition $\mathrm{Tr}[\rho] = 1$ removes one of them.) This situation, in which one is effectively trying to identify a completely unknown ρ from measurements on multiple copies, is known as *quantum state estimation* or *quantum tomography*.

A good review of this topic for D-dimensional systems is Ref. [DPS03]. Note that, in order to measure $D^2 - 1$ parameters using *projective* measurements, it is necessary to consider $D + 1$ different projective measurements, that is, measuring $D + 1$ different observables. Such a set of observables is called a *quorum* [Fan57]. That a quorum size of $D + 1$ observables is *necessary* is obvious from the fact that measuring one observable repeatedly can give only $D - 1$ parameters: the D probabilities (one for each outcome) minus one because they must sum to unity. That $D + 1$ is a *sufficient* size for a (suitably chosen) quorum was first proven by Fano [Fan57].

The term *quantum tomography* was coined in quantum optics for estimating the state of an electromagnetic field mode. Here the quorum consists of a collection of quadrature operators $\hat{X}_\theta = \cos\theta\ \hat{Q} + \sin\theta\ \hat{P}$ for different values of θ. This is analogous to the reconstruction of a two-dimensional tissue density distribution from measurements of density profiles along various directions in the plane in medical tomography. Of course, the quantum harmonic oscillator has an infinite Hilbert-space dimension D, so strictly an infinite quorum (infinitely many θs) must be used. In practice, it is possible to reconstruct an arbitrary ρ with a finite quorum, and a finite number of measurements for each observable, if certain assumptions or approximations are made. We refer the reader to the review [PR04] for a discussion of these techniques, including maximum-likelihood estimation.

2.8.2 Other work

A general treatment of the resources (state preparation, and measurement) required to attain the Heisenberg limit in parameter estimation is given by Giovanneti, Lloyd and Maccone [GLM06]. This is done using the language of quantum information processing, which is treated by us in Chapter 7. See the discussion of recent theory and experiment [HBB$^+$07] in optical phase estimation in Section 7.10. The theory in this work is closely related to the adaptive algorithm of Section 2.5.2.

Understanding the quantum limits to parameter estimation has also thrown light on the controversial issue of the time–energy uncertainty relation. It has long been recognized that the time–energy uncertainty relation is of a different character from the position–momentum uncertainty relation, since there is no time operator in quantum mechanics; see for example Ref. [AB61]. However, if $\hat{H}$ is taken as the generator $\hat{G}$ of the unitary, and t as the parameter X to be estimated, then the Holevo upper bound (2.9) gives a precise meaning to the time–energy uncertainty relation.

Developing physical systems to estimate time translations as accurately as possible is, of course, the business of the time-standards laboratories. Modern time standards are based on the oscillations of atomic dipoles, and their quantum description is very similar to that given in Section 2.2.1. However, in practice, clocks are limited in their performance by their

instability, which can be thought of as the relative variation in the time interval between ticks of the clock. In terms of the general theory of parameter estimation, it is as if the generator $\hat{G}$ itself varied randomly by a tiny amount from shot to shot. For a model based on a two-level atomic transition the instability is the ratio of the variation in the frequency of the transition $\delta\omega$ to the mean frequency ω. Currently, the best clocks have an instability $\delta\omega/\omega$ at or below 10^{-17} [HOW+05].

The topic of discriminating non-orthogonal quantum states has been reviewed by Chefles [Che00]; see also Ref. [PR04]. More recently, Jacobs [Jac07] has considered this problem in the context of continuous adaptive measurements similar to that discussed in Section 2.6 above. Jacobs shows that, by using an adaptive technique, one can increase the rate at which the information regarding the initial preparation is obtained. However, in the long-time limit, such an adaptive measurement actually *reduces* the total amount of information obtained, compared with the non-adaptive measurement that reproduces (in the long-time limit) the optimal projective measurement discussed in Section 2.7.1. That is, the adaptive measurement fails to attain the Helstrom lower bound. This is an instructive example of the fact that locally (in time) optimizing the rate of increase in some desired quantity (such as the information about the initial state preparation) does not necessarily lead to the globally optimal scheme. There is thus no reason to expect the adaptive measurement schemes discussed in Sections 2.5 and 2.6 to be optimal.