CLASSICAL AND QUANTUM PARAMETER ESTIMATION THEORY FOR OPTICAL SPECTROSCOPY AND IMAGING

ANG SHAN ZHENG

NATIONAL UNIVERSITY OF SINGAPORE

2016

CLASSICAL AND QUANTUM PARAMETER ESTIMATION THEORY FOR OPTICAL SPECTROSCOPY AND IMAGING

ANG SHAN ZHENG
(B.Eng.(Hons.), NUS)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2016

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Ang Shan Zheng
August 5, 2016

Acknowledgement

Before venturing any further, I want to thank my supervisor, Dr. Mankei Tsang, for his guidance throughout this research and his generosity in funding my study and research.

Special thanks also goes to my colleagues, Mr. Shilin Ng, Dr. Ranjith Nair, Dr. Xiao-Ming Lu, and Dr. Shibdas Roy for countless hours of academic discussions and inspirations. I am grateful for the pleasant research environment in our research group, and I would like to thank our former group members, Dr. Adam Chaudhry, Mr. Soham Saha, Ms. Dan Li, and Dr. Andy Chia. I also thank all my collaborators, Dr. Glen I. Harris, Professor Warwick P. Bowen, Dr. Trevor A. Wheatley, Dr. Hidehiro Yonezawa, Professor Akira Furusawa, and Professor Elanor H. Huntington. Without their efforts, experiments described in this thesis would not have been possible.

Foremost, I would like to thank my lovely wife ChinHui for the patience she has had with me throughout this journey and the love and encouragement she has given me to succeed. I wish to thank my sister Siao Wen and my parents in Malaysia for their support and trust in pursuing my interests.

Contents

Summary

In this thesis, we present a series of results about parameter estimation theory for optical systems. The thesis is subdivided into three parts. The first part concerns the problem of parameter estimation from a noisy optomechanical system. We derive analytic expressions for the Cramér-Rao lower bound on the estimation errors, and apply various estimation techniques to experimental data to estimate the parameters of an optomechanical system, including force-noise power, mechanical resonance frequency, damping rate and measurement noise power. The analytic results should be valuable to optomechanical experiment designs.

In the second part, we propose a theoretical framework for the estimation of spectral parameters with quantum dynamical systems, deriving simple analytic results for quantum limits to the estimation errors in terms of power spectral densities, such that they can be readily applied to optics and optomechanics experiments. To illustrate our theory, we analyze a recent experiment of continuous optical phase estimation and demonstrate that the experimental performance using homodyne detection is close to our quantum limit. We further propose a spectral photon counting method that can beat homodyne detection and attain quantum-optimal performance in low signal-to-noise regime.

The last part is dedicated to the field of optical imaging. Here we study the problem of two-dimensional localization of two incoherent optical point sources from the perspective of quantum estimation theory. We obtain

the fundamental limit to the estimation of the Cartesian components of the centroid and separation of the sources. We propose two measurement schemes that asymptotically attain the quantum Cramér-Rao bound for both components of the separation over many trials.

List of Tables

List of Figures

List of Symbols

$\|\psi\rangle$	Quantum state vector, 'ket'
$\langle\psi\|$	Quantum state vector, 'bra', the Hermitian conjugate of $\|\psi\rangle$
$\hat{O}$	Quantum operator
$\mathrm{tr}(\hat{O})$	Trace of operator $\hat{O}$
$\hat{O}^\dagger$	Hermitian conjugate of operator $\hat{O}$
$\mathbb{E}(X)$	Expectation value of X
$\check{\theta}$	Estimate of parameter θ
$A^\top$	Transpose of matrix A
$A^\dagger$	Conjugate transpose of matrix A
A^{-1}	Matrix inverse of matrix A
$\det(A)$	Determinant of matrix A
$\mathrm{Tr}(A)$	Trace of matrix A
z^*	Complex conjugate of complex number z
$\arg z$	Argument of complex number z
$\mathrm{Re}(z)$	Real part of complex number z
$\mathrm{Im}(z)$	Imaginary part of complex number z
$g * h$	Convolution of functions g and h

List of Abbreviations

CR *lower bound*	Cramér-Rao lower bound for the estimation error covariance matrix. *See* page 21, *and also* QCR
EM *algorithm*	Expectation-Maximization algorithm. *See* page 17
HG *basis*	Hermite-Gaussian basis
ML *estimator*	Maximum-likelihood estimator. *See* page 17
MSE	Mean-square error
POVM	Positive-operator valued measure. *See* page 14
PSD	Power spectral density
PSF	Point-spread function
QCR *bound*	Quantum Cramér-Rao lower bound for the estimation error covariance matrix. *See* page 24
QFI	Quantum Fisher information. *See* page 24 and Eq. (2.86) for definition
RTS *smoother*	Rauch-Tung-Striebel smoother
SPADE	Spatial-mode demultiplexing. *See* section 5.5
SLD	Symmetric logarithmic derivative. *See* page 25
SLIVER *interferometry*	Super localization by image inversion interferometry. *See* section 5.4
SNR	Signal-to-noise ratio
SPLOT	An assumption of stationary processes and a long observation time

Chapter 1

Introduction

Quantum mechanics is a fascinating source of probabilistic and statistical models. Isolated from outside, a quantum system evolves according to the Schrödinger's equation. At the instance the quantum system comes to contact with outside world (for example, when measurement is performed), the system state makes a random jump to a new state and gives an outcome with a certain probability distribution.

For many decades, the quantum randomness had been hard to observe owing to technological limitations. Recent spectacular technological developments for microscopic and quantum systems have, however, made the probabilistic nature of quantum measurement increasingly relevant to modern technology. Quantum metrology, in particular, promises much more sensitive parameter estimation than that offered by conventional techniques. To realize this, it is necessary to study the estimation theory for classical and quantum systems.

Estimation theory studies the inference of parameters from noisy experimental data. The strategy for finding the estimated parameters from the data is called an estimator [1, 2]; while an experimental design is for choosing the measurements used for the estimation. In classical systems, the noise comes from technological imperfections, while in quantum mechan-

ics, the noise is inherent. Regardless of the origin of the noise, the goal here and of estimation theory is to minimize the estimation error, defined as the difference between the estimate and the parameter, given a noise model. The key aim is to find a combination of estimator and experimental design that gives us a more precise estimate, using fewer measurement trials or shorter measurement time. To achieve this, we need a method in evaluating the size of the estimation error. We focus here on the mean-square error, a commonly used criterion in estimation theory.

In this thesis, we derive analytic expressions for fundamental lower bounds on the mean-square estimation error. On the topic of optical spectroscopy, we derive an asymptotic form of the bounds in terms of power spectral densities, such that they can be applied to optomechanics [3] and optics experiments. We analyze real experiments and compare the results with our bounds. On the topic of imaging, we derive the quantum bound for estimating the centroid and separation of two incoherent optical point sources. We also propose two experimental designs that approach the bound.

1.1 Outline of the Thesis

This thesis is organized as follows.

In Chapter 2, we give a brief review on key ideas and theories used in this thesis.

In Chapter 3, we develop a statistical framework used to study the problem of parameter estimation from a noisy optomechanical system. We obtain analytically a lower bound on the estimation errors and discuss the details of various estimators applied to the experimental data.

In Chapter 4, we focus on the parameter estimation of the spectral parameters of a classical stochastic process coupled to a quantum dynamical

system. We derive analytically the fundamental limits in terms of the power spectral densities and investigate measurement and data analysis techniques that approach the limits.

In Chapter 5, we tackle the problem of optical imaging resolution, challenging the long-standing Rayleigh's criterion. We obtain the quantum limit for estimating the Cartesian components of the centroid and separation of two incoherent optical point sources. We also propose two linear-optics based measurements that approach the bound for the estimation of separation parameters once the centroid has been located.

In the final chapter, we close with the concluding remarks and possible future research works.

Chapter 2

Theoretical Background

We assume that the reader is familiar with the basics of quantum mechanics, probability theory, random processes, and mathematics concepts such as Fourier transform and matrix operation. For quantum mechanics, one may read the sections on operators and states in Refs. [4, 5, 6]. Basic concepts about probability theory and random processes can be found in Ref. [7]. For a more advanced text on probability theory, *Probability Theory* by Jaynes and Bretthorst [8] is a great reference.

2.1 Background on Measurement Theory

We begin with a discussion of some elementary notions of measurement of a classical system followed by an introduction to quantum measurement theory.

2.1.1 Classical measurement theory

A classical system can be described by a set of *system variables*, also known as system configuration. The set of all possible values of the system variables forms the configuration space S. For example, there is a classical system that can be described by a single system variable X, the configuration space S would be $\mathbb{R}$, where $\mathbb{R}$ is the real line. We define the

state of this classical system as the probability distribution of X on S. The probability distribution that a discrete variable X has the value x is $P_X(x) = \Pr[X = x]$ and the state is represented by the function $P_X(x)$ for all possible $x \in S$. If X is a continuous variable, probability density $P_X(x)$ is given by $P_X(x)\mathrm{d}x = \Pr[x < X < x + \mathrm{d}x]$ [9]. More generally, X could be a vector of discrete or continuous random variables.

As we have defined earlier, the state of a system describes an observer's knowledge about the system variables. Unless the probability distribution is zero for all except one particular configuration in S, the state is a state of uncertainty (or incomplete knowledge). Note that different observers can assign a different state to a system if they have a different knowledge about it. This is possible as long as all the different states are *consistent* (where there is at least one configuration which every state assigns a nonzero probability). This ensures that there is at least one configuration, that describes the system, every observer agrees on.

With the definition of state explained, we can now consider measurement of a classical system. If a measurement of X is perfect, the observer would have its value, say x'. The state is a state of certainty (or complete knowledge) about this variable. The probability distribution for discrete variable X is given by

$$P(x) = \delta_{xx'}, \tag{2.1}$$

where $\delta_{xx'}$ is the Kronecker delta; or for continuous variable X

$$P(x) = \delta(x - x') \tag{2.2}$$

where $\delta(x - x')$ is the Dirac function. Now suppose that the measurement is taken through an apparatus with system variable Y. By some physical processes, the apparatus variable Y is correlated to system variable X of

interest and is measured. For a given value of system variable, say x, there is a probability distribution for the measurement result y, $P(y|x)$, which defines the measurement and is known as the conditional probability of y given x. In order to determine how the state of the classical system (about X) changes when the value y is obtained from the measurement, we introduce the concept of conditional state of the system or a-posteriori state, $P(x|y)$. For discrete case, Bayes' theorem gives the a-posteriori state as

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}, \tag{2.3}$$

where $P(x)$ is the system state prior to making the measurement and

$$P(y) = \sum_x P(y|x)P(x) \tag{2.4}$$

is the marginal distribution of y irrespective of the value of x. For continuous variables,

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}, \tag{2.5}$$

where

$$P(y) = \int_S \mathrm{d}x P(y|x)P(x). \tag{2.6}$$

The system state after taking the measurement is given by $P(x|y)$.

To conclude this section, for the situation of having complete knowledge about a classical system, the results of any ideal measurements on the system are deterministic. If the measurement is less ideal or there is lack of complete knowledge about the system, the state after measurement is given by Bayes' theorem. In the next section, we discuss the different situation in quantum mechanics.

2.1.2 Quantum measurement theory

In the formulation of quantum mechanics, a quantum *state vector* $|\psi\rangle$ is described as a vector in a complex vector space known as *Hilbert space*. If $\{|\phi_n\rangle\}, n = 0, 1, \ldots$ is an orthonormal basis for this vector space, the state vector can be written as

$$|\psi\rangle = \sum_n c_n |\phi_n\rangle, \tag{2.7}$$

for some complex coefficients c_n with normalization $\sum_n |c_n|^2 = 1$. The probability of getting the nth outcome in an experiment is given by $|c_n|^2$. This uncertainty about future measurement outcomes is fundamental to quantum mechanics theory. However, this quantum state vector is not a quantum analogy of the state of a classical system described in the previous section, as it does not represent our knowledge of the quantum system but the probabilistic nature of quantum mechanics. More precisely, it is called the quantum superposition.

Instead, the quantum version of state that represents our knowledge about a system is the quantum *density operator* $\hat{\rho}$ [5, 9]. For example, suppose a system may be found in state vector $|\psi_0\rangle$ with probability p_0 and $|\psi_1\rangle$ with probability $p_1(= 1 - p_0)$, to the observer the state is given by $\hat{\rho} = p_0|\psi_0\rangle\langle\psi_0| + p_1|\psi_1\rangle\langle\psi_1|$. Generally, the quantum density operator is given by

$$\hat{\rho} = \sum_n p_n |\psi_n\rangle\langle\psi_n|, \tag{2.8}$$

where $\{|\psi_n\rangle\}$ is a spanning set of normalized state vectors, each has a corresponding probability p_n. This mixture of states is known as a *mixed*

state; while a *pure state* is a state with density operator

$$\hat{\rho} = |\psi\rangle\langle\psi|. \tag{2.9}$$

Using the definition in Eq. (2.8) and nonnegativity property of probability function, a density operator $\hat{\rho}$ must be Hermitian, such that

$$\hat{\rho}^\dagger = \hat{\rho}, \tag{2.10}$$

where $\hat{O}^\dagger$ denotes the Hermitian conjugate of $\hat{O}$. Similar, $\hat{\rho}$ is positive-semidefinite, where its expectation

$$\langle\phi|\hat{\rho}|\phi\rangle \geq 0, \tag{2.11}$$

for any state vector $|\phi\rangle$. Also, $\hat{\rho}$ is of trace 1, which follows from the fact that probabilities sum to one. A density operator $\hat{\rho}$ describes a pure state if and only if

$$\hat{\rho}^2 = \hat{\rho}. \tag{2.12}$$

Similar to classical states, different observers may assign different density operators $\hat{\rho}_i$ to the same system based on their knowledge. Consistency is ensured if each different state from all the observers is a mixture of a pure state $|\psi\rangle\langle\psi|$ and some other states. Note that, although we treat the quantum state as a representation of our knowledge about a system, it is actually our knowledge about the outcomes of future measurements on the system.

Projective measurements

The traditional description of measurement in quantum mechanics is in terms of projective measurements. Let $\hat{O}$ be an observable with non-

degenerate eigenvalues $O_m, m = 1, 2, \ldots$ with corresponding eigenvectors $|\phi_m\rangle$, i.e.

$$\hat{O}|\phi_m\rangle = O_m|\phi_m\rangle, \qquad m = 1, 2, \ldots. \tag{2.13}$$

Alternatively, we write

$$\hat{O} = \sum_m O_m \hat{P}_m, \qquad m = 1, 2, \ldots \tag{2.14}$$

where

$$\hat{P}_m = |\phi_m\rangle\langle\phi_m| \tag{2.15}$$

associated with eigenvalue O_m is the projection operator onto the subspace of corresponding eigenstate $|\phi_m\rangle$ of $\hat{O}$. For an observable with degenerate eigenvalue O_m, we let $\hat{P}_m$ be a projection operator onto the subspace of eigenstates associated with the eigenvalue.

A system prepared in pure state $|\psi\rangle$ can be expressed as a superposition of the eigenstates $|\phi_i\rangle$ using Eq. (2.7), since the set $\{|\phi_i\rangle\}$ is a complete basis of the Hilbert space:-

$$|\psi\rangle = \sum_m c_m|\phi_i\rangle, \qquad m = 1, 2, \ldots. \tag{2.16}$$

The eigenvalues O_m are the possible outcomes of the measurement, with corresponding probabilities

$$\Pr[O_m] = |\langle\phi_m|\psi\rangle|^2 = |c_m|^2. \tag{2.17}$$

After the measurement, if the outcome is O_m then the system is in pure state $|\phi_m\rangle$. Any repeated measurement will give the same result O_m.

A mixed state with density operator $\hat{\rho}$ given by Eq. (2.8) is a statistical

ensemble of pure states $|\psi_n\rangle$ with probabilities p_n. Hence, the probability of getting measurement outcome O_m for the mixed state is

$$\begin{aligned}\Pr[O_m] &= \sum_n p_n \times \Pr[O_m, |\psi_n\rangle] \\ &= \sum_n p_n |\langle\phi_m|\psi_n\rangle|^2, \end{aligned} \tag{2.18}$$

where $\Pr[O_m, |\psi_n\rangle]$ is the probability of getting outcome O_m for pure state $|\psi_n\rangle$. We define outcome probability

$$\Pr[O_m] \equiv \mathrm{tr}(\hat{\rho}\hat{P}_m), \tag{2.19}$$

where tr is the trace of operator. Using cyclic property of trace operation (e.g. $\mathrm{tr}(\hat{A}\hat{B}\hat{C}) = \mathrm{tr}(\hat{B}\hat{C}\hat{A}) = \mathrm{tr}(\hat{C}\hat{A}\hat{B})$), one can show that

$$\begin{aligned}\Pr[O_m] &= \langle\phi_m|\hat{\rho}|\phi_m\rangle \\ &= \sum_n p_n |\langle\phi_m|\psi\rangle|^2, \end{aligned} \tag{2.20}$$

which gives the same result as Eq. (2.18).

The density operator after the measurement is

$$\hat{\rho}' = \frac{\hat{P}_m\hat{\rho}\hat{P}_m}{\mathrm{tr}(\hat{\rho}\hat{P}_m)} \tag{2.21}$$

for measurement result O_m. However, if one wishes to describe the unconditional state of the system, ignoring the measurement outcome, the density operator is

$$\hat{\rho}' = \sum_m \hat{P}_m\hat{\rho}\hat{P}_m. \tag{2.22}$$

From what we have discussed, in quantum mechanics the results of measurements are generally statistical. One of the aims of quantum mea-

surement theory is to be able to specify the probability of a particular measurement result and the state of the system immediately after the measurement, given the initial state of a system.

Generalized measurements

Although useful, projective measurements are inadequate in describing quantum measurements, as in a real-world experiment scientists never measure a system directly. Instead, the system-of-interest interacts with a probe or apparatus, and the scientists observe the effect of the system on the probe. For example, in photodetection there is simply no photon remains after the measurement. The state of photons is not projected into some other states post measurement as predicted in projective measurement. Hence, there is a need for a more general description of quantum measurements.

Suppose the initial state of the system is given by pure state $|\psi\rangle$, and the probe has initial state $|\eta\rangle$. The initial combined state which describes the system and probe is

$$|\Psi\rangle = |\eta\rangle|\psi\rangle. \tag{2.23}$$

Let the combined state interacts for time t, under a unitary evolution operator $\hat{U}_t$, the combined state becomes

$$|\Psi'\rangle = \hat{U}_t|\eta\rangle|\psi\rangle, \tag{2.24}$$

which cannot be written in the factorized form of Eq. (2.23).

The scientist observer measures projectively the probe, so that the operator $|m\rangle\langle m| \otimes \hat{\mathbb{I}}$ acts on the combined state. The set $\{|m\rangle\}$ forms an orthonormal basis for the probe Hilbert space and identity operator $\hat{\mathbb{I}}$ is in

the system Hilbert space. The combined state after measurement is

$$|\Psi''\rangle = \frac{|m\rangle\langle m|\hat{U}_t|\eta\rangle|\psi\rangle}{\mathcal{N}}, \tag{2.25}$$

where $\mathcal{N}$ is the normalization constant. The state can be written as [9]

$$|\Psi''\rangle = |m\rangle\frac{\hat{M}_m|\psi\rangle}{\mathcal{N}}, \tag{2.26}$$

where measurement operator $\hat{M}_m = \langle m|\hat{U}_t|\eta\rangle$ acts only in the system Hilbert space. And the system state immediately after the measurement is $\hat{M}_m|\psi\rangle$. Therefore, we can formulate the postulate for generalized measurements [5].

Postulate (measurements). Quantum measurements are described by a collection $\{\hat{M}_m\}$ of measurement operators. These are operators acting on the Hilbert space of system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $|\psi\rangle$ (or ρ for mixed state) immediately before the measurement then the probability that result m occurs is given by

$$\Pr[m] = \langle\psi|\hat{M}_m^\dagger\hat{M}_m|\psi\rangle \qquad \text{or} \qquad \Pr[m] = \mathrm{tr}(\hat{\rho}\hat{M}_m^\dagger\hat{M}_m), \tag{2.27}$$

and the state immediately after the measurement is

$$|\psi'\rangle = \frac{\hat{M}_m|\psi\rangle}{\sqrt{\langle\psi|\hat{M}_m^\dagger\hat{M}_m|\psi\rangle}} \qquad \text{or} \qquad \hat{\rho}' = \frac{\hat{M}_m\hat{\rho}\hat{M}_m^\dagger}{\mathrm{tr}(\hat{M}_m^\dagger\hat{M}_m\hat{\rho})}. \tag{2.28}$$

As shown in Eq. (2.27), the probabilities are given by the expectation of another operator,

$$\hat{E}_m = \hat{M}_m^\dagger\hat{M}_m, \tag{2.29}$$

known as effect or probability operator, and is defined in terms of the measurement operator. The set of $\{E_m\}$ constitutes a *positive-operator valued measure* (POVM) on the space of measurement results m [9, 10].

Properties of POVM

The measurement operators satisfy the completeness relation

$$\sum_m \hat{M}_m^\dagger \hat{M}_m = \hat{\mathbb{I}} \tag{2.30}$$

from the fact that the outcome probabilities add up to unity. Hence, we have $\{\hat{E}_m\}$ of a POVM sum to the identity operator:

$$\sum_m \hat{E}_m = \hat{\mathbb{I}}, \qquad \hat{E}_m = \hat{M}_m^\dagger \hat{M}_m. \tag{2.31}$$

Similarly, the projection operators of Eq. (2.15) also sum to the identity,

$$\sum_m \hat{P}_m = \hat{\mathbb{I}}, \qquad \hat{P}_m = |\phi_m\rangle\langle\phi_m|, \tag{2.32}$$

as $\{|\phi_m\rangle\}$ is an orthonormal basis of the Hilbert space they act in.

We can view POVM as a generalization of projective measurement. The main difference is that the elements of a POVM, $\hat{E}_m$, are not orthogonal in general; while the set of projector operators has the following property:-

$$\hat{P}_m \hat{P}_n = \hat{P}_m \delta_{mn}. \tag{2.33}$$

This implies that the total number of elements in the POVM, M can be greater than the dimension of the Hilbert space. It follows from the definition given by Eq. (2.29), the operators $\hat{E}_m$ are positive-semidefinite. In fact, there are only two restrictions on $\{\hat{E}_m\}$ as a POVM [9]. First, they must be positive-semidefinite operators. Second, they must be a resolution of the identity for the system Hilbert space, i.e. Eq. (2.31).

2.2 Background on Parameter Estimation Theory

Many quantities of interest in physics are not directly accessible. For example, in quantum mechanics quantities like entanglement and purity do not correspond to proper quantum observables. Facing these situations scientists have to resort to indirect measurement, inferring the values of these quantities from measurement of other observables. Hence, there is a need for a theory on parameter estimation.

In this section, we develop the basic ideas in classical and quantum estimation theory.

2.2.1 Parameter estimation theory

The goal of parameter estimation theory is to estimate the value of a *parameter* vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)^\top \in \mathbb{R}^m$ based on the data Y acquired by observing a system. The acquired data has a random component and its distribution depends on the true value of θ. We consider the cases where Y takes discrete or continuous values over a domain $\mathcal{Y}$. In discrete case, the domain $\mathcal{Y} = \{\mathcal{Y}_i : i \in \mathbb{N}\}$ is a countable set of discrete values $\mathcal{Y}_i \in \mathbb{R}^n$; while in the continuous case, the domain $\mathcal{Y} = \mathbb{R}^n$.

The observation data $y = (y_1, y_2, \ldots, y_n)^\top$ are random variables with joint probability density function,

$$P_Y(y|\theta) = P_Y(y_1, y_2, \ldots, y_n|\theta_1, \theta_2, \ldots, \theta_m), \tag{2.34}$$

depends on parameters $\theta_1, \theta_2, \ldots, \theta_m$. The estimate $\check{\theta} = (\check{\theta}_1, \check{\theta}_2, \ldots, \check{\theta}_m)^\top$ is a function

$$\check{\theta} = \check{\theta}(y), \tag{2.35}$$

of the observation data and takes values in $\mathbb{R}^m$. The function describes a strategy for finding the estimates from the data and is known as *estimator* [1, 2].

An estimator is *unbiased* when its expected value is equal to the true value of the parameter,

$$\mathbb{E}_Y \left[\check{\theta}(Y)|\theta\right] = \theta, \tag{2.36}$$

where

$$\mathbb{E}_Y[g(y)|\theta] = \sum_{\mathcal{Y}} g(y) P_Y(y|\theta) \tag{2.37}$$

for discrete case, and

$$\mathbb{E}_Y[g(y)|\theta] = \int_{\mathcal{Y}} \mathrm{d}^n y g(y) P_Y(y|\theta) \tag{2.38}$$

for continuous case. A biased estimate has nonzero bias given by

$$b = \mathbb{E}_Y[\check{\theta}(y)|\theta] - \theta. \tag{2.39}$$

To find the optimum strategy in constructing an estimator, we need to introduce a concept of *loss function*. A loss function or cost function is a function expressing the cost of estimating θ as $\check{\theta}(y)$ [1, 2]. The value of a loss function is random as it depends on the outcome of random variable $\check{\theta}(y)$. For single parameter or $m = 1$, a common loss function is the squared error,

$$C = (\check{\theta} - \theta)^2. \tag{2.40}$$

For multiple parameters, one may adopt loss functions, for example,

$$C = \sum_{i=0}^{m} \sum_{j=0}^{m} g_{ij} (\check{\theta}_i - \theta_i)(\check{\theta}_j - \theta_j). \tag{2.41}$$

where g_{ij} is the ij-element of a positive-definite matrix, or,

$$C' = \sum_{i=0}^{m} (\check{\theta}_i - \theta_i)^2, \tag{2.42}$$

a special case where g is the identity matrix. Most of the optimum estimators are constructed such that their expected loss is minimum.

2.2.2 Examples of estimators

In this section, we introduce some common estimators which we study in detail in the following three chapters. First is the most commonly studied estimator, the *maximum-likelihood (ML) estimator*.

Maximum-likelihood (ML)

To apply ML estimator [1, 2], we first specify the joint probability density function $P_Y(y|\theta)$ given by Eq. (2.34) for all observations. ML estimates the set of parameters θ_{ML} by finding value of θ that maximizes the log-likelihood function $\ln P_Y(y|\theta)$:

$$\check{\theta}_{\mathrm{ML}} = \arg\max_{\theta} \ln P_Y(y|\theta). \tag{2.43}$$

The main advantage of ML estimator is its efficiency, it achieves the Cramér-Rao lower bound (which we introduce in Sec. 2.2.3) on mean-square estimation error asymptotically [2]. However, in practice it might be difficult to derive or computationally expensive to evaluate.

Expectation-Maximization (EM) Algorithm

In ML estimation, we need to solve Eq. (2.43) for maxima which typically involves the derivatives of the likelihood function. However, there are cases where the equations involved cannot be solved directly and simultaneously. The *Expectation-Maximization (EM) algorithm* [2, 11, 12] can significantly

simplify this task if there exists hidden data Z that results in simplified expressions for $P(Z|Y,\theta)$ and $P(Y,Z|\theta)$.

Starting with a trial $\theta = \theta^0$, the EM algorithm iteratively applies the following two steps:

1. Expectation step: Consider the estimated log-likelihood function

$$Q(\theta,\theta^k) \equiv \int \mathrm{d}Z P(Z|Y,\theta^k) \ln P(Y,Z|\theta), \tag{2.44}$$

   where the superscript k is an index denoting the EM algorithm iteration.

2. Maximization step: By maximizing $Q(\theta,\theta^k)$, we find a new estimate θ^{k+1} for the next iteration, i.e.

$$\theta^{k+1} = \arg\max_{\theta} Q(\theta,\theta^k). \tag{2.45}$$

The iteration is halted if the difference or relative difference between θ^k and θ^{k+1} is smaller than a preset threshold value, and the EM estimate $\check{\theta}_{\mathrm{EM}}$ takes the value of the final θ^{k+1}. In general, EM algorithm converges to a local optimum and there is no guarantee of global convergence [2]. Techniques, such as random restart, could be useful for EM algorithm in escaping convergence to a local optimum.

Radiometer

For a more specific problem of estimating the amplitude A of the spectrum of a stationary process with power spectral density (PSD) in the form of

$$S_Y(\omega) \equiv \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\left|\int_{-T/2}^{T/2} \mathrm{d}t y(t)\exp(i\omega t)\right|^2\right] \tag{2.46}$$

$$= AS(\omega) + R, \tag{2.47}$$

where R is the power of an additive white noise. The 'radiometer' estimator described in Ref. [13] gives

$$\check{A}_{\text{rad}} = G\bigg[\int_{-T/2}^{T/2} \mathrm{d}t \int_{-T/2}^{T/2} \mathrm{d}t' y(t)h(t-t')y(t') - B\bigg]. \tag{2.48}$$

Function $h(t-t')$ filters $y(t')$ before correlating the results with $y^*(t)$, and G and B are the parameters chosen to enforce the unbiased condition:

$$G = \left[T\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} H(\omega)S(\omega)\right]^{-1}, \tag{2.49}$$

$$B = T\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} H(\omega)R, \tag{2.50}$$

$$H(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t h(t)\exp(-i\omega t). \tag{2.51}$$

Radiometer estimator is easy to implement, however, it is a single parameter estimator while other three estimators are for multiple parameters.

Whittle

The assumption of stationary processes and a long observation time T (relative to all other time scales in the problem) is known as the SPLOT assumption. In cases that likelihood function P_Y is complicated, there will be difficulty in deriving the ML estimate. The *Whittle estimator* [14] exploits the SPLOT assumption to simplify the likelihood function. Consider a real discrete-time series

$$\{y(t_k); k = 0, 1, \ldots, K-1\}, \qquad t_k = k\delta t, \tag{2.52}$$

with zero-mean Gaussian statistics conditioned on parameter θ. Define the discrete Fourier transform as

$$Y_m = \frac{\delta t}{\sqrt{T}} \sum_{k=0}^{K-1} y(t_k)\exp(i\omega_m t_k), \qquad \omega_m = \frac{2\pi m}{T}, \tag{2.53}$$

with integer m. Since $y(t_k)$ is real, $Y_m = Y^*_{K-m}$. In the SPLOT case, it can be shown that the positive components $\{y_m; 0 < m < K/2\}$ are independent complex Gaussian random variables with means zero and variances $S_Y(\omega_m|\theta)$ [12, 14]. This means that the log-likelihood function can be approximated as

$$\ln P(Y|\theta) \approx \mathcal{A} - \sum_{0<m<K/2} \left[\ln S_Y(\omega_m|\theta) + \frac{|Y_m|^2}{S_Y(\omega_m|\theta)} \right], \tag{2.54}$$

where $\mathcal{A}$ is independent of θ. Whittle estimator is then given by

$$\check{\theta}_{\mathrm{W}} = \arg\min_{\theta} f(\theta). \tag{2.55}$$

where

$$f(\theta) = \sum_{0<m<K/2} \left[\ln S_Y(\omega_m|\theta) + \frac{|Ym|^2}{S_Y(\omega_m|\theta)} \right]. \tag{2.56}$$

As Whittle likelihood is an approximation to the log-likelihood function, the efficiency of Whittle estimator depends on how exact this approximation is. Similar to ML estimator, Whittle estimator is asymptotically efficient [12] and typically less computational expensive.

2.2.3 Cramér-Rao (CR) lower bound

A useful figure of merit for evaluating an estimator performance is the mean-square estimation error matrix Σ,

$$\Sigma \equiv \mathbb{E}_Y\{[\check{\theta}(Y) - \theta][\check{\theta}(Y) - \theta]^\top\}. \tag{2.57}$$

The *Cramér-Rao (CR) bound* gives a lower bound on the mean-square estimation error matrix for any unbiased estimator $\check{\theta}$ [2, 10]:

$$\Sigma \geq j^{-1}(P_Y), \tag{2.58}$$

where the matrix inequality means that $\Sigma - j^{-1}$ is positive-semidefinite, or

$$u^\top(\Sigma - j^{-1})u \geq 0, \tag{2.59}$$

for any real column vector u. Note that the observation Y has probability distribution $P_Y(y|\theta)$ dependent on parameter vector θ given by Eq. (2.34), the classical *Fisher information* matrix $j(P_Y)$ characterizes the information in Y about θ. The Fisher information matrix is a $m \times m$ matrix, given by

$$j(P_Y) = \mathbb{E}_Y \left\{ \nabla[\ln P_Y(Y|\theta)] \nabla^\top [\ln P_Y(Y|\theta)] | \theta \right\}, \tag{2.60}$$

where

$$\nabla \equiv \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_m}, \right)^\top. \tag{2.61}$$

Proof: Given any estimator $\check{\theta}$, assuming unbiased condition defined in Eq. (2.36), we have

$$\int_{\mathcal{Y}} \mathrm{d}^n y (\check{\theta} - \theta) P_Y(y|\theta) = 0. \tag{2.62}$$

Differentiate with respect to $\theta_i, \quad i = 1, 2, \ldots, m,$

$$\int_{\mathcal{Y}} \mathrm{d}^n y (-I) P_Y(y|\theta) + \int_{\mathcal{Y}} \mathrm{d}^n y (\check{\theta} - \theta) \nabla^\top P_Y(y|\theta) = 0, \tag{2.63}$$

where I is an $m \times m$ identity matrix and ∇ is defined in Eq. (2.61). As the probability density $P_Y(y|\theta)$ is normalized, we have

$$\int_{\mathcal{Y}} \mathrm{d}^n y(\check{\theta} - \theta)\nabla^\top P_Y(y|\theta) = I,$$
$$\int_{\mathcal{Y}} \mathrm{d}^n y P_Y(y|\theta)(\check{\theta} - \theta)\nabla^\top \ln P_Y(y|\theta) = I. \tag{2.64}$$

Introduce arbitrary real column vectors u and v of the same dimension as θ, multiply Eq. (2.64) from the left by $u^\top$ and from the right by v,

$$\int_{\mathcal{Y}} \mathrm{d}^n y[\sqrt{P_Y(y|\theta)}u^\top(\check{\theta} - \theta)][\nabla^\top \ln P_Y(y|\theta)v\sqrt{P_Y(y|\theta)}] = u^\top v. \tag{2.65}$$

We square the equation and by the Cauchy-Schwarz inequality,

$$\begin{aligned} |u^\top v|^2 &\leq \int_{\mathcal{Y}} \mathrm{d}^n y P_Y(y|\theta)|u^\top(\check{\theta} - \theta)|^2 \int_{\mathcal{Y}} \mathrm{d}^n y|\nabla^\top \ln P_Y(y|\theta)v|^2 P_Y(y|\theta) \\ &= (u^\top \Sigma u)(v^\top j v), \end{aligned} \tag{2.66}$$

where Σ is given by Eq. (2.57) and j is given by Eq. (2.60). Choosing $v = j^{-1}u$, we have inequality

$$u^\top \Sigma u \geq u^\top j^{-1} u, \tag{2.67}$$

for any arbitrary column vector u, proving the CR lower bound.

Apart from Eq. (2.60), the Fisher information matrix has a second form:

$$j(P_Z) = -\mathbb{E}_Y[\nabla\nabla^\top \ln P_Y(Y|\theta)|\theta], \tag{2.68}$$

since we have

$$\nabla\nabla^\top \ln P_Y(Y|\theta) = -[\nabla \ln P_Y(Y|\theta)][\nabla^\top \ln P_Y(Y|\theta)] + \frac{\nabla\nabla^\top P_Y(Y|\theta)}{P_Y(Y|\theta)}, \tag{2.69}$$

and expectation

$$\mathbb{E}_Y\left[\frac{\nabla\nabla^\top P_Y(Y|\theta)}{P_Y(Y|\theta)}\right] = \nabla\nabla^\top \int_{\mathcal{Y}} \mathrm{d}^n y P_Y(Y|\theta) = 0. \tag{2.70}$$

In the asymptotic limit, ML estimation can attain the CR bound [12], so the bound is a meaningful indicator of estimation error. An *efficient* estimator is an estimator that saturates the CR bound. In this case, ML estimators are asymptotically efficient estimators.

2.2.4 Quantum estimation theory

In Sec. 2.1.2, we saw that in quantum mechanics the outcomes of measurements are random variables even when one has complete knowledge of the state. Here, we are concerned with estimating the classical parameter that is used in preparing the state.

Let $\hat{\rho} = \hat{\rho}(\theta)$ be the density operator of a quantum system depending on an unknown vector parameter θ. Consider the estimation of θ from the quantum measurement outcome Y on $\hat{\rho}$. The probability density of Y is given by

$$P_Y(Y|\theta) = \mathrm{tr}[\hat{E}(Y)\hat{\rho}(\theta)], \tag{2.71}$$

where $\hat{E}(Y)$ is the POVM that characterizes the statistics of the quantum measurement [9, 10]. Quantum estimation theory seeks POVMs that maximize the Fisher information, and minimize estimation errors.

The simplest way to characterize an estimation error is to use the estimation error covariance matrix. However, it is not easy to derive a POVM that minimize the errors, therefore, we find a lower bound for this estimation error covariance matrix for all possible POVMs in the following section.

2.2.5 Quantum Cramér-Rao (QCR) bound

Let $\hat{\rho}$ be the θ-dependence density operator that describes a quantum system and $\hat{E}(Y)$ be the POVM that models the quantum measurement on the system, the outcome Y of the measurement has probability distribution $P_Y(y)$ given by Eq. (2.71).

Given an estimate $\check{\theta}(Y)$ of θ, the estimation error covariance matrix Σ,

$$\Sigma_{\mu\nu} \equiv \mathbb{E}\left\{\left[\check{\theta}_\mu(Y) - \theta_\mu\right]\left[\check{\theta}_\nu(Y) - \theta_\nu\right]\right\}, \tag{2.72}$$

$$\mathbb{E}[g(y)] \equiv \int \mathrm{d}y P_Y(y) g(y). \tag{2.73}$$

In the case of unbiased estimator, defined in Eq. (2.36), Σ is bounded by the classical and quantum CR bounds [1, 10, 15, 16],

$$\Sigma \geq j^{-1} \geq J^{-1}. \tag{2.74}$$

Here, the inequalities mean that matrices $\Sigma - j^{-1}$, $\Sigma - J^{-1}$ and $j^{-1} - J^{-1}$ are positive-semidefinite. Matrix j is the classical Fisher information matrix defined in Eq. (2.60) and matrix J is the *quantum Fisher information* (QFI) matrix.

Proof: We follow closely the method used by Helstrom [10]. For unbiased estimator with probability density given by Eq. (2.71),

$$\int_{\mathcal{Y}} d^n y (\check{\theta}_\mu - \theta_\mu) \operatorname{tr}[\hat{E}(Y)\hat{\rho}] = 0. \tag{2.75}$$

Differentiate with respect to θ_ν,

$$\int_{\mathcal{Y}} d^n y (\check{\theta}_\mu - \theta_\mu) \operatorname{tr}[\hat{E}(Y)\frac{\partial \hat{\rho}}{\partial \theta_\nu}] = \delta_{\mu\nu}, \tag{2.76}$$

making use of the fact that the probability density integrates to unity.

Introduce the *symmetric logarithmic derivative* (SLD) operator $\hat{\mathcal{L}}_\nu$ as the Hermitian operator satisfying the equation

$$\frac{\partial\hat{\rho}}{\partial\theta_\nu} = \frac{1}{2}(\hat{\mathcal{L}}_\nu\hat{\rho} + \hat{\rho}\hat{\mathcal{L}}_\nu), \tag{2.77}$$

Eq. (2.76) becomes

$$\mathrm{Re}\{\mathrm{tr}[\int_{\mathcal{Y}} d^n y(\check{\theta}_\mu - \theta_\mu)\hat{\rho}\hat{\mathcal{L}}_\nu\hat{E}(Y)]\} = \delta_{\mu\nu}, \tag{2.78}$$

by swapping the trace and integration operation, and using the property of Hermitian operators $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_n$,

$$\mathrm{tr}(\hat{B}_1\hat{B}_2\ldots\hat{B}_n) = [\mathrm{tr}(\hat{B}_n\ldots\hat{B}_2\hat{B}_1)]^*, \tag{2.79}$$

where c^* denotes complex conjugate of a complex number c. Let $u = (u_1, u_2, \ldots u_m)^\top$ and $v = (v_1, v_2, \ldots v_m)^\top$ be two arbitrary real column vectors, multiply Eq. (2.78) by the real number $u_\mu v_\nu$ and sum over μ and ν, we obtain

$$\mathrm{Re}\{\mathrm{tr}[\int_{\mathcal{Y}} d^n y A\hat{\rho}\hat{E}(Y)\hat{B}]\} = u^\top v, \tag{2.80}$$

where

$$\begin{aligned} A &= \sum_{\mu=1}^{m} u_\mu(\check{\theta}_\mu - \theta_\mu), \\ \hat{B} &= \sum_{\nu=1}^{m} v_\nu\hat{\mathcal{L}}_\nu. \end{aligned} \tag{2.81}$$

Before we continue, it can be easily shown that the Cauchy-Schwarz

inequality

$$\mathrm{tr}[\hat{\rho}\int_{\mathcal{Y}}\mathrm{d}^n y\hat{C}^\dagger\hat{E}(Y)\hat{C}]\,\mathrm{tr}[\hat{\rho}\int_{\mathcal{Y}}\mathrm{d}^n y\hat{D}^\dagger\hat{E}(Y)\hat{D}]\geq\big|\,\mathrm{tr}[\hat{\rho}\int_{\mathcal{Y}}\mathrm{d}^n y\hat{D}^\dagger\hat{E}(Y)\hat{C}]\big|^2, \tag{2.82}$$

for any $\hat{C}$ and $\hat{D}$. The proof follows from the expansion of

$$\mathrm{tr}[\hat{\rho}\int_{\mathcal{Y}}\mathrm{d}^n y(\hat{C}-\lambda\hat{D})^\dagger\hat{E}(Y)(\hat{C}-\lambda\hat{D})]\geq 0 \tag{2.83}$$

for complex λ, and Hermitian $\hat{\rho}$ and $\hat{E}(Y)$. Finally, set

$$\lambda=\frac{\int_{\mathcal{Y}}\mathrm{d}^n y\hat{D}^\dagger\hat{E}(Y)\hat{C}}{\int_{\mathcal{Y}}\mathrm{d}^n y\hat{D}^\dagger\hat{E}(Y)\hat{D}}. \tag{2.84}$$

Squaring both sides of Eq. (2.80), we have

$$\begin{aligned}(u^\top v)^2&\leq\big|\,\mathrm{tr}[\hat{\rho}\int_{\mathcal{Y}}d^n yA\hat{E}(Y)\hat{B}]\big|^2\\&\leq\mathrm{tr}[\hat{\rho}\int_{\mathcal{Y}}d^n yA\hat{E}(Y)A]\,\mathrm{tr}[\hat{\rho}\int_{\mathcal{Y}}d^n y\hat{B}\hat{E}(Y)\hat{B}]\\&=[\int_{\mathcal{Y}}d^n yA^2P(Y|\theta)]\,\mathrm{tr}[\hat{\rho}\hat{B}\hat{B}]\\&=(u^\top\Sigma u)(v^\top Jv)\end{aligned} \tag{2.85}$$

where

$$J_{\mu\nu}=\frac{1}{2}\,\mathrm{tr}[\hat{\rho}(\hat{\mathcal{L}}_\mu\hat{\mathcal{L}}_\nu+\hat{\mathcal{L}}_\nu\hat{\mathcal{L}}_\mu)], \tag{2.86}$$

and the second inequality is given by the Cauchy-Schwarz inequality. As u and v are arbitrary real vectors, we can choose $u=Jv$. The result is

$$\Sigma\geq J^{-1}. \tag{2.87}$$

The second part of the inequalities relating classical and quantum Fisher

information was proven by Braunstein and Caves [17] for scalar θ. It can be generalized for multiple parameters by transforming

$$\frac{\partial}{\partial\theta} \to u_\mu \frac{\partial}{\partial\theta_\mu}, \tag{2.88}$$

where u is any real vector of the same dimension as vectorial parameter θ. From Eq. (2.71) and applying Eq. (2.77), we have

$$\begin{aligned} u_\mu \frac{\partial P_Y(y|\theta)}{\partial\theta_\mu} &= \mathrm{tr}[\hat{E}(Y) u_\mu \frac{\partial\hat{\rho}}{\partial\theta_\mu}] = \mathrm{Re}\{\mathrm{tr}[\hat{E}(Y) u_\mu \hat{L}_\mu \hat{\rho}]\}, \\ \left[u_\mu \frac{\partial P_Y(y|\theta)}{\partial\theta_\mu}\right]^2 &\le |\,\mathrm{tr}[\hat{E}(Y) u_\mu \hat{L}_\mu \hat{\rho}]|^2 \\ &= \left|\,\mathrm{tr}\left[\sqrt{\hat{\rho}}\sqrt{\hat{E}(Y)}\sqrt{\hat{E}(Y)} u_\mu \hat{L}_\mu \sqrt{\hat{\rho}}\right]\right|^2 \\ &\le \mathrm{tr}\left[\sqrt{\hat{\rho}}\hat{E}(Y)\sqrt{\hat{\rho}}\right] \mathrm{tr}\left[\sqrt{\hat{\rho}} u_\mu \hat{L}_\mu \hat{E}(Y) \hat{L}_\nu u_\nu \sqrt{\hat{\rho}}\right], \end{aligned} \tag{2.89}$$

where the second inequality makes use of Cauchy-Schwarz inequality. With cyclic property of trace,

$$u_\mu \frac{\partial P_Y(y|\theta)}{\partial\theta_\mu} \frac{\partial P_Y(y|\theta)}{\partial\theta_\nu} u_\nu \le P_Y(y|\theta) u_\mu \,\mathrm{tr}\left[E(Y)\hat{L}_\nu \hat{\rho} \hat{L}_\mu\right] u_\nu, \tag{2.90}$$

$$u_\mu j_{\mu\nu} u_\nu \le u_\mu \,\mathrm{tr}\left[\hat{\rho}\hat{L}_\mu \hat{L}_\nu\right] u_\nu. \tag{2.91}$$

Since u can be any vector, we obtain matrix inequality

$$j \le J. \tag{2.92}$$

Recent theoretical studies in quantum parameter estimation have proved the existence of a POVM that attains the QCR bound for scalar parameter estimation [18, 19], though the quantum bound may not be attainable for vectorial parameter.

Apart from the SLD given by Eq. (2.77), there is another way of defining the derivatives of the quantum density operator $\hat{\rho}$. The right logarithmic

derivative operators $\hat{L}_j$ is defined by [10, 20]

$$\frac{\partial \hat{\rho}}{\partial \theta_\nu} = \hat{\rho}\hat{L}_\nu = \hat{L}_\nu^\dagger \hat{\rho}. \tag{2.93}$$

These non-Hermitian operators $\hat{L}_j$ may not exist for some density operators. The alternative form of the quantum bound is

$$u^\dagger \Sigma u \geq u^\dagger A^{-1} u, \tag{2.94}$$

where u is an arbitrary complex vector and

$$A_{\mu\nu} = \mathrm{tr}(\hat{\rho}\hat{L}_\mu \hat{L}_\nu). \tag{2.95}$$

However, in this thesis we will focus solely on the QCR bound based on the SLD. We will present a series of results about parameter estimation theory for optical systems in the following three chapters.

Chapter 3

Parameter Estimation with Optomechanical Systems

In this chapter, we introduce a statistical framework to study the problem of parameter estimation from a noisy optomechanical system. This problem is especially relevant to the recent optomechanics experiments reported in Refs. [21, 22]. We derive analytic expressions for the CR lower bound on the estimation errors in term of PSD of the signal in Sec. 3.4, and apply various estimation techniques to the experimental data to estimate the parameters of an optomechanical system in Sec. 3.6, including force-noise power, mechanical resonance frequency, damping rate and measurement noise power. The analytic results provide convenient expressions of the estimation errors as a function of system parameters and measurement time, and should be valuable to optomechanical experiment designs. This chapter is based on the work published in Ref. [23].

3.1 Motivation

Recent spectacular advances in optomechanical oscillators [3] for force-sensing applications enable ultra-sensitive force measurements of charge, single spin, mass, acceleration and magnetic field [24, 25, 26, 27, 28]. These

advances in experimental physics allow scientists to study quantum light-matter interactions in macroscopic structures [29, 30], hence pave way towards new technologies for quantum information science and quantum metrology [31, 32, 33, 34].

While thermal and measurement noises impose major limitations to the accuracy of mechanical force sensors, the higher quality and lower mass optomechanical oscillators with increased force sensitivity [21] has higher sensitivity to their environment. This introduces additional noises that can cause fluctuations in system parameters such as effective oscillator temperature and mechanical resonance frequency. For optomechanical technology to continue to advance, there is a need to develop methods that can characterize, monitor and control these new noise sources, along with thermal and measurement noises.

3.2 Experiment

Before we develop our theoretical model and statistical framework, we first describe the experiment presented in Ref. [22]. The setup of the microtoroidal cavity optomechanical system consists of a room temperature microtoroidal resonator that supports simultaneously mechanical modes and whispering gallery optical modes. The excited mechanical mode is sensitive to external forces, having fundamental frequency $\Omega_m = 40.33$ MHz, damping rate $\gamma = 23$ kHz, and effective mass of $m_{\text{eff}} = 7$ ng. A shot-noise limited 1550 nm laser was coupled evanescently into the whispering gallery mode of the microtoroid via a tapered optical fiber, permitting ultra-precise readout of the mechanical displacement. The excitation of the mechanical mode induces phase fluctuations on the transmitted light which is measured by shot-noise limited homodyne detection.

To specifically demonstrate power estimation, a small incoherent elec-

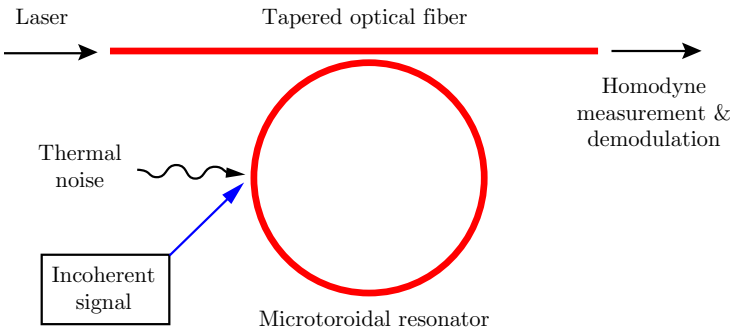


Figure 3.1: Conceptual schematics of the experiment. A mechanical oscillator with thermal noise and applied signal acting on it.

trostatic gradient force is applied to the mechanical oscillator by a nearby electrode driven with a white-noise signal [35] as depicted in Fig. 3.1. The measurement record is acquired from the homodyne signal by electronic lock-in detection which involves demodulation of the photocurrent at the mechanical resonance frequency to allow real time measurement of the slowly evolving quadratures of motion, denoted by $I(t)$ and $Q(t)$ where $x(t) = I(t)\cos\Omega_m t + Q(t)\sin\Omega_m t$. The room temperature thermal fluctuations of the mechanical mode are observed with a signal-to-noise ratio of 37 dB and calibrated via the optical response to a known reference modulation [36].

The resulting force sensitivity will depend on the specific protocol used to extract the information from the measurement record. Here we evaluate the force sensitivity of the three parameter estimation protocols presented in Sec. 3.5 relative to the CR lower bound on estimation error introduced in Sec. 2.2.3.

3.3 Theoretical Model

As the measurement record can be expressed in the form of

$$x(t) = \frac{1}{2}[I(t) - iQ(t)]\exp(i\Omega_m t) + \frac{1}{2}[I(t) + iQ(t)]\exp(-i\Omega_m t), \qquad (3.1)$$

we let $y(t) = I(t) + iQ(t)$ be the new complex measurement record.

We model the mechanical mode of the oscillator using a simple linear Gaussian model. $z(t)$ is the complex analytic signal of the mechanical-mode displacement with the following equation of motion:-

$$\frac{\mathrm{d}z(t)}{\mathrm{d}t} = -\gamma z(t) + i\Omega z(t) + \xi(t), \tag{3.2}$$

where Ω is the mechanical resonance frequency relative to Ω_m, γ is the damping rate and $\xi(t)$ is the stochastic force as a sum of the thermal noise and the signal. $\xi(t)$ is assumed to be a complex zero-mean white Gaussian noise with power A and covariance function

$$\mathbb{E}[\xi(t)\xi^*(t')] = A\delta(t - t'), \tag{3.3}$$

$$\mathbb{E}[\xi(t)\xi(t')] = 0. \tag{3.4}$$

The measurements can be modeled in continuous time as

$$y(t) = Cz(t) + \eta(t), \tag{3.5}$$

where C is a real parameter and $\eta(t)$ is the measurement noise, assumed to be a complex white Gaussian noise with power R and covariance function

$$\mathbb{E}[\eta(t)\eta^*(t')] = R\delta(t - t'), \tag{3.6}$$

$$\mathbb{E}[\eta(t)\eta(t')] = 0. \tag{3.7}$$

We further assume that the parameters

$$\theta = (\Omega, \gamma, A, C, R)^\top \tag{3.8}$$

are constant in time, such that $z(t), \xi(t), y(t)$ and $\eta(t)$ are stationary stochas-

tic processes given θ. In particular, $y(t)$ has a PSD given by

$$S_Y(\omega|\theta) = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\left|\int_{-T/2}^{T/2} \mathrm{d}t y(t)\exp(-i\omega t)\right|^2\right]$$
$$= AS(\omega) + R. \tag{3.9}$$
$$S(\omega) = \frac{C^2}{(\omega-\Omega)^2+\gamma^2}. \tag{3.10}$$

Although this simple model suffices to describe the experiment, it is not difficult to generalize our entire formalism to describe more complicated dynamics and colored noise [1]. This can be done by generalizing $z(t)$ to a vector of state variables for more mechanical and optical modes, Eq. (3.2) to a vectoral equation of motion, and the parameters $(\Omega, \gamma, A, C, R)$ to matrices that describe the coupled-mode dynamics and the noise statistics.

3.4 Review of CR Lower Bound

We now consider the estimation of θ defined in Eq. (3.8) from our measurement data Y given by Eq. (3.5). Let the estimate be $\check{\theta}(Y)$ and the probability density of Y be $P_Y(y|\theta)$. The mean-square estimation error matrix Σ is defined in Eq. (2.57),

$$\Sigma \equiv \mathbb{E}_Y\left\{[\check{\theta}(Y)-\theta][\check{\theta}(Y)-\theta]^\top|\theta\right\}. \tag{3.11}$$

Assuming $\check{\theta}$ satisfies the unbiased condition, the CR lower bound on Σ is

$$\Sigma \geq j^{-1}(P_Y), \tag{3.12}$$

where $j(P_Y)$ is the Fisher information matrix. For dynamical systems, e.g. Y of Eq. (3.5), $j(P_Y)$ might be difficult to evaluate directly using the definition given by Eq. (2.60) or Eq. (2.68).

It turns out that the Fisher information matrix can be related to the

Bhattacharyya distance in hypothesis testing problem. Bhattacharyya distance between two probability densities $P(x|\theta)$ and $P(x|\theta')$ is [37]

$$B(\theta, \theta') = -\ln \rho, \tag{3.13}$$

where ρ is the Bhattacharyya coefficient between the two probability densities

$$\rho \equiv \int \mathrm{d}x \sqrt{P(x|\theta)P(x|\theta')}. \tag{3.14}$$

Taylor series of $\rho(\theta') = \exp[B(\theta, \theta')]$ at θ is the power series

$$\begin{aligned}\rho = \exp[B(\theta, \theta')]|_{\theta'=\theta} + (\theta' - \theta)^\top [-(\nabla B)\exp(-B)]_{\theta'=\theta} \\ + \frac{1}{2}(\theta' - \theta)^\top [-(\nabla\nabla^\top B)\exp(-B) + (\nabla B)(\nabla)^\top \exp(-B)]_{\theta'=\theta} \\ \times (\theta' - \theta) + \dots,\end{aligned} \tag{3.15}$$

where

$$\begin{aligned}\theta &\equiv (\theta_1, \theta_2, \dots)^\top, \\ \nabla &\equiv \left(\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2}, \dots\right)^\top.\end{aligned} \tag{3.16}$$

From the definitions in Eqs. (3.13) and (3.14), it can be shown that the value of $B(\theta, \theta)$ is zero and $\nabla B(\theta, \theta')|_{\theta'=\theta}$ is a zero vector. Assuming Einstein summation convention from here onwards, we have for $\theta' = \theta + \epsilon u$

$$\rho(\theta, \theta + \epsilon u) = 1 - \frac{\epsilon^2}{2} u_\mu [(\nabla\nabla^\top B)_{\theta'=\theta}]_{\mu\nu} u_\nu + o(\epsilon^2) \tag{3.17}$$

where ϵ is a scalar, $u = (u_1, u_2, \dots)^\top$ is any real vector with the same dimension with θ, and $o(\epsilon^2)$ denotes terms asymptotically smaller than ϵ^2.

Similarly, it is also known that [13, 37]

$$\int \mathrm{d}x \sqrt{P_X(x|\theta) P_X(x|\theta+\epsilon u)} = 1 - \frac{\epsilon^2}{8} u_\mu j_{\mu\nu}(P_X) u_\nu + o(\epsilon^2), \tag{3.18}$$

by expanding the left-hand side with Taylor's series and applying the definition of the Fisher information in Eq. (2.68). Applying Eqs. (3.17) and (3.18) to left-hand and right-hand sides of Eq. (3.14) respectively, and comparing the ϵ^2 terms on both sides, we obtain

$$j(P_X) = 4 \nabla \nabla^\top B(\theta, \theta')|_{\theta'=\theta}. \tag{3.19}$$

In the case of SPLOT assumption, the Bhattacharyya distance is known to be [2]

$$\lim_{T\to\infty} \frac{1}{T} B(\theta, \theta') = \frac{1}{2} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \ln \frac{S_X(\omega|\theta) + S_X(\omega|\theta')}{2\sqrt{S_X(\omega|\theta) S_X(\omega|\theta)}}, \tag{3.20}$$

where $S_X(\omega|\theta)$ is the PSD of signal $x(t)$ given θ. For a complex signal $y(t)$, Eq. (3.20) needs to be modified. This can be done by assuming that $y(t)$ is band-limited in $[-\pi b, \pi b]$ and considering a real signal $x(t)$ given by

$$x(t) \equiv y(t) \exp(i\omega_0 t) + y^*(t) \exp(-i\omega_0 t), \tag{3.21}$$

where ω_0 is a carrier frequency assumed to be greater than πb. We then have

$$S_X(\omega|\theta) = S_Y(\omega - \omega_0|\theta) + S_Y(-\omega - \omega_0|\theta). \tag{3.22}$$

Using this expression in Eq. (3.20) leads to the expression for the Bhattacharyya distance

$$\lim_{T\to\infty} \frac{1}{T} B(\theta, \theta') = \int_{-\pi b}^{\pi b} \frac{\mathrm{d}\omega}{2\pi} \ln \frac{S_Y(\omega|\theta) + S_Y(\omega|\theta')}{2\sqrt{S_Y(\omega|\theta) S_Y(\omega|\theta)}} \tag{3.23}$$

where $y(t)$ is a complex signal.

3.4.1 Fisher information

In Sec. 3.3, measurement record $y(t)$ in the linear Gaussian model is a realization of a stationary process given θ, so the Fisher information matrix for the estimation of θ from $Y = \{y(t) : -T/2 \leq t \leq T/2\}$ in the SPLOT case can be obtained by combining Eqs. (3.19) and (3.23):

$$j(P_Y) = 4T\nabla\nabla^\top \int_{-\pi b}^{\pi b} \frac{\mathrm{d}\omega}{2\pi} \ln \frac{S_Y(\omega|\theta) + S_Y(\omega|\theta')}{2\sqrt{S_Y(\omega|\theta)S_Y(\omega|\theta)}}\bigg|_{\theta'=\theta}, \tag{3.24}$$

where the PSD $S_Y(\omega|\theta)$ is given by Eq. (3.9). Although the preceding formalism is applicable to the estimation of any of the parameters, in the following we focus on the force noise power, A. The CR bound on the mean-square error Σ_A is

$$\Sigma_A \equiv \mathbb{E}\left\{ \left[\check{A}(Y) - A\right]^2 \big|\theta \right\} \geq j_A^{-1} \tag{3.25}$$

$$j_A = T \int_{-\pi b}^{\pi b} \frac{\mathrm{d}\omega}{2\pi} \frac{S^2(\omega)}{[AS(\omega) + R]^2}, \tag{3.26}$$

assuming SPLOT condition. This bound allows us to study the efficiency of the estimators introduced in the next section.

3.5 Parameter Estimation Algorithms

3.5.1 Averaging

We first consider the estimator used in Refs. [21, 22]

$$\check{A}_{\mathrm{avg}} = G \int_{-T/2}^{T/2} \mathrm{d}t |y(t)|^2, \tag{3.27}$$

$$G = \left[T \int_{-\pi b}^{\pi b} \frac{\mathrm{d}\omega}{2\pi} S(\omega) \right]^{-1}. \tag{3.28}$$

The rationale for this simple averaging estimator is that, in the absence of measurement noise, $R = 0$, it is an unbiased estimate for $T \to \infty$:

$$\lim_{T\to\infty} \mathbb{E}_Y \left(\check{A}_{\text{avg}} | \theta, R = 0 \right) = A. \tag{3.29}$$

However, the unbiasedness breaks down in the presence of measurement noise, and we are therefore motivated to find a better estimator.

3.5.2 Radiometer

The radiometer estimator in Sec. 2.2.2 can be generalized for complex variables. The result is

$$\check{A}_{\text{rad}} = G \left[\int_{-T/2}^{T/2} \mathrm{d}t \int_{-T/2}^{T/2} \mathrm{d}t' y^*(t) h(t - t') y(t') - B \right], \tag{3.30}$$

where $y(t)$ is now complex, $h(t - t')$ is a filter function, G and B are parameters chosen to enforce the unbiased condition. For $T \to \infty$,

$$G = \left[T \int_{-\pi b}^{\pi b} \frac{\mathrm{d}\omega}{2\pi} H(\omega) S(\omega) \right]^{-1}, \tag{3.31}$$

$$B = T \int_{-\pi b}^{\pi b} \frac{\mathrm{d}\omega}{2\pi} H(\omega) R, \tag{3.32}$$

$$H(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t h(t) \exp(-i\omega t). \tag{3.33}$$

The averaging estimator $\check{A}_{\text{avg}}$ given by Eq. (3.27) is a special case of radiometer estimator.

The mean-square error has the asymptotic expression

$$\lim_{T\to\infty} \Sigma_A T = G^2 \int_{-\pi b}^{\pi b} \frac{\mathrm{d}\omega}{2\pi} H^2(\omega) S_Y^2(\omega|\theta). \tag{3.34}$$

This expression coincides with the CR bound given by Eqs. (3.25) and

(3.26) if we choose

$$H(\omega) = \frac{S(\omega)}{[DS(\omega) + R]^2} \tag{3.35}$$

and D happens to be equal to A. For any other value of D, the radiometer is suboptimal.

3.5.3 EM algorithm

In order to apply the iterative EM algorithm introduced in Sec. 2.2.2 to our model in Sec. 3.3, we consider a complex discrete-time Gauss-Markov model

$$z_{j+1} = f z_j + w_j, \tag{3.36}$$

$$y_j = c z_j + v_j, \qquad j = 0, 1, \ldots, J. \tag{3.37}$$

In general, z_j and y_j can be column vectors, and f and c are matrices. w_j and v_j are complex independent zero-mean Gaussian random variables with covariances given by

$$\mathbb{E}(w_j w_k^\dagger) = q\delta_{jk}, \qquad \mathbb{E}(w_j w_k^\top) = 0, \tag{3.38}$$

$$\mathbb{E}(v_j v_k^\dagger) = r\delta_{jk}, \qquad \mathbb{E}(v_j v_k^\top) = 0, \tag{3.39}$$

where q and r are covariance matrices. The parameters of interest θ are the components of f, c, q and r. The EM algorithm for a real Gauss-Markov model in Refs. [2, 12, 38] is generalized for complex variables here. Treating

Z as the hidden data, we have

$$-\ln P(Y,Z|\theta) = \sum_{j=0}^{J-1}(z_{j+1}-fz_j)^\dagger q^{-1}(z_{j+1}-fz_j) + J\ln\det q + \sum_{j=0}^{J}(y_j-cz_j)^\dagger r^{-1}(y_j-cz_j) + (J+1)\ln\det r + \mathcal{A}, \tag{3.40}$$

where $\mathcal{A}$ does not depend on θ and is discarded. To compute the estimated log-likelihood function $Q(\theta,\theta^k)$ of Eq. (2.44), we need estimate

$$\check{z}_j^k \equiv \mathbb{E}(z_j|Y,\theta^k), \tag{3.41}$$

and we define

$$\epsilon \equiv z_j - \check{z}_j^k, \tag{3.42}$$

$$\Pi_j^k \equiv \mathbb{E}(\epsilon_j^k \epsilon_j^{k\dagger}|Y,\theta^k), \tag{3.43}$$

$$\Pi_{j,j-1}^k \equiv \mathbb{E}(\epsilon_j^k \epsilon_{j-1}^{k\dagger}|Y,\theta^k). \tag{3.44}$$

Using the Rauch-Tung-Striebel (RTS) smoother [2, 38], we can compute these; starting with stationary initial conditions for $\check{z}_{-1}^{+k}$ and Π_{-1}^{+k}, the smoother consists of a forward Kalman filter:-

$$\check{z}_j^{-k} = f^k \check{z}_{j-1}^{+k}, \tag{3.45}$$

$$\Pi_j^{-k} = f^k \Pi_{j-1}^{+k} f^{k\dagger} + q, \tag{3.46}$$

$$K_j^{+k} = \Pi_j^{-k} c^{k\dagger}(c^k \Pi_j^{-k} c^{k\dagger} + r^k)^{-1}, \tag{3.47}$$

$$\check{z}_j^{+k} = \check{z}_j^{-k} + K_j^{+k}(y_j - c^k \check{z}_j^{-k}), \tag{3.48}$$

$$\Pi_j^{+k} = (I - K_j^{+k} c^k)\Pi_j^{-k}(I - K_j^{+k} c^k)^\dagger + K_j^{+k} r^k K_j^{+k\dagger}, \tag{3.49}$$

until $j = J$, and a backward propagation:-

$$\check{z}_J^k = \check{z}_J^{+k}, \tag{3.50}$$

$$\Pi_J^k = \Pi_J^{+k}, \tag{3.51}$$

$$K_j^k = \Pi_j^{+k} f^{k\dagger} (\Pi_{j+1}^{-k})^{-1}, \tag{3.52}$$

$$\check{z}_j^k = \check{z}_j^{+k} + K_j^k (\check{z}_{j+1}^k - \check{z}_{j+1}^{-k}), \tag{3.53}$$

$$\Pi_j^k = \Pi_j^{+k} - K_j^k (\Pi_{j+1}^{-k} - \Pi_{j+1}^k) K_j^{k\dagger}, \tag{3.54}$$

$$\Pi_{j,j-1}^k = \Pi_j^k K_{j-1}^{k\dagger}, \tag{3.55}$$

until $j = 0$. Applying Eqs. (3.41)–(3.44) into Eq. (3.40), we can then write $Q(\theta, \theta^k)$ as

$$\begin{aligned} -Q(\theta, \theta^k) = \mathrm{Tr}[q^{-1}(\Phi^k - f\Psi^{k\dagger} - \Psi^k f^\dagger + f\Theta^k f^\dagger) + J \ln q \\ + r^{-1}(\Upsilon - c\Xi^{k\dagger} - \Xi^k c^\dagger + c\Delta^k c^\dagger) + (J+1) \ln r], \end{aligned} \tag{3.56}$$

where Tr is the trace of a matrix,

$$\Phi^k \equiv \sum_{j=1}^{J} (\check{z}_j^k \check{z}_j^{k\dagger} + \Pi_j^k), \tag{3.57}$$

$$\Psi^k \equiv \sum_{j=1}^{J} (\check{z}_j^k \check{z}_{j-1}^{k\dagger} + \Pi_{j,j-1}^k), \tag{3.58}$$

$$\Theta^k \equiv \sum_{j=0}^{J-1} (\check{z}_j^k \check{z}_j^{k\dagger} + \Pi_j^k), \tag{3.59}$$

$$\Upsilon \equiv \sum_{j=0}^{J} y_j y_j^\dagger, \tag{3.60}$$

$$\Xi^k \equiv \sum_{j=0}^{J} y_j \check{z}_j^{k\dagger}, \tag{3.61}$$

$$\Delta^k \equiv \sum_{j=0}^{J} (\check{z}_j^k \check{z}_j^{k\dagger} + \Pi_j^k). \tag{3.62}$$

Maximizing $Q(\theta, \theta^k)$ given by Eq. (3.56) with respect to θ, we find new estimates

$$f^{k+1} = \Psi^k (\Theta^k)^{-1}, \tag{3.63}$$

$$c^{k+1} = \Xi^k (\Delta^k)^{-1}, \tag{3.64}$$

$$q^{k+1} = \frac{1}{J}[\Phi^k - \Psi^k (\Theta^k)^{-1} \Psi^{k\dagger}], \tag{3.65}$$

$$r^{k+1} = \frac{1}{J+1}[\Upsilon - \Xi^k (\Delta^k)^{-1} \Xi^{k\dagger}]. \tag{3.66}$$

The iteration process is repeated till a prescribed threshold is reached. The complex EM algorithm turns out to be the same as the real version with all transpose operations $^\top$ replaced by conjugate transpose †.

The problem may become ill-conditioned if there are too many unknown parameters and multiple ML solutions exist [2, 12, 39], so we choose a parameterization with known q:

$$f = \exp[(i\Omega - \omega)\delta t], \tag{3.67}$$

$$c = C\sqrt{A\frac{1 - \exp(-2\gamma\delta t)}{2\gamma\delta t}}, \tag{3.68}$$

$$q = \delta t, \tag{3.69}$$

$$r = \frac{R}{\delta t}, \tag{3.70}$$

where δt is the sampling period. With the EM estimates $\check{f}_{\mathrm{EM}}$, $\check{c}_{\mathrm{EM}}$, and $\check{r}_{\mathrm{EM}}$ and assuming that δt and C are known by independent calibrations, we can retrieve estimates of Ω, γ, A, and R:

$$\check{\Omega}_{\mathrm{EM}} = \frac{\arg \check{f}_{\mathrm{EM}}}{\delta t}, \tag{3.71}$$

$$\check{\gamma}_{\mathrm{EM}} = -\frac{\ln |\check{f}_{\mathrm{EM}}|}{\delta t}, \tag{3.72}$$

$$\check{A}_{\mathrm{EM}} = \frac{\check{c}^2_{\mathrm{EM}}}{C^2} \frac{2\check{\gamma}_{\mathrm{EM}}\delta t}{1 - \exp(-2\check{\gamma}_{\mathrm{EM}}\delta t)}, \tag{3.73}$$

$$\check{R}_{\mathrm{EM}} = \check{r}_{\mathrm{EM}}\delta t. \tag{3.74}$$

The same algorithm is also applicable to the quantum Gauss-Markov model [9], as the RTS smoother is equivalent to the linear quantum smoother [40, 41, 42]. The possibility of using the EM algorithm for quantum systems is also mentioned in Ref. [43].

3.6 Experimental Data Analysis

3.6.1 Procedure

There are two records of experimental data, one with thermal noise in $\xi(t)$ and one with additional applied white noise in $\xi(t)$, leading to a different A for each record, denoted by $A^{(0)}$ and $A^{(1)}$. Each record contains $J_{\max}+1 = 3,750,001$ points of $y_j^{(n)}$. With a sampling frequency $b = 1/\delta t = 15$ MHz, the total time for each record is $T_{\max} = J_{\max}\delta t = 0.25$ s. From independent calibrations, we also obtain $C = 2.61 \times 10^{-2}$ $(\mathrm{fN}/\sqrt{\mathrm{Hz}})^{-1}$.

To investigate the errors with varying T, we divide each record into slices of records with various T, resulting in $M(T) = \mathrm{floor}(T_{\max}/T)$ number of trials for each T. Using a desktop computer (Intel Core i7-2600 CPU@3.4GHz with 16GB RAM) and Matlab® we apply each of the three estimators in Sec. 3.5 to each trial to produce an estimate $\check{A}_{m,l}^{(n)}(T)$, where m denotes the trial and l denotes the estimator. The EM iteration is stopped when the fractional difference between the current estimate of A and the previous value is less than 10^{-7}. For the averaging and radiometer estimators, true values for Ω, γ, and R are needed, and since we do not know them, we estimate them by applying the EM algorithm to the whole records. This is reasonable because $T_{\max} \gg 4$ ms $\geq T$, and we expect $\check{\theta}_{\mathrm{EM}}^{(n)}(T_{\max})$ to be much closer to the true values θ^n than the short-time estimates. The EM algorithm for each T, on the other hand, does not use $\check{\theta}_{\mathrm{EM}}^{(n)}(T_{\max})$ at all and produces its own estimates each time. The parameter D in Eq. (3.35) of radiometer estimator is taken to be $\check{A}_{\mathrm{EM}}^{(0)}(T_{\max})$.

The estimation errors are computed by

$$\Sigma_l^{(n)}(T) = \frac{1}{M(T)} \sum_{m=1}^{M(T)} \left[\check{A}_{m,l}^{(n)}(T) - A^{(n)} \right]^2, \tag{3.75}$$

and compared with the SPLOT CR bound j_A^{-1} of Eq. (3.26) by assuming $\theta^{(n)} = \check{\theta}_{\mathrm{EM}}^{(n)}(T_{\max})$. Note that the estimation error in general contains two components:

$$\Sigma = \frac{1}{M} \sum_{m=1}^{M} (\check{A}_m - \bar{A})^2 + (\bar{A} - A)^2, \tag{3.76}$$

where

$$\bar{A} \equiv \frac{1}{M} \sum_{m=1}^{M} \check{A}_m \tag{3.77}$$

is the sample mean of the estimate. The first component is the sample variance, and the second component is the square of the estimate bias with respect to the true value A. Unlike Refs. [21, 22], our error analysis is able to account for the bias component more accurately by referencing with the much more accurate EM estimates.

3.6.2 Results

Applied to the two records, the EM algorithm produces the following estimates listed in Table 3.1. The algorithm takes ≈ 3.3 hours to run for each record. These values are then used as references to analyze the estimators at shorter times.

Figure 3.2 plots the root-mean-square errors $\sqrt{\Sigma_l^{(n)}(T)}$ and the SPLOT CR bound $j_A^{-1/2}$ versus time T in log-log scale. The two plots show very similar behavior. A few observations can be made:

1. The averaging estimator is more accurate than the radiometer for

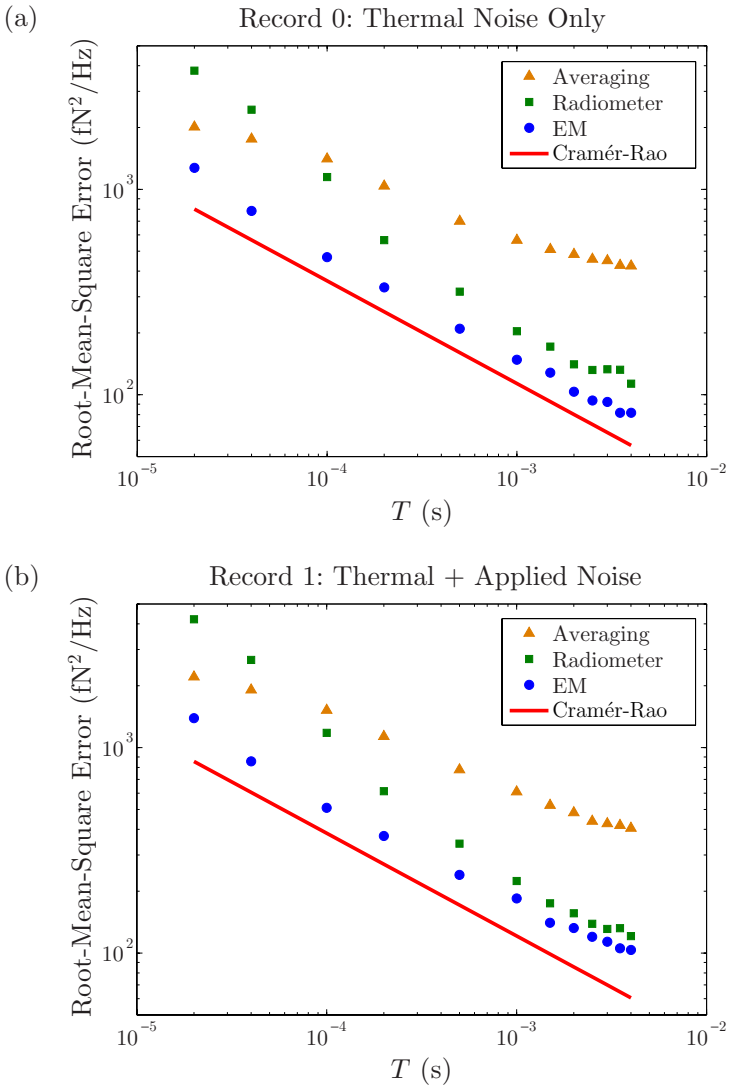


Figure 3.2: Root-mean-square force-noise-power estimation errors and the asymptotic CR bound versus time in log-log scale. (a) The force contains thermal noise only. (b) The force contains thermal noise and an applied noise.

Table 3.1: Estimates of θ taken to be true values of θ, by applying the EM algorithm to the whole records ($T_{\max}$).

Estimates	Values
$\hat{A}^{(0)}_{\mathrm{EM}}(T_{\max}) = 2.4748/C^2$	3.64×10^3 fN^2 Hz^{-1}
$\hat{\Omega}^{(0)}_{\mathrm{EM}}(T_{\max})$	-1.8582×10^4 rad s^{-1}
$\hat{\gamma}^{(0)}_{\mathrm{EM}}(T_{\max})$	5.5730×10^4 rad s^{-1}
$\hat{R}^{(0)}_{\mathrm{EM}}(T_{\max})$	1.4532×10^{-13} Hz^{-1}
$\hat{A}^{(1)}_{\mathrm{EM}}(T_{\max}) = 2.6926/C^2$	3.96×10^3 fN^2 Hz^{-1}
$\hat{\Omega}^{(1)}_{\mathrm{EM}}(T_{\max})$	-1.8668×10^4 rad s^{-1}
$\hat{\gamma}^{(1)}_{\mathrm{EM}}(T_{\max})$	5.6156×10^4 rad s^{-1}
$\hat{R}^{(1)}_{\mathrm{EM}}(T_{\max})$	1.4703×10^{-13} Hz^{-1}

short times but becomes much worse for longer times. We cannot explain the short-time errors because our analytic results rely on the long-time limit. The large long-time errors can be attributed to the bias and suboptimality of the averaging estimator.

2. The radiometer beats the averaging estimator and approaches the CR bound for longer times. This is consistent with our SPLOT analysis, as we have chosen $D = \check{A}^{(0)}_{\mathrm{EM}}(T_{\max})$ and the radiometer should be near-optimal.

3. The EM estimator beats the other estimators at all times and follows the CR bound more closely, even though we allow the averaging and radiometer estimators to have the unfair advantage of accessing more accurate values of Ω, γ, and R. This may be explained by the fact that the EM algorithm is formulated to perform ML estimation on discrete measurements for any finite T, unlike the other estimators that rely only on asymptotic arguments.

4. The EM estimator takes a much longer time to compute (computation time ≈ 200 s for one trial with $J + 1 = 60,001$ points and $T = 4$ ms) than the other estimators (≈ 0.3 ms for the averaging estimator, ≈ 16 ms for the radiometer). If computation time is a concern, the radiometer estimator may be preferred, though its performance

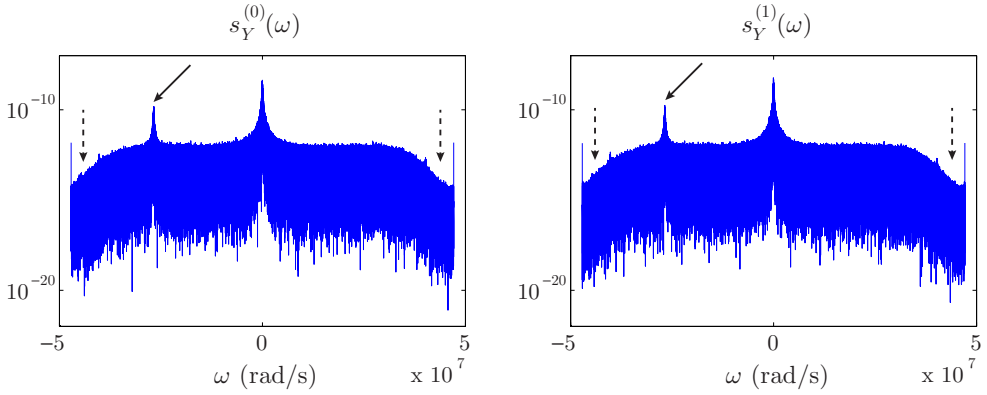


Figure 3.3: Raw spectra $s_Y^{(n)}(\omega)$ of the measurement records $y_j^{(n)}$ in log scale. In both records, there exists a second weaker resonance peak denoted by solid arrow. At higher frequencies region, the flat noise floor of the data rolls off (dotted arrow). (Left) Record 0: Thermal noise only. (Right) Record 1: Thermal + applied noise.

depends heavily on the accuracy of the other assumed parameters, and the EM method can still be useful for estimating such parameters in offline system identification.

To gain further insight into the finite gap between the errors and the CR bound, in Fig. 3.3 we plot the raw spectrum of $y_j^{(n)}$, defined as

$$s_Y^{(n)}(\omega) \equiv \frac{1}{T_{\max}} \left| \delta t \sum_{j=0}^{J_{\max}} y_j^{(n)} \exp(i\omega j \delta t) \right|^2 . \tag{3.78}$$

The figures show that for both records our model does not exactly match the experiment in two ways:

1. The data shows a second weaker resonance peak.

2. The noise floor of the data rolls off at higher frequencies due to the presence of an RF notch filter in the experiment prior to data acquisition.

Despite the mismatch, our results are in reasonable agreement with the theory. To further improve the estimation accuracy, the weaker resonance can be modeled by including another mode in our linear Gaussian model,

while the noise-floor roll-off may be removed by a whitening filter before applying the estimators.

3.7 Outlook

We investigate parameter estimation for an optomechanical system, focusing on unbiased and ML estimators and the CR bound. For detection applications [44] with uncertain parameters, the ML estimator can form the basis of more advanced hypothesis testing techniques, such as the generalized likelihood-ratio test [2]. The assumption of static parameters means that the presented techniques are most suited to system identification purposes. For sensing applications, the parameters are often time-varying, and Bayesian estimators, such as the extended and unscented Kalman filters for continuous variables [38], the generalized-pseudo-Bayesian and interacting-multiple-model algorithms for finite-state dynamical hypotheses [45], and particle filtering [46], may be more suitable.

Since the Gauss-Markov model often remains valid for quantum systems [9], a quantum extension of our study is straightforward. This means that the presented techniques are potentially useful for future quantum sensing and system identification applications, such as optomechanical force sensing [47, 48, 49, 50] and atomic magnetometry [40, 51]. We expect our parametric methods to lead to more accurate quantum sensing and control than robust quantum control methods [51, 52], which may be too conservative for the highly controlled environment of typical quantum experiments. There also exist quantum versions of the CR bound that impose fundamental limits to the parameter estimation accuracy for a quantum system with any measurement [10, 53, 54], and it may be interesting to explore how close the classical bounds presented here can get to the quantum limits.

The continued improvement of optomechanical devices for applications

and fundamental science requires precise engineering of the mechanical resonance frequency, dissipation rate and effective mass. This necessitates a deep understanding of how these mechanical properties depend on various materials and fabrication techniques. The mechanical resonance frequency is easily predicted via a numerical eigenmode analysis using the geometry of the structure and the Young's modulus of the material. However, it is much more challenging to predict the level of mechanical dissipation, where numerical models are not as well established and multiple decay channels usually exist. Effective experimental characterization of such dissipation channels requires high precision force estimation to accurately quantify the oscillators coupling to the environment. This is critical to advancing optomechanics in applications such as quantum memories and quantum information [55, 56]. A more immediate application for high precision force estimation is that of temperature sensing and bolometry where small relative changes of the signal power are of interest, for example, in detecting submillimeter wavelengths in radio astronomy [57] or even to search for low energy events in particle physics [58]. Given the demonstrated success of our statistical techniques, we envision them to be similarly useful for all these applications.

Chapter 4

Estimation of Spectral Parameters with Quantum Dynamical Systems

In this chapter, we propose a theoretical framework for estimation of spectral parameters with quantum dynamical systems, proving the fundamental limits to estimation errors and investigating measurement and data analysis techniques that approach the limits. The key result is a simple analytic expression of the limits in terms of basic PSDs in the problem, such that they can be readily applied to optics and optomechanics [3] experiments. To illustrate our theory, we analyze a recent experiment of continuous optical phase estimation and demonstrate that the experimental performance using homodyne detection is close to our quantum limit. We further propose a spectral photon counting method that can beat homodyne detection and attain quantum-optimal performance for weak modulation and a coherent-state input. The advantage is especially significant when the signal-to-noise ratio (SNR) is low, thus demonstrating the importance of quantum-optimal measurements and coherent optical information processing in the low-SNR regime. The results are reported in Ref. [59].

4.1 Motivation

Recent spectacular development of theory and experiment, especially in optomechanics, suggests that quantum noise will soon be the major limiting factor in many metrological applications [60]. Many tasks in optomechanics force sensing, including thermometry, estimation of stochastic gravitational-wave background [61], and testing-spontaneous-wavefunction-collapse [62, 63], involve the spectrum analysis of a stochastic force. The effect of quantum noise on such tasks has also been of recent interest [62, 63]. Therefore, to study the quantitative effect of experimental design on estimation accuracy, it is important to use a rigorous statistical inference framework to investigate the parameter estimation error. While there exist many theoretical studies of quantum parameter estimation for thermometry (see, for example, Refs. [64, 65, 66, 67, 68]), their application to more complex dynamical systems with broadband measurements such as optomechanics remains unclear.

4.2 Quantum Metrology

Consider a quantum dynamical system with Hamiltonian $\hat{H}(X,t)$ as a functional of a c-number hidden stochastic process $X(t)$, such as a classical force. Also, assume that the prior probability measure of $X(t)$ depends on an unknown vectorial parameter θ.

4.2.1 Review of parameter estimation theory

Let Y be the quantum measurement outcome and $\check{\theta}(Y)$ be an estimate of θ given Y. The mean-square estimation error matrix Σ is defined in Eq. (2.57),

$$\Sigma_{\mu\nu} \equiv \mathbb{E}_Y\{[\check{\theta}_\mu(Y) - \theta_\mu][\check{\theta}_\nu(Y) - \theta_\nu]\}. \tag{4.1}$$

Given a quantum system, let $\hat{\rho} = \hat{\rho}(\theta)$ be a θ-dependent density operator and $\hat{E}(y)$ be the POVM that models the measurement, such that the observation probability measure

$$P_Y(y|\theta) = \mathrm{tr}[\hat{E}(y)\hat{\rho}(\theta)]. \tag{4.2}$$

For dynamical systems, $\hat{\rho}(\theta)$ can be obtained using the principles of purification and deferred measurements [5, 53, 54, 69]. For the purpose of spectrum-parameter estimation, we model $\hat{\rho}$ as

$$\hat{\rho}(\theta) = \mathbb{E}_{X|\theta}\left\{\hat{U}[X,T]|\psi\rangle\langle\psi|\hat{U}^\dagger[X,T]\right\}, \tag{4.3}$$

where $\hat{U}[X,T] = \mathcal{T}\exp\{-(i/\hbar)\int_0^T dt\hat{H}[X,t]\}$ is the unitary time-ordered exponential of $\hat{H}$ with total evolution time T, $|\psi\rangle$ is the initial quantum state, and the expectation is with respect to the hidden process $X(t)$, the prior probability measure of which depends on θ. θ is called hyperparameter in this context [70].

In Sec. 2.2.5, we see that for any POVM, a QCR bound states that

$$\Sigma \geq j^{-1}(P_Y) \geq J^{-1}(\hat{\rho}), \tag{4.4}$$

where $j(P_Y)$ is the classical Fisher information matrix with respect to P_Y and $J(\hat{\rho})$ is the QFI matrix with respect to the SLD of $\hat{\rho}$ [10, 15, 71].

4.2.2 Extended convexity of QCR bound

While quantum parameter estimation bounds for dynamical systems have been studied previously in the context of low-dimensional systems such as qubits (see, for example, Refs. [72, 73, 74]), J is much more difficult to evaluate analytically for multimode high-dimensional dynamical systems under continuous measurements. To proceed, we will exploit a recently

discovered property of J known as the extended convexity [75], which states that

$$J(\hat{\rho}) \leq \mathcal{J}\{\hat{\sigma}, P_Z\} \equiv \mathbb{E}_{Z|\theta}\left[J(\hat{\sigma})\right] + j(P_Z), \tag{4.5}$$

where $\{\hat{\sigma}, P_Z\}$ is any ensemble of $\hat{\rho}$ with elements $\hat{\sigma}$ and mixing probability measure P_Z such that $\hat{\rho}(\theta) = \mathbb{E}_{Z|\theta}[\hat{\sigma}(Z|\theta)]$. Ref. [75] proves $J \leq \mathcal{J}$ for one parameter but assumes a POVM attaining $j = J$ that may not exist [76].

To prove $J \leq \mathcal{J}$, here we use instead the strong concavity of Uhlmann fidelity [5]. Let $\{\hat{\sigma}, P_Z\}$ be an ensemble for $\hat{\rho}(\theta)$ such that

$$\hat{\rho}(\theta) = \int \mathrm{d}z P_Z(z|\theta)\hat{\sigma}(z|\theta). \tag{4.6}$$

Define the *Uhlmann fidelity* as

$$F[\hat{\rho}, \hat{\rho}'] \equiv \mathrm{tr}[\sqrt{\sqrt{\hat{\rho}}\hat{\rho}'\sqrt{\hat{\rho}}}]. \tag{4.7}$$

The strong concavity states that [5]

$$F[\hat{\rho}(\theta), \hat{\rho}(\theta')] \geq \int \mathrm{d}z \sqrt{P_Z(z|\theta)P_Z(z|\theta')} \times F[\hat{\sigma}(z|\theta), \hat{\sigma}(z|\theta')]. \tag{4.8}$$

Ref. [5] proves this property for a discrete z, but the proof also applies to continuous z if we purify P_Z in the form $\int \mathrm{d}z\sqrt{P_Z(z)}|z\rangle$ with continuous eigenket $|z\rangle$ that satisfies $\langle z|z'\rangle = \delta(z - z')$. To relate F to J, we use the fact [18]

$$F[\hat{\rho}(\theta), \hat{\rho}(\theta + \epsilon u)] = 1 - \frac{\epsilon^2}{8} u_\mu J_{\mu\nu}(\hat{\rho}) u_\nu + o(\epsilon^2), \tag{4.9}$$

where ϵ is a scalar, u is any real vector with the same dimension as θ, and $o(\epsilon^2)$ denotes terms asymptotically smaller than ϵ^2. We have derived earlier

in Eq. (3.18), relating Bhattacharyya coefficient with j,

$$\int \mathrm{d}z \sqrt{P_Z(z|\theta) P_Z(z|\theta + \epsilon u)} = 1 - \frac{\epsilon^2}{8} u_\mu j_{\mu\nu}(P_Z) u_\nu + o(\epsilon^2). \tag{4.10}$$

Expanding $F[\hat{\rho}(\theta), \hat{\rho}(\theta')]$ and $F[\hat{\sigma}(z|\theta), \hat{\sigma}(z|\theta')]$ in Eq. (4.8) using Eq. (4.9), applying Eq. (4.10) to the right-hand side of Eq. (4.8), and comparing the ϵ^2 terms on both sides, we obtain

$$u_\mu J_{\mu\nu}(\hat{\rho}) u_\nu \leq u_\mu \{\mathbb{E}_{Z|\theta}[J_{\mu\nu}(\hat{\sigma})] + j_{\mu\nu}(P_Z)\} u_\nu. \tag{4.11}$$

Since this holds for any u, we obtain the matrix inequality in Eq. (4.5). The classical simulation technique proposed in Ref. [77] can be regarded as a special case of extended convexity when $J(\hat{\sigma}) = 0$.

4.2.3 Tighter bounds

To compute simple analytic results for dynamical systems, we make further assumptions. Assume that $X(t)$ is zero-mean, Gaussian, and stationary, with a PSD given by

$$S_X(\omega|\theta) \equiv \int_{-\infty}^{\infty} \mathrm{d}\tau \mathbb{E}_{X|\theta}[X(t)X(t+\tau)] \exp(i\omega\tau). \tag{4.12}$$

And for the quantum system, we assume Hamiltonian of the form

$$\hat{H} = \hat{H}_0 - \hat{Q} X(t), \tag{4.13}$$

where $\hat{Q}$ is the quantum generator and $\hat{H}_0$ is the rest of the Hamiltonian. For example, in the context of optical phase modulation (see Sec. 4.3), $X(t)$ is the phase modulation on the optical beam and $\hat{Q}$ is proportional to the photon-flux operator. For mechanical force measurements, $X(t)$ can be the classical force on a mechanical oscillator and $\hat{Q}$ is the quantum position

operator. In the case of linear optomechanical system [3], a modified purification technique can transform the Hamiltonian in the interaction picture and produce a possible tighter bound in terms of the optical statistical alone [54, 59].

As the extended convexity holds for any ensemble of $\hat{\rho}$, tighter bounds can be obtained by choosing the ensemble meticulously. To transform the original ensemble $\{\hat{\sigma}_x, P_X\}$ into a new ensemble $\{\hat{\sigma}_z, P_Z\}$, we define a new stochastic process $Z(t)$ by

$$X(t) = \int_{-\infty}^{\infty} \mathrm{d}\tau g(t-\tau|\theta) Z(\tau), \tag{4.14}$$

where g is an impulse-response function to be chosen later. $\hat{\rho}$ can now be expressed as

$$\hat{\rho}(\theta) = \mathbb{E}_{Z|\theta}\left\{\hat{U}[g*Z,T]|\psi\rangle\langle\psi|\hat{U}^{\dagger}[g*Z,T]\right\}, \tag{4.15}$$

where $*$ denotes convolution. We choose

$$\hat{\sigma}_z = \hat{U}[g*Z,T]|\psi\rangle\langle\psi|\hat{U}^{\dagger}[g*Z,T], \tag{4.16}$$

and we have an ensemble $\{\hat{\sigma}_z, P_Z\}$ for $\hat{\rho}$, which is parameterized by g.

Assuming the Hamiltonian in Eq. (4.13), it can be shown that [53, 78]

$$\begin{aligned} J_{\mu\nu} = \frac{4}{\hbar^2}\int_0^T \mathrm{d}t \int_0^T \mathrm{d}t' K_Q(t,t') \int_{-\infty}^{\infty} \mathrm{d}\tau \frac{\partial g(t-\tau|\theta)}{\partial \theta_\mu} Z(\tau) \\ \times \int_{-\infty}^{\infty} \mathrm{d}\tau' \frac{\partial g(t'-\tau'|\theta)}{\partial \theta_\nu} Z(\tau'), \end{aligned} \tag{4.17}$$

where $K_Q(t,t')$ is the quantum covariance of the generator in the Heisenberg picture, defined as

$$K_Q(t,t') \equiv \mathrm{Re}\left[\langle\psi|\Delta\hat{Q}(t)\Delta\hat{Q}(t')|\psi\rangle\right], \tag{4.18}$$

$$\Delta\hat{Q}(t) \equiv \hat{Q}(t) - \langle\psi|\hat{Q}(t)|\psi\rangle, \tag{4.19}$$

$$\hat{Q}(t) \equiv \hat{U}^{\dagger}(X,t)\hat{Q}\hat{U}(X,t). \tag{4.20}$$

We now assume that $K_Q(t,t')$ is independent of $X(t)$; such an assumption is commonly satisfied in optical-phase-modulation and linear optomechanical systems. The expected $J(\sigma)$ becomes

$$\begin{aligned}\mathbb{E}_{Z|\theta}[J_{\mu\nu}(\hat{\sigma}_z)] = \frac{4}{\hbar^2}\int_0^T \mathrm{d}t \int_0^T \mathrm{d}t' K_Q(t,t') \int_{-\infty}^{\infty} \mathrm{d}\tau \int_{-\infty}^{\infty} \mathrm{d}\tau' K_Z(\tau,\tau'|\theta) \\ \times \left[\frac{\partial g(t-\tau|\theta)}{\partial\theta_\mu}\right]\left[\frac{\partial g(t'-\tau'|\theta)}{\partial\theta_\nu}\right]\end{aligned} \tag{4.21}$$

where

$$K_Z(\tau,\tau'|\theta) \equiv \mathbb{E}_{Z|\theta}[Z(\tau)Z(\tau')] \tag{4.22}$$

is the prior covariance of $Z(t)$. Assume further that the quantum statistics of $\Delta Q(t)$ are stationary, with a PSD given by

$$S_Q(\omega) \equiv \int_{-\infty}^{\infty} \mathrm{d}\tau K_Q(t,t+\tau)\exp(i\omega\tau). \tag{4.23}$$

Defining a transfer function as

$$G(\omega|\theta) \equiv \int_{-\infty}^{\infty} \mathrm{d}t g(t|\theta)\exp(i\omega t). \tag{4.24}$$

restricting G to be nonzero for all frequencies of interest. Note that the PSD of $Z(t)$ is $S_X/|G|^2$, making the SPLOT assumption, Eq. (4.21) can be rewritten as

$$\mathbb{E}_{Z|\theta}[J_{\mu\nu}(\hat{\sigma}_z)] = T\int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi}\frac{4S_Q S_X}{\hbar^2}\frac{\partial \ln G}{\partial\theta_\mu}\frac{\partial \ln G}{\partial\theta_\nu} \tag{4.25}$$

The Fisher information $j(P_Z)$ can be obtained by substituting Eq. (3.20)

into Eq. (3.19). The result is

$$j_{\mu\nu}(P_Z) = \frac{T}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{1}{S_Z^2}\frac{\partial S_Z}{\partial\theta_\mu}\frac{\partial S_Z}{\partial\theta_\nu} \tag{4.26}$$

$$= T\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{1}{2}\frac{\partial}{\partial\theta_\mu}\left(\ln\frac{S_X}{|G|^2}\right)\frac{\partial}{\partial\theta_\nu}\left(\ln\frac{S_X}{|G|^2}\right). \tag{4.27}$$

Combining Eqs. (4.25) and (4.27) according to Eq. (4.5), we obtain

$$u_\mu\mathcal{J}_{\mu\nu}u_\nu = T\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\left[\frac{4S_QS_X}{\hbar^2}|\lambda|^2 + \frac{1}{2}(\Lambda-\lambda-\lambda^*)^2\right], \tag{4.28}$$

where

$$\lambda \equiv u_\mu\frac{\partial\ln G}{\partial\theta_\mu}, \tag{4.29}$$

$$\Lambda \equiv u_\mu\frac{\partial\ln S_X}{\partial\theta_\mu}. \tag{4.30}$$

Since Eq. (4.28) is quadratic with respect to λ, the λ and thus G that minimizes Eq. (4.28) for each u can be found analytically by taking its derivatives with respect to the real and imaginary parts of λ and setting them to zero. This gives

$$\lambda = \frac{\Lambda}{2+4S_QS_X/\hbar^2}, \tag{4.31}$$

and a variational upper bound on the QFI matrix,

$$J \le \tilde{\mathcal{J}}, \tag{4.32}$$

which elements are given by

$$\tilde{\mathcal{J}}_{\mu\nu} = T\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{(\partial\ln S_X/\partial\theta_\mu)(\partial\ln S_X/\partial\theta_\nu)}{2+\hbar^2/S_QS_X}. \tag{4.33}$$

Note that the quantum state $|\psi\rangle$ need not be Gaussian for the result to hold.

For continuous optical phase modulation [9, 48, 49, 50], S_X is the phase PSD and $S_Q/\hbar^2$ is the photon-flux PSD. For mechanical force measurement, S_X is the classical force PSD and S_Q is the mechanical position PSD. For linear cavity optomechanics, using the interaction-picture purification technique, S_X is the PSD of the forced displacement and S_Q is proportional to the cavity photon number PSD [59]. The frequency-domain integral given by Eq. (4.33), together with the matrix inequalities

$$\Sigma \geq j^{-1} \geq J^{-1} \geq \tilde{\mathcal{J}}^{-1} \tag{4.34}$$

that follow from Eqs. (4.4) and (4.32), represent a novel form of uncertainty relations and indicate a nontrivial interplay between the classical noise characterized by S_X and a frequency-domain SNR given by $S_Q S_X/\hbar^2$ in bounding the estimation error and the Fisher information quantities.

Note also that $\tilde{\mathcal{J}}$ is proportional to the total time T, as are all the Fisher information quantities we derive here and in Chapter 3. This suggests that a longer observation time can improve the parameter estimation, as is well known in statistics [12] but missed by some of the previous quantum studies [62, 63].

4.3 Continuous Optical Phase Modulation

Consider the optics experiment depicted in Fig. 4.1(a) or (b). An external stochastic source $X(t)$, such as a moving mirror or an electro-optic modulator, modulates the phase of a continuous optical beam, which is then measured to obtain information about the source. The Hamiltonian is

$$\hat{H} = \hbar \hat{I}(t) X(t), \tag{4.35}$$

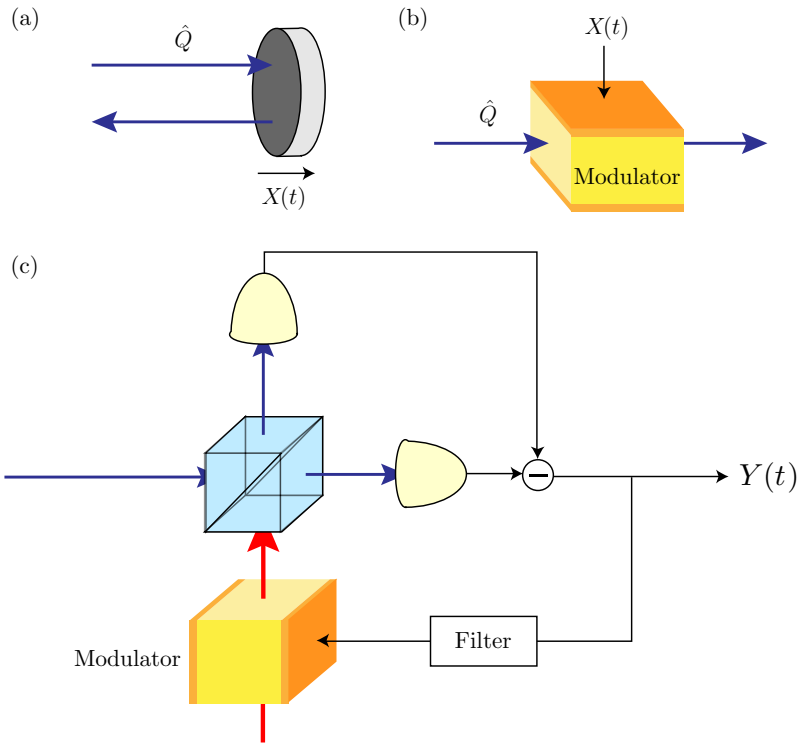


Figure 4.1: (a) and (b) Optical phase modulation, where $X(t)$ is the phase modulation and $\hat{Q}$ is proportional to photon flux. (c) Adaptive homodyne detection.

where $\hat{I}(t)$ is the photon-flux operator, $S_I(\omega) = S_Q(\omega)/\hbar^2$ is the photon-flux PSD, and $S_X(\omega|\theta)$ is the source PSD. The quantum limit given by Eq. (4.33) becomes

$$\tilde{\mathcal{J}}_{\mu\nu} = T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2 + 1/(S_I S_X)}, \tag{4.36}$$

where $\partial_\mu \equiv \partial/\partial\theta_\mu$.

4.3.1 Homodyne detection

We can compare our bound with the Fisher information for homodyne detection, a standard experimental phase measurement method [9, 48, 49, 50], as illustrated in Fig. 4.1(c). If the mean field is strong, and the modulation is weak or tight phase locking is achieved, the output process can be

linearized as

$$Y(t) \approx X(t) + \eta(t), \tag{4.37}$$

where $\eta(t)$ is the phase-quadrature noise. As discussed in Sec. 3.3, if η is Gaussian and stationary with PSD $S_\eta(\omega)$ such that Y is also Gaussian and stationary, the information $j[P_Y^{(\mathrm{hom})}]$ can be computed analytically [13]. Using Eq. (4.26), the result with the SPLOT assumption is

$$j_{\mu\nu}[P_Y^{(\mathrm{hom})}] = T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2(1 + S_\eta/S_X)^2}. \tag{4.38}$$

The classical CR bound $\Sigma \geq j^{-1}[P_Y^{(\mathrm{hom})}]$ is asymptotically attainable for long T using ML estimation [12].

With the quadrature uncertainty relation $S_\eta(\omega)S_I(\omega) \geq 1/4$ for the optical beam [79], the optimal homodyne information $\tilde{j}$ is

$$j[P_Y^{(\mathrm{hom})}] \leq \tilde{j}, \tag{4.39}$$

and

$$\tilde{j}_{\mu\nu} = T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2 + 1/(S_I S_X) + 1/(8S_I^2 S_X^2)}. \tag{4.40}$$

We can compare this homodyne limit with the quantum limit in Eq. (4.36); the expressions are similar, apart from an extra factor of $1/(8S_I^2 S_X^2)$ that makes the homodyne limit strictly worse than our quantum limit, especially if $S_I S_X$ is small.

4.3.2 Spectral photon counting

Although Eq. (4.34) sets rigorous lower bounds on the estimation error Σ, there is no guarantee the existence of a measurement that can attain

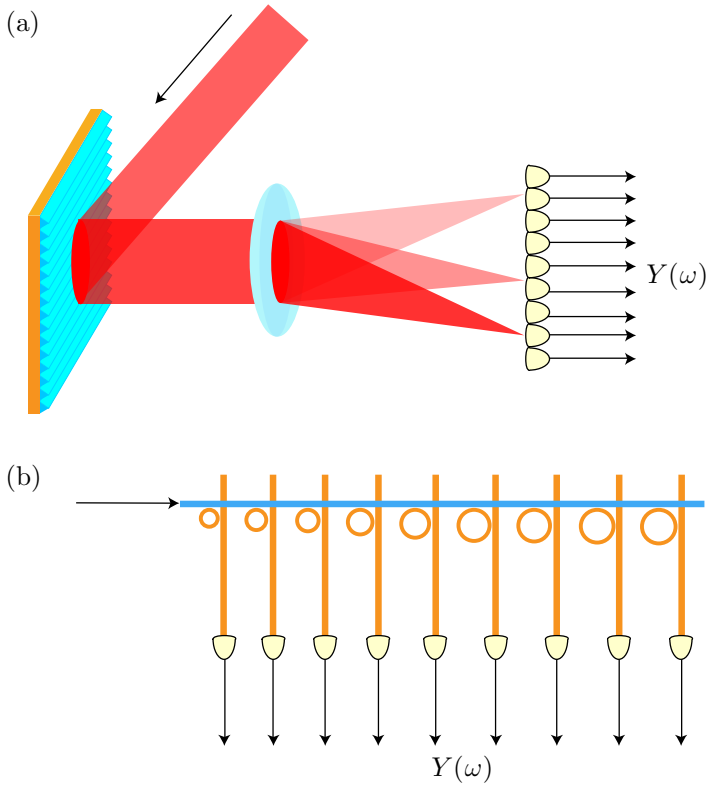


Figure 4.2: (a) Spectral photon counting with a diffraction grating and a lens. (b) Spectral photon counting with an optical ring-resonator array.

the final bound $\tilde{\mathcal{J}}^{-1}$. Inspired by Ref. [68], we propose a photon counting method which we name spectral photon counting. It consists of a conventional optical spectrometer with photon counting for each spectral mode [80, 81]. For the phase spectral parameter estimation problem with weak modulation and a coherent-state input, this method turns out to have an information $j[P_Y^{(\mathrm{spc})}]$ coincides with $\tilde{\mathcal{J}}$ for all parameters.

The first step of spectral photon counting is the coherent optical Fourier transform via a dispersive optical element, such as a diffraction grating or a prism and a Fourier-transform lens [80] as depicted in Fig. 4.2(a), or an array of optical ring resonators with different resonant frequencies coupled to a cross grid of waveguides [82] as depicted in Fig. 4.2(b). The second step is a measurement of the photon numbers in the spectral modes, and the final step is a ML estimation of θ from the spectral photon counting results.

Let the positive-frequency electric field at the input of the phase modulator be

$$\hat{E}^{(+)}(t) = \hat{A}(t)\exp(-i\Omega t), \tag{4.41}$$

where $\hat{A}(t)$ is an annihilation operator for the slowly varying envelope with commutation relation

$$[\hat{A}(t), \hat{A}^{\dagger}(t')] = \delta(t - t') \tag{4.42}$$

and Ω is the optical carrier frequency. With a strong mean field $\alpha \equiv \langle\psi|\hat{A}(t)|\psi\rangle$ and weak phase modulation, the output field can be linearized as

$$\hat{B}(t) \approx \hat{A}(t) + i\alpha X(t). \tag{4.43}$$

To model the optical Fourier transform, we follow Shapiro [80] to express each frequency mode in terms of the mode annihilation operator as

$$\hat{b}_m = (1/\sqrt{T})\int_0^T dt\hat{B}(t)\exp(i\omega_m t), \tag{4.44}$$

with sideband frequencies $\omega_m = 2\pi m/T$, $m \in \{\dots, -2, -1, 0, 1, 2, \dots\}$, and $[\hat{b}_m, \hat{b}_n^{\dagger}] = \delta_{mn}$. Assuming α to be time-constant,

$$\hat{b}_m \approx \hat{a}_m + i\alpha x_m, \tag{4.45}$$

where $\hat{a}_m$ is the Fourier transform of $\hat{A}(t)$ and x_m is that of $X(t)$ in the same way as $\hat{b}_m$.

The strong mean field is contained in the $m = 0$ mode only, and if the spectrum of x_m is wide, negligible information is lost if we neglect the $m = 0$ mode. The other modes are coherent states for a given displacement

$i\alpha x_m$ if the input beam is a coherent state [80]. For a given x_m, the photon-counting distribution for $\hat{n}_m \equiv \hat{b}_m^\dagger \hat{b}_m$ in each mode is therefore Poissonian with mean $|\alpha|^2|x_m|^2$ and independent from one another.

Since $X(t)$ is a hidden stochastic process, we must average the Poissonian distribution over the prior of $X(t)$ to obtain the final likelihood function. For a Gaussian $X(t)$ with the SPLOT assumption, $\{x_m; m > 0\}$ are independent complex Gaussian random variables with variances $S_X(\omega_m|\theta)$ [12], but since $X(t)$ is real, the sidebands are symmetric with $x_m = x_{-m}^*$. This means that, averaged over x, the photon numbers at opposite sideband frequencies become correlated.

To simplify the analysis, suppose that, for each $m > 0$, we sum the pair of measured photon numbers n_m and n_{-m} at opposite sidebands and use a reduced set of measurement record $\{N_m \equiv n_m + n_{-m}; m > 0\}$ for estimation. It can be shown that each N_m is also Poissonian conditioned on the mean $2|\alpha|^2|x_m|^2$, but now they remain independent from one another in the set after averaging over $\{x_m; m > 0\}$.

With x_m being complex Gaussian and N_m being conditionally Poissonian with mean $2|\alpha|^2|x_m|^2$, it can be shown that the marginal distribution of N_m is a Bose-Einstein distribution with mean number [79]

$$\bar{N}_m = 2|\alpha|^2 S_X(\omega_m|\theta), \tag{4.46}$$

and probability

$$P_n^{(m)}(n_m|\theta) = \frac{1}{\bar{N}_m + 1}\left(\frac{\bar{N}_m}{\bar{N}_m + 1}\right)^{n_m}. \tag{4.47}$$

The classical Fisher information for each $P_n^{(m)}$ is given by Eq. (2.60),

$$j_{\mu\nu}[P_n^{(m)}] = \sum_{n_m} P_n^{(m)} \frac{\partial \ln P_n^{(m)}}{\partial \theta_\mu} \frac{\partial \ln P_n^{(m)}}{\partial \theta_\nu}$$

$$= \frac{\partial \bar{N}_m}{\partial \theta_\mu} \frac{\partial \bar{N}_m}{\partial \theta_\nu} \sum_{n_m} P_n^{(m)} \left[\frac{\partial \ln P_n^{(m)}}{\partial \theta_\mu} \right]^2$$
$$= \frac{\partial \bar{N}_m}{\partial \theta_\mu} \frac{\partial \bar{N}_m}{\partial \theta_\nu} \frac{1}{\bar{N}_m(\bar{N}_m + 1)}. \tag{4.48}$$

The Fisher information $j(P_Y^{(\mathrm{spc})})$ is obtained by summing the classical information from each mode,

$$j[P_Y^{(\mathrm{spc})}] = \sum_{m>0} j[P_n^{(m)}],$$
$$j_{\mu\nu}[P_Y^{(\mathrm{spc})}] = \sum_{m>0} \frac{(\partial \ln \bar{N}_m/\partial\theta_\mu)(\partial \ln \bar{N}_m/\partial\theta_\nu)}{1 + 1/\bar{N}_m}, \tag{4.49}$$

which leads to

$$j_{\mu\nu}[P_Y^{(\mathrm{spc})}] = T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2 + 1/(\mathcal{N} S_X)}, \tag{4.50}$$

if we use the SPLOT assumption to replace $\sum_{m>0}$ with $T\int_0^\infty d\omega/(2\pi)$ [10] and use the symmetry of the integrand to replace $T\int_0^\infty d\omega/(2\pi)$ with $(T/2)\int_{-\infty}^\infty d\omega/(2\pi)$. $\mathcal{N}$ denotes the average input photon flux. Since $S_I(\omega) = \mathcal{N}$ for a coherent state, Eq. (4.50) coincides with the quantum bound in Eq. (4.36). Comparing Eq. (4.50) with the homodyne limit given by Eq. (4.40), we can expect that spectral photon counting becomes significantly better than homodyne detection when $\mathcal{N} S_X$ is small.

4.3.3 Ornstein-Uhlenbeck spectrum analysis

For a more specific example, consider the experiments in Refs. [48, 49], which can be modeled as the continuous-optical-phase-modulation problem depicted in Fig. 4.1(b), with adaptive homodyne detection depicted in Fig. 4.1(c) and $X(t)$ given by an Ornstein-Uhlenbeck process with PSD

$$S_X(\omega|\theta) = \frac{2\theta_1\theta_2}{\omega^2 + \theta_2^2}, \tag{4.51}$$

where $\theta_1 = \mathbb{E}_{X|\theta}[X^2(t)]$ is the area under S_X and θ_2 is the spectral width. The experimental S_I can be assumed to be constant for all frequencies of interest, and the quantum limit given by Eq. (4.36) on the estimation of θ_1 and θ_2 can be computed analytically:

$$\begin{aligned}
\tilde{\mathcal{J}}_{11} &= \frac{\theta_2 T}{8\theta_1^2} \frac{C}{\sqrt{1+C/2}}, \\
\tilde{\mathcal{J}}_{22} &= \frac{2T}{\theta_2} \frac{1+C/4}{C} \left(\frac{1+C/4}{\sqrt{1+C/2}} - 1 \right), \\
\tilde{\mathcal{J}}_{12} = \tilde{\mathcal{J}}_{21} &= \frac{T}{2\theta_1} \left(\frac{1+C/4}{\sqrt{1+C/2}} - 1 \right),
\end{aligned} \tag{4.52}$$

where

$$C \equiv \frac{8\theta_1 S_I}{\theta_2} = 4 S_I S_X(0|\theta) \tag{4.53}$$

is an SNR quantity. For comparison, the homodyne limit given by Eq. (4.40) is

$$\begin{aligned}
\tilde{j}_{11} &= \frac{\theta_2 T}{8\theta_1^2} \frac{C^2}{(1+C)^{3/2}}, \\
\tilde{j}_{22} &= \frac{2T}{\theta_2} \frac{1}{C} \left[\frac{(1+C/2)(1+5C/4+C^2/8)}{(1+C)^{3/2}} - \left(1 + \frac{C}{4}\right) \right], \\
\tilde{j}_{12} = \tilde{j}_{21} &= \frac{T}{2\theta_1} \left[\frac{1+3C/2+C^2/4}{(1+C)^{3/2}} - 1 \right].
\end{aligned} \tag{4.54}$$

For homodyne detection, C is an upper limit on the ratio between the peak of S_X and the homodyne noise floor S_η in the frequency domain.

Fig. 4.3 plots the quantum ($\tilde{\mathcal{J}}^{-1}$) and homodyne ($\tilde{j}^{-1}$) bounds on the estimation errors Σ_{11} and Σ_{22} versus C. Both plots show similar behaviors, and the $C \gg 1$ and $C \ll 1$ limits are of special interest. In the high-SNR

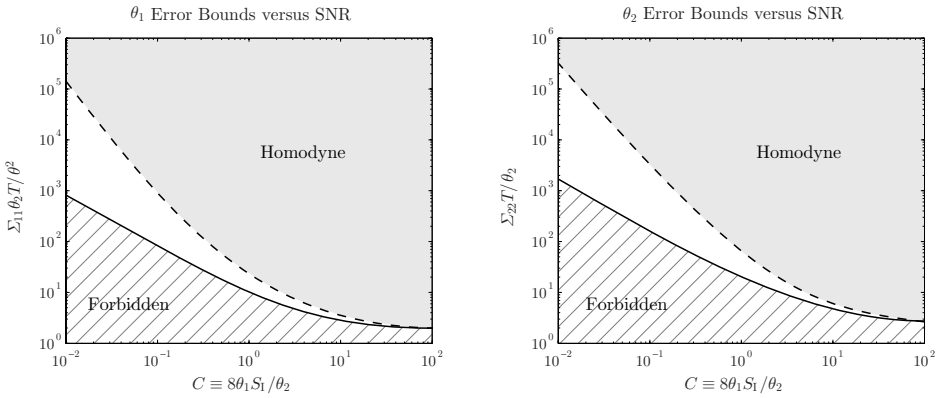


Figure 4.3: Log-log plots of the homodyne limit $\tilde{j}^{-1}$ (dashed line, given by the inverse of Eq. (4.54)) and quantum limit $\tilde{\mathcal{J}}^{-1}$ (solid line, given by the inverse of Eq. (4.52)) on the mean-square estimation errors versus an SNR quantity $C \equiv 8\theta_1 S_I/\theta_2$. (Left): limits on normalized Σ_{11} (in unit of $\theta_1^2/(\theta_2 T)$), (Right): limits on normalized Σ_{22} (in unit of θ_2/T). No measurement can achieve an error below the quantum limit (the "forbidden" region), and no homodyne measurement can achieve an error below the homodyne limit. For $C \gg 1$, the limits approach constants, while for $C \ll 1$ the homodyne limit has significantly worse error scaling.

regime ($C \gg 1$), both $\tilde{\mathcal{J}}^{-1}$ and $\tilde{j}^{-1}$ approach a C-independent limit:

$$\lim_{C\to\infty} \tilde{\mathcal{J}}^{-1} = \lim_{C\to\infty} \tilde{j}^{-1} = \frac{2}{\theta_2 T} \begin{pmatrix} \theta_1^2 & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_2^2 \end{pmatrix}, \tag{4.55}$$

and the homodyne performance is near-quantum-optimal. This asymptotic behavior is different from that of the bounds for single-parameter estimation, as both $1/\tilde{\mathcal{J}}_{\mu\mu}$ and $1/\tilde{j}_{\mu\mu}$ scale as $C^{-1/2}$ and decrease indefinitely for increasing C. The matrix bounds thus demonstrate the detrimental effect of having two unknown parameters that act as noise to each other. The C-independent limits also suggest that, once an experiment is in the high-SNR regime, no significant improvement can be made by increasing S_I and reducing the noise floor via photon-flux increase, squeezing, or changing the measurement method.

In the low-SNR regime ($C \ll 1$), on the other hand, it can be shown

that

$$\tilde{\mathcal{J}}^{-1} \approx \frac{8}{\theta_2 T} C^{-1} \begin{pmatrix} \theta_1^2 & 0 \\ 0 & 2\theta_2^2 \end{pmatrix}, \tag{4.56}$$

$$\tilde{j}^{-1} \approx \frac{16}{\theta_2 T} C^{-2} \begin{pmatrix} \theta_1^2 & \theta_1\theta_2 \\ \theta_1\theta_2 & 2\theta_2^2 \end{pmatrix}, \tag{4.57}$$

where the homodyne bounds on Σ_{11} and Σ_{22} diverge from the quantum bounds by a large factor of $2/C \gg 1$. The diverging bounds demonstrate the importance of quantum-optimal measurement in the low-SNR limit: at least for a coherent-state input and weak modulation, the quantum-optimal performance of spectral photon counting can exhibit a superior error scaling and offer significant improvements over homodyne detection.

4.4 Experimental Data Analysis

To compare our theory with actual experimental performance, we analyze the data from the experiment reported in Ref. [48], which is in a high-SNR regime ($C \geq 23.5$) where the adaptive homodyne performance is expected to be close to our quantum limit. We focus on the experiment with coherent states and not the one with squeezed states reported in Ref. [49], as Eq. (4.55) implies that squeezing offers insignificant improvement in this high-SNR regime. The details of our statistical data analysis are described in Sec. 4.4.1, and the experimental estimation errors are plotted in Fig. 4.4, together with the quantum limit given by the inverse of $\tilde{\mathcal{J}}$ in Eq. (4.52) and the homodyne limit given by the inverse of $\tilde{j}$ in Eq. (4.54). The plots demonstrate estimation errors close to both the homodyne limit and the fundamental quantum limit, despite experimental imperfections such as imperfect phase locking.

The experiment reported in Ref. [48] used a beam where its phase is

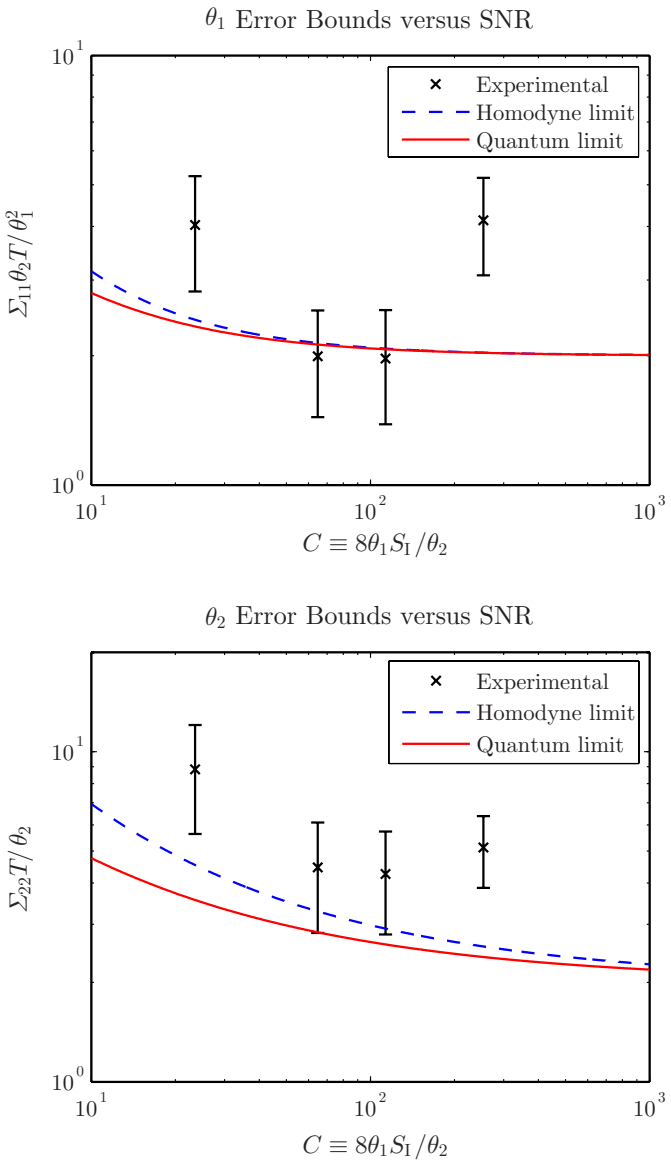


Figure 4.4: Log-log plots of the experimental mean-square estimation error Σ of Eq. (4.64) with error bar V of Eq. (4.65), the theoretical homodyne limit j^{-1} and the quantum limit $\tilde{\mathcal{J}}$ versus the SNR quantity $C \equiv 8\theta_1 S_I/\theta_2$. (Top) Experimental $\Sigma_{11} = \{4.0 \pm 1.2, 2.0 \pm 0.6, 2.0 \pm 0.6, 4.4 \pm 1.1\}$ (in unit of $\theta_1^2/(\theta_1 T)$) versus $C = \{23.5, 64.8, 113, 254\}$. (Bottom) Experimental $\Sigma_{22} = \{8.7 \pm 3.2, 4.4 \pm 1.6, 5.2 \pm 1.7, 6.4 \pm 1.4\}$ (in unit of $\theta_1^2/(\theta_1 T)$) versus C of the same value.

modulated with an Ornstein-Uhlenbeck signal $X(t)$ depicted in Fig. 4.1(b). The phase-modulated beam was measured by adaptive homodyne detection described in Sec. 4.3.1 and depicted in Fig. 4.1(c). The experiment used four different mean photon fluxes and for each photon flux $\mathcal{N}_k$, M_k traces of $X(t)$ and M_k traces of $Y(t)$ were recorded as listed in Table 4.1. Each trace of $Y(t)$ was obtained using a different feedback gain for the filter in the phase-locked loop, such that the phase locking might not be optimal. The original purpose of varying the feedback gains was to demonstrate the existence of an optimal filter for phase estimation in Ref. [48], but it is appropriate for our present analysis, as θ_1, θ_2 and the optimal filter are supposed to be unknown.

Table 4.1: The mean photon fluxes $\mathcal{N}_k$ and the number of records M_k for the experiment

Experiments	Mean photon flux ($\mathcal{N}_k$)	Number of records (M_k)
$k = 1$	1.315×10^6 s^{-1}	21
$k = 2$	3.616×10^6 s^{-1}	23
$k = 3$	6.327×10^6 s^{-1}	24
$k = 4$	1.418×10^7 s^{-1}	27

We assume further that the phase locking remained tight even if the filter was suboptimal, such that we can still use the linearized model

$$Y(t) = \sin[X(t) - \check{X}(t)] + \eta(t) + \check{X}(t) \approx X(t) + \eta(t), \tag{4.58}$$

where $\check{X}(t)$ is the feedback phase modulation on the local oscillator. Comparisons of the experimental $X(t)$ with $\check{X}(t)$ show that $\mathbb{E}[X(t) - \check{X}(t)]^2 \lesssim 0.3$ and the linearized model is thus reasonable. In Sec. 4.4.2, we describe further calibrations that ensure Eq. (4.58) is accurate. Most metrological experiments, such as gravitational-wave detectors, deal with extremely weak phase modulation, so the linearized model is expected to be even more accurate in those cases.

4.4.1 Procedure

In the last chapter, we perform the ML estimation for any observation time T using an EM algorithm [12, 23]. However, numerical simulations suggest that in this case it is safe to use Whittle estimator [14] given by Eq. (2.56) which is faster. For discrete-time measurement record $\{y(t_k)\}$ where $t_k = k\delta t, k = 0, 1, \ldots, K-1$, Whittle method approximates the likelihood function under SPLOT assumption, up to a θ-independent additive constant $\mathcal{A}$:

$$\ln P_Y \approx \mathcal{A} - \sum_{0<m<L/2} \left[\ln S_Y(\omega_m|\theta) + \frac{|\tilde{Y}_m|^2}{S_Y(\omega|\theta)} \right], \tag{4.59}$$

where $\tilde{Y}_m$ is the discrete Fourier transform of y defined in Eq. (2.53). Approximate ML estimation can then be performed by Fourier-transforming the time series into $\{\tilde{Y}_m\}$ and finding the parameters that maximize Eq. (4.59). We use Matlab® and its `fft` and `fminunc` functions to implement this procedure on the same desktop PC used in previous chapter. With $T = 0.01$ s for each $Y(t)$ trace, we expect the SPLOT assumption to be reasonable. We also perform numerical simulations throughout our analysis to ensure that our SPLOT and unbiased-estimator assumptions are valid and our results are expected.

To prevent technical noise and model mismatch at higher frequencies from contaminating our analysis, we consider only the spectral components up to 6×10^5 rad/s $\sim 10\theta_2$, rather than the full measurement bandwidth $\pi/\delta t = \pi \times 10^8$ rad/s. To estimate the true parameters more accurately, we apply the Whittle method to the collective record of all $\sum_k M_k = 95$ experimental $X(t)$ traces, assuming the spectrum given by Eq. (4.51), and obtain

$$\theta_1 = 0.1323$$

$$\theta_2 = 5.909 \times 10^4 \text{ rad/s}. \tag{4.60}$$

We take these to be the true parameters, as the estimates from such a large number of $X(t)$ traces are expected to be much more accurate than those from each $Y(t)$ trace.

We apply the Whittle method to each $Y(t)$ trace and evaluate the estimation errors by comparing the estimates with the true parameters. For each photon flux we assume a noise floor that is estimated from high-frequency data, and then we estimate θ using spectral components of Y up to $\omega = 6 \times 10^5$ rad/s. Let the resulting estimates be

$$\{\check{\theta}_{\mu k}^{(m_k)}; \mu = 1, 2; k = 1, 2, 3, 4; m_k = 1, \ldots, M_k\}, \tag{4.61}$$

where μ is the index for the two parameters, k is the index for the photon fluxes, and m_k is the index for the traces, and let the squared distance of each estimate from the true parameter be

$$\varepsilon_{\mu k}^{(m_k)} \equiv (\check{\theta}_{\mu k}^{(m_k)} - \theta_\mu)^2. \tag{4.62}$$

$\varepsilon_{\mu k}^{(m_k)}$ can be regarded as an outcome for a random variable $\varepsilon_{\mu k}$, so we can use the sample mean

$$\bar{\varepsilon}_{\mu k} \equiv \frac{1}{M_k} \sum_{m_k=1}^{M_k} \varepsilon_{\mu k}^{(m_k)} \tag{4.63}$$

to estimate the expected error

$$\Sigma_{\mu\mu} = \mathbb{E}_Y(\varepsilon_{\mu k}). \tag{4.64}$$

To find the deviation of the sample mean $\bar{\varepsilon}_{\mu k}$ from the expected value, we

use an unbiased estimate of the variance of $\varepsilon_{\mu k}$, that is,

$$V_{\mu k} \equiv \frac{1}{M_k - 1} \sum_{m_k=1}^{M_k} (\varepsilon_{\mu k}^{(m_k)} - \bar{\varepsilon}_{\mu k})^2, \tag{4.65}$$

and divide it by the number of samples M_k. Our final results $\{\bar{\varepsilon}_{\mu k} \pm \sqrt{V_{\mu k}/M_k}; \mu = 1, 2; k = 1, 2, 3, 4\}$ are plotted in normalized units in Fig. 4.4.

4.4.2 Data recalibration

In the experiment described in Ref. [48], calibration procedures were used to convert applied and measured voltages to the various physical quantities defined throughout Ref. [48]. In the course of analyzing that experimental data for the purpose of the new estimation task described here, we found that the data gives non-negligible bias in the estimation of θ_1. It turns out that the original calibration of experimental data was not accurate enough for the new task of estimating θ_1 (note that θ_2 is robust against this inaccuracy). The systematic calibration error had insignificant effects on the phase estimation task in Ref. [48] – making the estimate slightly worse than it would have been without the bias but generally within the uncertainty of the experiment as reported in Ref. [48]. The bias might have been caused by non-linearity or saturation of electronic circuits during the calibration phase of the experiment or long timescale drift. For our purpose here, we refine the calibration of the data from Ref. [48] so that we can achieve an accurate estimate. To do this fairly and independently, we use two extra data sets ($k = 5, 6$), which were not shown in Ref. [48] but recorded by the same experimental setup with different experimental parameters. Mean photon fluxes of these data sets are $\mathcal{N}_5 = 6.198 \times 10^6$ s^{-1} and $\mathcal{N}_6 = 5.986 \times 10^6$ s^{-1}. Number of traces are $M_5 = 24$ and $M_6 = 24$. Note that we use these "training" data only for the purposes of refining the experimental calibration. We apply the Whittle method to the two extra

data sets to obtain the true θ_1 from the collective record of $X(t)$, and a mean value of the estimated θ_1 from the collective record of $Y(t)$ traces using the coarse calibration from Ref. [48]. We determine that a refined calibration factor of 0.8945 is required to cancel the unwanted bias in the estimate of θ_1 for the extra data sets $k = 5, 6$. We then apply the refinement factor to $Y(t)$ of the original data sets ($k = 1$ to 4). By this method, we can refine the calibration of the original data presented in Ref. [48] by making use of independent, but contemporaneously recorded data.

4.5 Discussion

We have presented three key results: a measurement-independent quantum limit to the estimation of spectral parameters, the optimality of spectral photon counting, and an experimental data analysis. The quantum limit applies to a wide range of experiments and is particularly relevant to optomechanics, where the spectrum parameters of a stochastic force are often of interest to metrological applications [61, 62, 63, 83]. The proposed spectral photon counting method will be useful whenever the problem can be modeled as weak phase modulation of a coherent state and the SNR is low. Most metrological experiments, including gravitational-wave detectors, involve extremely weak phase modulation and low SNR. So the potential improvement over homodyne or heterodyne detection without the need of squeezed light is an important discovery. Our experimental data analysis further demonstrates the relevance of our theory to current technology and provides a recipe for future spectrum-analysis experiments.

There are many interesting potential extensions of our theory. Although quantum baths can often be modeled classically, a generalization of our formalism to account for nonclassical baths will make our theory applicable to an even wider range of experiments. A generalization for nonstationary

processes and finite observation time will be valuable for the study of unstable systems, which are potentially more sensitive than stable systems [84]. Tighter quantum limits that explicitly account for decoherence may be derived by applying the techniques in Refs. [54, 77, 85]. A Bayesian formulation that removes the unbiased-estimator assumption should be possible [1, 53, 54, 86, 87]. A more detailed study of our theory in the context of optomechanics can serve as an extension of Refs. [62, 63] and enable a more rigorous analysis of quantum limits to testing wavefunction-collapse models. Application of our theory to spin systems will provide a more rigorous foundation for stochastic magnetometry [88].

The actual performance of spectral photon counting depends on the bandwidth and spectral resolution of the Fourier-transform device, as well as the quantum efficiency and dark counts of the photodetectors in practice. While a more detailed analysis of such practical concerns is needed before one can judge the realistic performance of spectral photon counting with current technology, the large potential improvement in the low-SNR regime indicates the fundamental importance of coherent optical information processing for sensing applications and should motivate further technological advances in coherent quantum optical devices [89, 90]. In the high-SNR regime, on the other hand, our theory and experimental data analysis suggest that current technology can already approach the quantum limits with homodyne or even heterodyne detection. In this regime, our quantum limit primarily serves as a no-go theorem, proving that no other measurement can offer significant improvement. The challenge for actual metrological experiments will be to reach the high-SNR regime for weak signals, in which case our theory should serve as a rigorous foundation to guide future experimental designs.

Chapter 5

Imaging

In this chapter, we reconsider the problem of resolution in the field of optical imaging from the perspective of quantum estimation theory using the QCR lower bound [10, 15].

Rayleigh's criterion for resolution of two incoherent point sources [91] has been the most accepted criterion for optical resolution since its formulation in year 1879. In the past few decades, advances in far-field super-resolution techniques in microscopy [92, 93, 94] (see Ref. [95] for a review) enable us to sidestep Rayleigh's limit. Still, as they require that nearby sources are not emitting at the same time, those technologies do not challenge Rayleigh's criterion fundamentally for independently emitting sources; see Appendix A for details.

Here we study the problem of two-dimensional localization of two incoherent optical point sources. In Sec. 5.3, we obtain the fundamental limit for the estimation of the Cartesian components of the centroid and separation of the sources, and compare the QCR bound with the classical limit for direct imaging method. We propose two measurement schemes which asymptotically attain the QCR bound for both components of the separation over many trials in Sections 5.4 and 5.5.

This chapter is based on the preprint listed in Ref. [96]

5.1 Motivation

Being rooted in the optical measurement technology of its era, Rayleigh's criterion neglects the discrete and stochastic nature of the photodetection process. By adopting a stochastic framework, the studies [97, 98, 99] gave a modern formulation of the criterion for two sources radiating independently and incoherently. Using the CR bound of classical estimation theory [1], they showed that the localization accuracy of any unbiased estimator based on image-plane photon counting deteriorates rapidly on approaching sub-Rayleigh separations.

Following a preliminary study of the fundamental localization limit in Ref. [100], Tsang *et al.* [101] obtained the quantum limit on localizing two incoherent optical point sources in one-dimensional imaging. Their quantum bound is independent of separation and shows no deterioration when the two sources are closer than the conventional Rayleigh limit of the imaging system. Similar conclusions were reproduced using a semiclassical photodetection theory under a Poisson model [102]. In Ref. [101], a linear optical measurement – SPADE (SPAtial-mode DEmultiplexing)—was also proposed and shown to attain the QCR bound for any separation. Another measurement scheme proposed in Ref. [103]—SLIVER (Super Localization by Image inVERsion) interferometry—approaches the QCR bound for sub-Rayleigh separations.

Motivated by the work in Refs. [101, 102, 103], we study the problem of two-dimensional localization of two incoherent optical point sources. We first obtain the full 4-parameter quantum Fisher information matrix characterizing the precision of estimating all four (transverse) Cartesian coordinates of the two sources. We then focus on estimating the separations in the x and y directions, which the QCR bounds are shown to be independent of. Recent theoretical studies in quantum parameter estimation proved the existence of a POVM [10, 15] that achieves the QCR bound for

estimation of a single parameter [18, 19], though the quantum bound may not be attainable for two or more parameters. Here we propose two measurement schemes which asymptotically attain the QCR bound for both components of the separation over many trials. The first is an extension of the SLIVER scheme for estimating both separation parameters and approaches the QCR bound for small values of source separation. The second is a two-dimensional version of the SPADE scheme that attains the bound regardless of the distance between the two sources.

5.2 Source and System Model

We first lay out the source and imaging system model used in this study, the former being identical to that in Ref. [101]. We assume that two incoherent optical point sources with equal intensities are located on the object plane. Far-field radiation from these sources is collected at the entrance pupil of an optical imaging system such as a microscope or telescope. We assume that the paraxial approximation is valid for the field propagation from object plane to entrance pupil and consider a single polarization only. We further assume that the radiation from the sources is quasi-monochromatic and excites only a single temporal mode in order to focus on the spatial aspects of the resolution problem. We assume also that the imaging system is spatially-invariant [104], and that the image-plane coordinates (x, y) have been divided by the magnification factor and that $\psi(x, y)$ is the two-dimensional point-spread function (PSF) of the optical system satisfying the normalization condition $\int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y |\psi_s(x, y)|^2 = 1$ on the image plane.

Given two incoherent optical point sources with equal intensities located at coordinates (X_1, Y_1) and (X_2, Y_2) on the object plane, we assume the two sources are weak enough so that the probability ϵ of one or the other source

(but not both) emitting one photon satisfies

$$\epsilon \ll 1. \tag{5.1}$$

We further assume that the probability of together emitting more than one photon is negligible. Under the above assumptions, the quantum density operator of the optical field on the image plane can be written as [101]

$$\hat{\rho} = (1-\epsilon)|\text{vac}\rangle\langle\text{vac}| + \frac{\epsilon}{2}(|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|), \tag{5.2}$$

where $|\text{vac}\rangle$ denotes the vacuum state and states $\{|\psi_s\rangle\}_{s=1}^2$ are given by

$$|\psi_s\rangle = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y\, \psi_s(x,y)|x,y\rangle, \quad s = 1,2, \tag{5.3}$$

with the wavefunctions

$$\psi_s(x,y) = \psi(x - X_s, y - Y_s), \quad s = 1,2, \tag{5.4}$$

and $|x,y\rangle$ denotes the state with one photon in the mode corresponding to position (x,y) alone such that $\langle x,y|x',y'\rangle = \delta(x-x')\delta(y-y')$.

Eq. (5.2) means that a photon arrives with equal probability $\epsilon/2$ from either of the two sources. If a photon arrives from the first source, it is in the state $|\psi_1\rangle$ with wavefunction $\psi_1(x,y)$; if it comes from the other source, it is in state $|\psi_2\rangle$ with wavefunction $\psi_2(x,y)$. The two states are not orthogonal in general, with overlap

$$\delta \equiv \langle\psi_1|\psi_2\rangle = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \psi_1^*(x,y)\psi_2(x,y) \neq 0. \tag{5.5}$$

The parameters we are interested in estimating are the four components

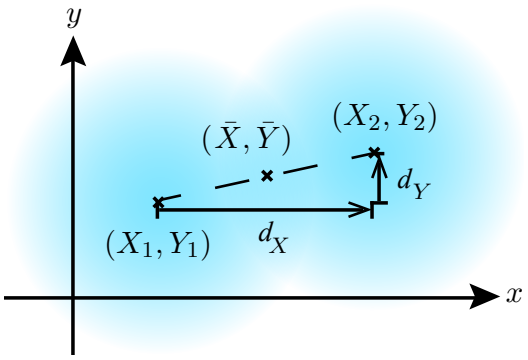


Figure 5.1: An illustration of the focused image of two point sources centered at (X_1, Y_1) and (X_2, Y_2). The shading indicates the approximate extent of the PSFs.

of the vector

$$\theta \equiv (\bar{X}, \bar{Y}, d_X, d_Y)^\top, \tag{5.6}$$

consisting of the centroid vector

$$(\bar{X}, \bar{Y}) = [(X_1, Y_1) + (X_2, Y_2)]/2, \tag{5.7}$$

and the separation vector

$$(d_X, d_Y) = (X_2, Y_2) - (X_1, Y_1), \tag{5.8}$$

as depicted in Fig. 5.1.

We also make a realistic simplifying assumption on the (possibly complex-valued) PSF $\psi(x, y)$ of the imaging system, namely that it is symmetric about the origin (or inversion-symmetric), viz.,

$$\psi(x, y) = \psi(-x, -y). \tag{5.9}$$

This assumption is satisfied for most imaging systems of interest, including spatially-invariant systems whose entrance aperture is rectangular or (hard or apodized) circular in shape [104], and is more general than the

assumption of a circularly-symmetric PSF used in Ref. [103]. Although the PSF may be complex-valued, the overlap δ is always real-valued. Using the definition of δ in Eq. (5.5), we have

$$\begin{aligned}\delta^* &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \psi^*(x - X_2, y - Y_2)\psi(x - X_1, y - Y_1) \\ &= \int_{-\infty}^{\infty} \mathrm{d}x' \int_{-\infty}^{\infty} \mathrm{d}y' \psi^*(x' - \frac{d_X}{2}, y' - \frac{d_Y}{2})\psi(x' + \frac{d_X}{2}, y' + \frac{d_Y}{2}) \end{aligned} \tag{5.10}$$

as we shift the coordinates, such that the centroid $(\bar{X}, \bar{Y})$ of Eq. (5.7) is the new origin. The terms d_X and d_Y are defined by Eq. (5.8). Using the assumption, $\psi(x, y)$ is symmetric about the origin, we have

$$\begin{aligned}\delta^* &= \int_{-\infty}^{\infty} \mathrm{d}x' \int_{-\infty}^{\infty} \mathrm{d}y' \psi^*(-x' + \frac{d_X}{2}, -y' + \frac{d_Y}{2})\psi(-x' - \frac{d_X}{2}, -y' - \frac{d_Y}{2}) \\ &= \int_{-\infty}^{\infty} \mathrm{d}x'' \int_{-\infty}^{\infty} \mathrm{d}y'' \psi^*(x'' + \frac{d_X}{2}, y'' + \frac{d_Y}{2})\psi(x'' - \frac{d_X}{2}, y'' - \frac{d_Y}{2}) \\ &= \delta, \end{aligned} \tag{5.11}$$

where in the second equality we apply a change of variables $x'' = -x'$ and $y'' = -y'$. Since the conjugate of δ is equal to itself, we show that δ is real-valued.

5.3 Quantum Limit on Two-source Localization

5.3.1 Review of the QCR bound

Let $\hat{\rho} = \hat{\rho}(\theta)$ be the density operator of a quantum system depending on an unknown vector parameter $\theta = (\theta_1, \theta_2, \dots)^\top$. Consider the estimation of θ from the quantum measurement outcome Y on M copies of $\hat{\rho}$. The

probability distribution of Y is given by

$$P(Y) = \mathrm{tr}[\hat{E}(Y)\hat{\rho}^{\otimes M}], \tag{5.12}$$

where $\hat{E}(Y)$ is the POVM that characterizes the measurement and $\hat{\rho}^{\otimes M}$ denotes the tensor product of M copies of $\hat{\rho}$. The estimation error covariance matrix Σ defined in Eq. (2.58) is bounded by the classical and quantum CR bound:

$$\Sigma \geq j^{-1} \geq J^{-1}, \tag{5.13}$$

which means that matrices $\Sigma - j^{-1}$, $\Sigma - J^{-1}$ and $j^{-1} - J^{-1}$ are positive-semidefinite. j is the classical Fisher information matrix given by Eq. (2.60) and J is the QFI matrix which can be expressed in terms of the SLD operators $\{\hat{\mathcal{L}}_\mu\}$ as

$$J_{\mu\nu} = M\mathrm{tr}\frac{\hat{\mathcal{L}}_\mu\hat{\mathcal{L}}_\nu + \hat{\mathcal{L}}_\nu\hat{\mathcal{L}}_\mu}{2}\hat{\rho}. \tag{5.14}$$

If $\hat{\rho}$ is diagonalized in an orthogonal basis $\{|e_n\rangle\}$, viz.,

$$\hat{\rho} = \sum_n D_n|e_n\rangle\langle e_n|, \tag{5.15}$$

$\hat{\mathcal{L}}_\mu$, can be expressed as [16]

$$\hat{\mathcal{L}}_\mu = \sum_{\substack{m,n \\ D_m+D_n\neq 0}} \frac{2}{D_m + D_n}\langle e_m|\frac{\partial\hat{\rho}_\theta}{\partial\theta_\mu}|e_n\rangle|e_m\rangle\langle e_n|. \tag{5.16}$$

5.3.2 Quantum Fisher information (QFI) matrix for two-source localization

We now consider the problem of estimation of the centroid and separation vectors for two incoherent point sources under the model of Section 5.2. Assuming the quantum density operator of Eq. (5.2) and the inversion-

symmetry of the PSF (Eq. (5.9)), the QFI matrix J can be evaluated using Eqs. (5.14) and (5.16).

We first diagonalize $\hat{\rho}$ as in Eq. (5.15), including enough eigenvectors to span the combined support of $\hat{\rho}$ and $\{\partial\hat{\rho}/\partial\theta_\mu\}$. The partial derivatives of $\hat{\rho}$ with respect to the object-plane source coordinates X_μ and Y_μ are

$$\frac{\hat{\partial}\rho}{\partial X_\mu} = \frac{\partial D_1}{\partial X_\mu}|e_1\rangle\langle e_1| + \frac{\partial D_2}{\partial X_\mu}|e_2\rangle\langle e_2| + \left(D_1\frac{\partial|e_1\rangle}{\partial X_\mu}\langle e_1| + D_2\frac{\partial|e_2\rangle}{\partial X_\mu}\langle e_2| + \text{H.c}\right), \tag{5.17}$$

$$\frac{\hat{\partial}\rho}{\partial Y_\mu} = \frac{\partial D_1}{\partial Y_\mu}|e_1\rangle\langle e_1| + \frac{\partial D_2}{\partial Y_\mu}|e_2\rangle\langle e_2| + \left(D_1\frac{\partial|e_1\rangle}{\partial Y_\mu}\langle e_1| + D_2\frac{\partial|e_2\rangle}{\partial Y_\mu}\langle e_2| + \text{H.c.}\right), \tag{5.18}$$

for $\mu = 1, 2$, where H.c. denotes the Hermitian conjugate.

After some algebra it can be shown that a possible set of eigenvectors of $\hat{\rho}$ is

$$|e_0\rangle = |\text{vac}\rangle, \tag{5.19}$$

$$|e_1\rangle = \frac{1}{\sqrt{2(1-\delta)}}(|\psi_1\rangle - |\psi_2\rangle), \tag{5.20}$$

$$|e_2\rangle = \frac{1}{\sqrt{2(1+\delta)}}(|\psi_1\rangle + |\psi_2\rangle), \tag{5.21}$$

$$|e_3\rangle = \frac{1}{c_3}\left[\Delta k_X(|\psi_{1X}\rangle + |\psi_{2X}\rangle) + r_+\Delta k_Y(|\psi_{1Y}\rangle + |\psi_{2Y}\rangle) - \frac{2(\gamma_X + r_+\gamma_Y)}{\sqrt{2(1-\delta)}}|e_1\rangle\right], \tag{5.22}$$

$$|e_4\rangle = \frac{1}{c_4}\left[\Delta k_X(|\psi_{1X}\rangle + |\psi_{2X}\rangle) - r_+\Delta k_Y(|\psi_{1Y}\rangle + |\psi_{2Y}\rangle) - \frac{2(\gamma_X - r_+\gamma_Y)}{\sqrt{2(1-\delta)}}|e_1\rangle\right], \tag{5.23}$$

$$|e_5\rangle = \frac{1}{c_5}\left[\Delta k_X(|\psi_{1X}\rangle - |\psi_{2X}\rangle) + r_-\Delta k_Y(|\psi_{1Y}\rangle - |\psi_{2Y}\rangle) + \frac{2(\gamma_X + r_-\gamma_Y)}{\sqrt{2(1+\delta)}}|e_2\rangle\right], \tag{5.24}$$

$$|e_6\rangle = \frac{1}{c_6}\Bigg[\Delta k_X(|\psi_{1X}\rangle - |\psi_{2X}\rangle) - r_-\Delta k_Y(|\psi_{1Y}\rangle - |\psi_{2Y}\rangle) + \frac{2(\gamma_X - r_-\gamma_Y)}{\sqrt{2(1+\delta)}}|e_2\rangle\Bigg], \tag{5.25}$$

where δ is given by Eq. (5.5),

$$|\psi_{1X}\rangle \equiv \frac{1}{\Delta k_X}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi(x-X_1,y-Y_1)}{\partial X_1}|x,y\rangle, \tag{5.26}$$

$$|\psi_{2X}\rangle \equiv \frac{1}{\Delta k_X}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi(x-X_2,y-Y_2)}{\partial X_2}|x,y\rangle, \tag{5.27}$$

$$|\psi_{1Y}\rangle \equiv \frac{1}{\Delta k_Y}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi(x-X_1,y-Y_1)}{\partial Y_1}|x,y\rangle, \tag{5.28}$$

$$|\psi_{2Y}\rangle \equiv \frac{1}{\Delta k_Y}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi(x-X_2,y-Y_2)}{\partial Y_2}|x,y\rangle, \tag{5.29}$$

$$\Delta k_X^2 \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\left|\frac{\partial\psi(x,y)}{\partial x}\right|^2, \tag{5.30}$$

$$\Delta k_Y^2 \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\left|\frac{\partial\psi(x,y)}{\partial y}\right|^2, \tag{5.31}$$

$$\gamma_X \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\,\psi^*(x-d_X,y-d_Y)\frac{\partial\psi(x,y)}{\partial x}, \tag{5.32}$$

$$\gamma_Y \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\,\psi^*(x-d_X,y-d_Y)\frac{\partial\psi(x,y)}{\partial y}, \tag{5.33}$$

$$a \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi^*(x,y)}{\partial x}\frac{\partial\psi(x,y)}{\partial y}, \tag{5.34}$$

$$a_s \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi^*(x,y)}{\partial x}\frac{\partial\psi(x-d_X,y-d_Y)}{\partial y}, \tag{5.35}$$

$$b_X^2 \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi^*(x-X_1,y-Y_1)}{\partial X_1}\frac{\partial\psi(x-X_2,y-Y_2)}{\partial X_2}, \tag{5.36}$$

$$b_Y^2 \equiv \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{\partial\psi^*(x-X_1,y-Y_1)}{\partial Y_1}\frac{\partial\psi(x-X_2,y-Y_2)}{\partial Y_2}, \tag{5.37}$$

$$r_\pm \equiv \left[\frac{\Delta k_X^2 \pm b_X^2 - \gamma_X^2/(1\mp\delta)}{\Delta k_Y^2 + b_Y^2 - \gamma_Y^2/(1\mp\delta)}\right]^{1/2}\exp\left[-i\arg\left(a \pm a_s - \frac{\gamma_X\gamma_Y}{1\mp\delta}\right)\right], \tag{5.38}$$

and normalization constants,

$$c_3 \equiv 2\sqrt{\Delta k_X^2 + b_X^2 - \frac{\gamma_X^2}{1-\delta} + |r_+|\left|a + a_s - \frac{\gamma_X\gamma_Y}{1-\delta}\right|}, \tag{5.39}$$

$$c_4 \equiv 2\sqrt{\Delta k_X^2 + b_X^2 - \frac{\gamma_X^2}{1-\delta} - |r_+|\left|a + a_s - \frac{\gamma_X\gamma_Y}{1-\delta}\right|}, \tag{5.40}$$

$$c_5 \equiv 2\sqrt{\Delta k_X^2 - b_X^2 - \frac{\gamma_X^2}{1+\delta} + |r_-| \left| a - a_s - \frac{\gamma_X \gamma_Y}{1+\delta} \right|}, \tag{5.41}$$

$$c_6 \equiv 2\sqrt{\Delta k_X^2 - b_X^2 - \frac{\gamma_X^2}{1+\delta} - |r_-| \left| a - a_s - \frac{\gamma_X \gamma_Y}{1+\delta} \right|}. \tag{5.42}$$

The eigenvalues of $\hat{\rho}$ (D_n corresponding to $|e_n\rangle$) are

$$\begin{aligned} &D_0 = 1 - \epsilon, \quad D_1 = \frac{\epsilon}{2}(1-\delta), \quad D_2 = \frac{\epsilon}{2}(1+\delta), \\ &D_3 = D_4 = D_5 = D_6 = 0. \end{aligned} \tag{5.43}$$

To verify the orthogonality of $\{|e_n\rangle\}, n = 1, \ldots, 6$, we first realize that $\langle e_1|e_2\rangle$, $(\langle\psi_{1X}| + \langle\psi_{2X}|)(|\langle\psi_{1X}\rangle - |\psi_{2X}\rangle)$ and $(\langle\psi_{1Y}| + \langle\psi_{2Y}|)(|\langle\psi_{1Y}\rangle - |\psi_{2Y}\rangle)$ are equal to zero. Then we show that

$$\begin{aligned} (\langle\psi_{1X}| + \langle\psi_{2X}|)(|\langle\psi_{1Y}\rangle - |\psi_{2Y}\rangle) &= 0, \\ (\langle\psi_{1Y}| + \langle\psi_{2Y}|)(|\langle\psi_{1X}\rangle - |\psi_{2X}\rangle) &= 0, \\ \langle e_1|(|\psi_{1X}\rangle - |\psi_{2X}\rangle) = \langle e_1|(|\psi_{1Y}\rangle - |\psi_{2Y}\rangle) &= 0, \end{aligned} \tag{5.44}$$

and

$$\langle e_2|(|\psi_{1X}\rangle + |\psi_{2X}\rangle) = \langle e_2|(|\psi_{1Y}\rangle + |\psi_{2Y}\rangle) = 0, \tag{5.45}$$

by using the fact that $\psi(x, y)$ is symmetric about the origin (see Eq. (5.9)). With these conditions, we prove that

1. $|e_1\rangle$ is orthogonal to both $|e_5\rangle$ and $|e_6\rangle$,
2. $|e_2\rangle$ is orthogonal to both $|e_3\rangle$ and $|e_4\rangle$,
3. $|e_3\rangle$ is orthogonal to both $|e_5\rangle$ and $|e_6\rangle$, and
4. $|e_4\rangle$ is orthogonal to both $|e_5\rangle$ and $|e_6\rangle$.

Next, the value of r_+ is chosen such that $\langle e_3|e_4\rangle = 0$ and the value of

r_- such that $\langle e_5|e_6\rangle = 0$. Finally, the last terms in $|e_3\rangle$, $|e_4\rangle$, $|e_5\rangle$, and $|e_6\rangle$ given by Eqs. (5.22)–(5.25) are chosen to ensure that $|e_3\rangle$ and $|e_4\rangle$ are orthogonal to $|e_1\rangle$, and $|e_5\rangle$ and $|e_6\rangle$ are orthogonal to $|e_2\rangle$.

With an orthogonal basis $\{|e_n\rangle\}$ that diagonalizes $\hat{\rho}$ in Eqs. (5.19)–(5.25), the SLDs in Eq. (5.16) with respect to the derivatives in Eqs. (5.17) and (5.18) can be expressed as:-

$$\begin{aligned}
\hat{\mathcal{L}}_\mu^{(X)} &= \sum_{\substack{n,m \\ D_n+D_m\neq 0}} \frac{2}{D_n+D_m}\langle e_n|\frac{\partial\rho}{\partial X_\mu}|e_m\rangle|e_n\rangle\langle e_m| = \sum_{n,m}\mathcal{L}_{\mu,nm}^{(X)}|e_n\rangle\langle e_m|, \\
\hat{\mathcal{L}}_\mu^{(Y)} &= \sum_{\substack{n,m \\ D_n+D_m\neq 0}} \frac{2}{D_n+D_m}\langle e_n|\frac{\partial\rho}{\partial Y_\mu}|e_m\rangle|e_n\rangle\langle e_m| = \sum_{n,m}\mathcal{L}_{\mu,nm}^{(Y)}|e_n\rangle\langle e_m|,
\end{aligned} \tag{5.46}$$

for $\mu = 1, 2$, with Hermitian matrices

$$\mathcal{L}_{\mu,nm}^{(X)} = \left(\mathcal{L}_{\mu,mn}^{(X)}\right)^* \quad \text{and} \quad \mathcal{L}_{\mu,nm}^{(Y)} = \left(\mathcal{L}_{\mu,mn}^{(Y)}\right)^*, \qquad \mu = 1, 2, \tag{5.47}$$

the nonzero and unique elements of which are found to be

$$\begin{aligned}
&\mathcal{L}_{1,11}^{(X)} = -\mathcal{L}_{2,11}^{(X)} = \gamma_X/(1-\delta), && \mathcal{L}_{1,11}^{(Y)} = -\mathcal{L}_{2,11}^{(Y)} = \gamma_Y/(1-\delta), \\
&\mathcal{L}_{1,12}^{(X)} = \mathcal{L}_{2,12}^{(X)} = \gamma_X\delta/\sqrt{1-\delta^2}, && \mathcal{L}_{1,12}^{(Y)} = \mathcal{L}_{2,12}^{(Y)} = \gamma_Y\delta/\sqrt{1-\delta^2}, \\
&\mathcal{L}_{1,13}^{(X)} = -\mathcal{L}_{2,13}^{(X)} = c_3/[2\sqrt{2(1-\delta)}], && \mathcal{L}_{1,13}^{(Y)} = -\mathcal{L}_{2,13}^{(Y)} = c_3/[2r_+\sqrt{2(1-\delta)}], \\
&\mathcal{L}_{1,14}^{(X)} = -\mathcal{L}_{2,14}^{(X)} = c_4/[2\sqrt{2(1-\delta)}], && \mathcal{L}_{1,14}^{(Y)} = -\mathcal{L}_{2,14}^{(Y)} = -c_4/[2r_+\sqrt{2(1-\delta)}], \\
&\mathcal{L}_{1,15}^{(X)} = \mathcal{L}_{2,15}^{(X)} = c_5/[2\sqrt{2(1-\delta)}], && \mathcal{L}_{1,15}^{(Y)} = \mathcal{L}_{2,15}^{(Y)} = c_5/[2r_-\sqrt{2(1-\delta)}], \\
&\mathcal{L}_{1,16}^{(X)} = \mathcal{L}_{2,16}^{(X)} = c_6/[2\sqrt{2(1-\delta)}], && \mathcal{L}_{1,16}^{(Y)} = \mathcal{L}_{2,16}^{(Y)} = -c_6/[2r_-\sqrt{2(1-\delta)}], \\
&\mathcal{L}_{1,22}^{(X)} = -\mathcal{L}_{2,22}^{(X)} = -\gamma_X/(1+\delta), && \mathcal{L}_{1,22}^{(Y)} = -\mathcal{L}_{2,22}^{(Y)} = -\gamma_Y/(1+\delta), \\
&\mathcal{L}_{1,23}^{(X)} = \mathcal{L}_{2,23}^{(X)} = c_3/[2\sqrt{2(1+\delta)}], && \mathcal{L}_{1,23}^{(Y)} = \mathcal{L}_{2,23}^{(Y)} = c_3/[2r_+\sqrt{2(1+\delta)}], \\
&\mathcal{L}_{1,24}^{(X)} = \mathcal{L}_{2,24}^{(X)} = c_4/[2\sqrt{2(1+\delta)}], && \mathcal{L}_{1,24}^{(Y)} = \mathcal{L}_{2,24}^{(Y)} = -c_4/[2r_+\sqrt{2(1+\delta)}], \\
&\mathcal{L}_{1,25}^{(X)} = -\mathcal{L}_{2,25}^{(X)} = c_5/[2\sqrt{2(1+\delta)}], && \mathcal{L}_{1,25}^{(Y)} = -\mathcal{L}_{2,25}^{(Y)} = c_5/[2r_-\sqrt{2(1+\delta)}],
\end{aligned}$$

$$\mathcal{L}_{1,26}^{(X)} = -\mathcal{L}_{2,26}^{(X)} = c_6/[2\sqrt{2(1+\delta)}], \qquad \mathcal{L}_{1,26}^{(Y)} = -\mathcal{L}_{2,26}^{(Y)} = -c_6/[2r_-\sqrt{2(1+\delta)}]. \tag{5.48}$$

In terms of the centroid and separation parameters θ of Eq. (5.6), the SLDs are

$$\begin{aligned} \hat{\mathcal{L}} &= \hat{\mathcal{L}}_1^{(X)} + \hat{\mathcal{L}}_2^{(X)}, \qquad \hat{\mathcal{L}}_2 = \hat{\mathcal{L}}_1^{(Y)} + \hat{\mathcal{L}}_2^{(Y)}, \\ \hat{\mathcal{L}}_3 &= \frac{\hat{\mathcal{L}}_2^{(X)} - \hat{\mathcal{L}}_1^{(X)}}{2}, \qquad \hat{\mathcal{L}}_4 = \frac{\hat{\mathcal{L}}_2^{(Y)} - \hat{\mathcal{L}}_1^{(Y)}}{2}. \end{aligned} \tag{5.49}$$

Substitute Eqs. (5.43), (5.46), and (5.49), into Eq. (5.14) to evaluate the QFI. The QFI matrix in terms of θ defined in Eq. (5.6) is found to be

$$J = N \begin{pmatrix} 4\left(\Delta k_X^2 - \gamma_X^2\right) & 4\left[\mathrm{Re}(a) - \gamma_X\gamma_Y\right] & 0 & 0 \\ 4\left[\mathrm{Re}(a) - \gamma_X\gamma_Y\right] & 4\left(\Delta k_Y^2 - \gamma_Y^2\right) & 0 & 0 \\ 0 & 0 & \Delta k_X^2 & \mathrm{Re}(a) \\ 0 & 0 & \mathrm{Re}(a) & \Delta k_Y^2 \end{pmatrix}, \tag{5.50}$$

where $N = M\epsilon$ is the average photon number collected over M trials, Δk_X, Δk_Y, γ_X, γ_Y and a are defined in Eqs. (5.30)–(5.34), and $\mathrm{Re}(z)$ denotes the real part of z. The terms Δk_X and Δk_Y are related to the spatial spectral width of the PSF in x- and y-direction, respectively and, along with a, are independent of the source parameters θ. γ_X and γ_Y depend on the separation coordinates (d_X, d_Y) but not on the centroid coordinates $(\bar{X}, \bar{Y})$. Thus, J as a whole is independent of $(\bar{X}, \bar{Y})$, as may be expected from our assumption of a spatially-invariant imaging system. Note that J has a block-diagonal form with respect to the centroid and separation coordinate pairs, and that the matrix elements related to the estimation errors of separations d_X and d_Y—J_{33} and J_{44}—are independent of (d_X, d_Y) as well.

The QFI matrix can be simplified further for the case of a circularly-

symmetric PSF $\psi(x, y)$ for which

$$\Delta k_X = \Delta k_Y \equiv \Delta k. \tag{5.51}$$

Furthermore, as the circularly-symmetric $\psi(x, y)$ is also symmetric along y-axis, viz., $\psi(x, y) = \psi(-x, y)$, the integrand of a

$$\frac{\partial \psi^*(x,y)}{\partial x}\frac{\partial \psi(x,y)}{\partial y} = -\frac{\partial \psi^*(-x,y)}{\partial x}\frac{\partial \psi(-x,y)}{\partial y}. \tag{5.52}$$

Hence, upon integration in the xy-plane, the term a vanishes. The QFI matrix J becomes

$$J = N \begin{pmatrix} 4\left(\Delta k^2 - \gamma_X^2\right) & -4\gamma_X\gamma_Y & 0 & 0 \\ -4\gamma_X\gamma_Y & 4\left(\Delta k^2 - \gamma_Y^2\right) & 0 & 0 \\ 0 & 0 & \Delta k^2 & 0 \\ 0 & 0 & 0 & \Delta k^2 \end{pmatrix}. \tag{5.53}$$

5.3.3 Comparison to direct imaging

The QCR bound can be compared with the classical CR bound of the conventional direct imaging method. For direct imaging measurement, the statistics is expressed in terms of the mean intensity [101]

$$\Lambda(x, y) = \frac{1}{2}\left[|\psi_1(x, y)|^2 + |\psi_2(x, y)|^2\right], \tag{5.54}$$

such that classical Fisher information matrix

$$j_{\mu\nu}^{(\mathrm{dir})} = N \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \frac{1}{\Lambda(x,y)} \frac{\partial \Lambda(x,y)}{\partial \theta_\mu} \frac{\partial \Lambda(x,y)}{\partial \theta_\nu}. \tag{5.55}$$

For any PSD $\psi(x, y)$, let

$$I(x, y) \equiv |\psi(x, y)|^2 \tag{5.56}$$

be the intensity point spread function. We assume that the centroid $(\bar{X}, \bar{Y})$ is located at the origin and we estimate the separation parameter $\eta = (d_X, d_Y)^\top$. The mean intensity in Eq. (5.54) becomes

$$\Lambda(x,y) = \frac{1}{2}\left[I\left(x + \frac{d_X}{2}, y + \frac{d_Y}{2}) + I(x - \frac{d_X}{2}, y - \frac{d_Y}{2}\right)\right], \tag{5.57}$$

For small d_X and d_Y, we can expand $\Lambda(x,y)$ to the second order to obtain

$$\Lambda(x,y) = I(x,y) + \frac{d_X^2}{8}\frac{\partial^2 I(x,y)}{\partial x^2} + \frac{d_X d_Y}{4}\frac{\partial^2 I(x,y)}{\partial x \partial y} + \frac{d_Y^2}{8}\frac{\partial^2 I(x,y)}{\partial y^2} + o(d^2), \tag{5.58}$$

where $o(d^2)$ denotes terms asymptotically smaller than d_X^2, $d_X d_Y$, and d_Y^2. Substituting this equation into Eq. (5.55), we have the Fisher information matrix $j^{(\mathrm{dir})}$ in terms of η:-

$$\begin{aligned}
j_{11}^{(\mathrm{dir})} &= \frac{N}{16}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{1}{I(x,y)}\left[d_X\frac{\partial^2 I(x,y)}{\partial x^2} + d_Y\frac{\partial^2 I(x,y)}{\partial x \partial y}\right]^2 + o(d^2),\\
j_{22}^{(\mathrm{dir})} &= \frac{N}{16}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{1}{I(x,y)}\left[d_X\frac{\partial^2 I(x,y)}{\partial x \partial y} + d_Y\frac{\partial^2 I(x,y)}{\partial y^2}\right]^2 + o(d^2),\\
j_{12}^{(\mathrm{dir})} &= \frac{N}{16}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\frac{1}{I(x,y)}\left[d_X\frac{\partial^2 I(x,y)}{\partial x^2} + d_Y\frac{\partial^2 I(x,y)}{\partial x \partial y}\right]\\
&\quad\times\left[d_X\frac{\partial^2 I(x,y)}{\partial x \partial y} + d_Y\frac{\partial^2 I(x,y)}{\partial y^2}\right] + o(d^2).
\end{aligned} \tag{5.59}$$

These expressions can be further simplified for circularly symmetric PSD $\psi(x,y)$, thus circularly symmetric intensity $I(x,y)$. The Fisher information matrix becomes

$$j_{11}^{(\mathrm{dir})} = \frac{N}{16}(d_X^2\kappa_1 + d_Y^2\kappa_2), \tag{5.60}$$

$$j_{22}^{(\mathrm{dir})} = \frac{N}{16}(d_X^2\kappa_2 + d_Y^2\kappa_1), \tag{5.61}$$

$$j_{12}^{(\mathrm{dir})} = \frac{N}{16}d_X d_Y(\kappa_1 + \kappa_2). \tag{5.62}$$

where

$$\begin{aligned}\kappa_1 &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \frac{1}{I(x,y)} \left[\frac{\partial^2 I(x,y)}{\partial x^2}\right]^2 \\ &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \frac{1}{I(x,y)} \left[\frac{\partial^2 I(x,y)}{\partial y^2}\right]^2, \\ \kappa_2 &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \frac{1}{I(x,y)} \left[\frac{\partial^2 I(x,y)}{\partial x \partial y}\right]^2. \end{aligned} \tag{5.63}$$

Note that $\kappa_1, \kappa_2 > 0$ as $I(x,y)$ is nonnegative. The second equality is due to circularly symmetry property of $I(x,y)$; while the two terms

$$\int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \frac{1}{I(x,y)} \frac{\partial^2 I(x,y)}{\partial x^2} \frac{\partial^2 I(x,y)}{\partial x \partial y} = 0$$

and

$$\int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \frac{1}{I(x,y)} \frac{\partial^2 I(x,y)}{\partial x \partial y} \frac{\partial^2 I(x,y)}{\partial y^2} = 0$$

as their integrands have odd parity of the form $g(x,y) = -g(-x,y)$, hence, the integrations in the xy-plane result in zero. This parity follows from the fact that $I(x,y)$ is symmetric about y-axis (since it is circularly symmetry), viz., $I(x,y) = I(-x,y)$, which leads to

$$\frac{\partial^2 I(x,y)}{\partial x \partial y} = -\frac{\partial^2 I(-x,y)}{\partial x \partial y}. \tag{5.64}$$

For direct imaging method, the CR bound terms related to the estimation of separation d_X and d_Y are

$$\begin{aligned}\{[j^{(\mathrm{dir})}]^{-1}\}_{11} &= \frac{16}{N} \frac{d_X^2 \kappa_2 + d_Y^2 \kappa_1}{(d_X^2 - d_Y^2)^2 \kappa_1 \kappa_2}, \\ \{[j^{(\mathrm{dir})}]^{-1}\}_{22} &= \frac{16}{N} \frac{d_X^2 \kappa_1 + d_Y^2 \kappa_2}{(d_X^2 - d_Y^2)^2 \kappa_1 \kappa_2}, \end{aligned} \tag{5.65}$$

which approach infinity as $d_X, d_Y \to 0$.

To illustrate the above results, we assume a circular Gaussian PSF of

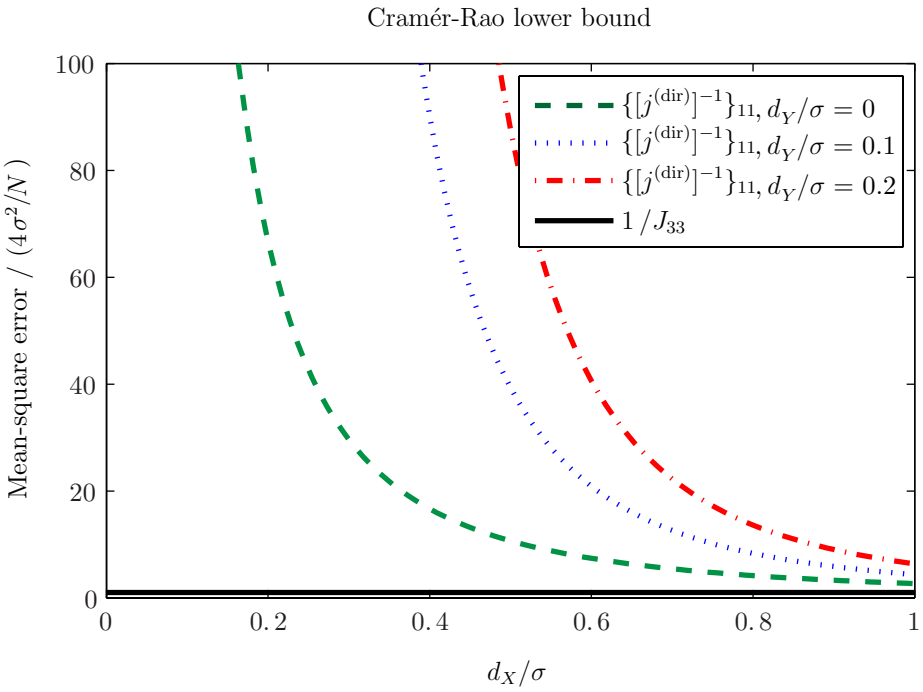


Figure 5.2: Quantum $(1/J_{33})$ and classical $(\{[j^{(\mathrm{dir})}]^{-1}\}_{11})$ CR bound versus normalized separation d_X/σ for a circular Gaussian PSD of Eq. (5.66). The classical bounds are plotted for different value of $d_Y/\sigma = 0, 0.1, 0.2$. The bounds are normalized with respect to the quantum value $4\sigma^2/N$.

the form

$$\psi(x,y) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left(-\frac{x^2+y^2}{4\sigma^2}\right). \tag{5.66}$$

The PSF-dependent terms are now

$$\Delta k = \frac{1}{2\sigma}, \qquad \kappa_1 = \frac{3}{2\sigma^4}, \qquad \text{and} \qquad \kappa_2 = \frac{1}{4\sigma^4}. \tag{5.67}$$

Eq. (5.65) becomes

$$\begin{aligned}\{[j^{(\mathrm{dir})}]^{-1}\}_{11} &= \frac{4\sigma^2}{N}\frac{8(d_X/\sigma)^2 + 48(d_Y/\sigma)^2}{3[(d_X/\sigma)^2 - (d_Y/\sigma)^2]^2},\\ \{[j^{(\mathrm{dir})}]^{-1}\}_{22} &= \frac{4\sigma^2}{N}\frac{48(d_X/\sigma)^2 + 8(d_Y/\sigma)^2}{3[(d_X/\sigma)^2 - (d_Y/\sigma)^2]^2},\end{aligned} \tag{5.68}$$

and Fig. 5.2 plots the QCR bound $1/J_{33}$ given by Eq. (5.53) and the CR bound $\{[j^{(\mathrm{dir})}]^{-1}\}_{11}$ for the estimation of d_X as function of separation parameters d_X and d_Y. From the symmetry in Eqs. (5.68), the CR bound for

the estimation of d_Y — $\{[j^{(\mathrm{dir})}]^{-1}\}_{22}$ — displays the same trend. The plot shows a huge divergence of the CR bound for direct imaging method from the quantum limit as d_X decreases; while the QCR bound stays constant regardless of separation. This means that a considerable improvement can be obtained if a quantum-optimal measurement can be implemented.

In Sections 5.4 and 5.5, we discuss concrete measurement schemes to simultaneously estimate the separation parameters $\eta = (d_X, d_Y)^\top$. For these schemes, we assume that the centroid $(\bar{X}, \bar{Y})$ has already been located, and compare their performance to the quantum bound obtained in Sec. 5.3.2.

5.4 Super Localization by Image Inversion (SLIVER) Interferometry

We now propose a two-stage interferometric scheme for estimation of d_X and d_Y adapting the SLIVER scheme of Ref. [103]. The scheme is analyzed in the framework of quantum theory under weak-source condition; while the analysis in Ref. [103] is of semiclassical photodetection theory [105, 106].

Assuming that the centroid of the sources is imaged at the origin of image-plane coordinates, the images of the sources are centered at $\mp\frac{1}{2}(d_X, d_Y)$ in the image plane. The quantum density operator of Eq. (5.2) is thus

$$\hat{\rho} = (1 - \epsilon_1 - \epsilon_2)|\mathrm{vac}\rangle\langle\mathrm{vac}| + \epsilon_1|\psi_1\rangle\langle\psi_1| + \epsilon_2|\psi_2\rangle\langle\psi_2|, \tag{5.69}$$

where ϵ_1 and ϵ_2 are the source intensities for source 1 and 2, respectively, $|\mathrm{vac}\rangle$ denotes the vacuum state and states $\{|\psi_s\rangle\}_{s=1}^2$ are given by Eq. (5.3) with the wavefunctions

$$\psi_1(x, y) = \psi\left(x + \frac{d_X}{2}, y + \frac{d_Y}{2}\right), \tag{5.70}$$

$$\psi_2(x, y) = \psi\left(x - \frac{d_X}{2}, y - \frac{d_Y}{2}\right). \tag{5.71}$$

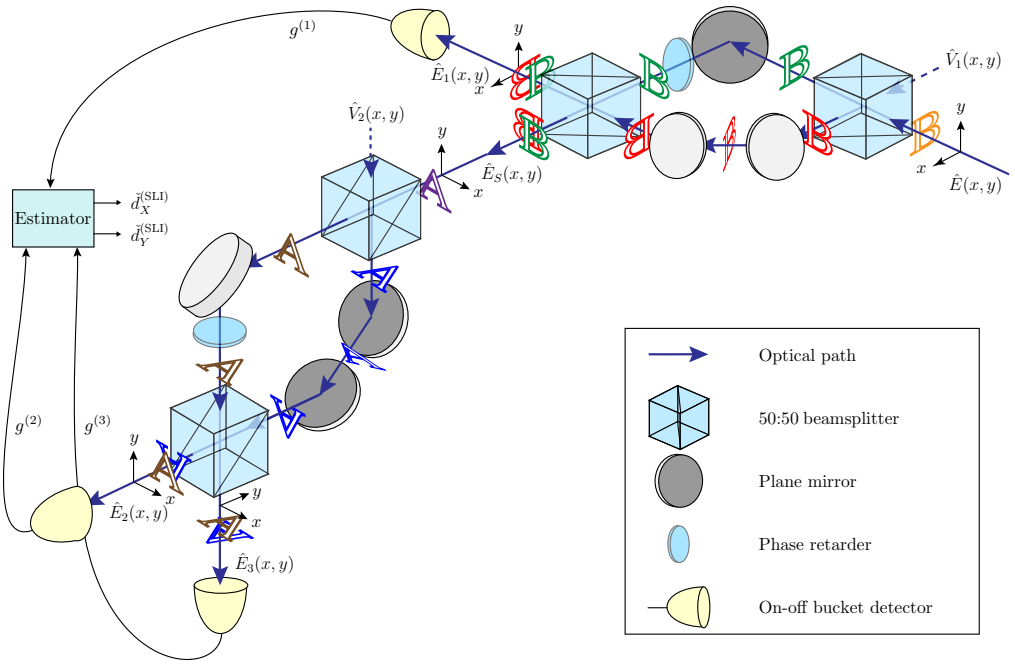


Figure 5.3: A proposed schematic of 2-stage SLIVER. The input field $\hat{E}(x,y)$ is separated into symmetric ($\hat{E}_S$) and antisymmetric ($\hat{E}_X$) components with respect to reflection about the y-axis. The $\hat{E}_X$ component impinges upon a bucket photodetector. The $\hat{E}_S$ component is separated again into symmetric and antisymmetric ($\hat{E}_Y$) components with respect to reflection about the x-axis, and the $\hat{E}_Y$ component impinges upon a second bucket detector. These field transformations are realized by the extra reflection at an appropriately aligned plane mirror in one arm of the balanced Mach-Zehnder interferometers. The set of binary outcomes – $g^{(1)}$, $g^{(2)}$ and $g^{(3)}$ – observed in the detectors over a series of M measurements is processed to give estimates $\check{d}_X$ and $\check{d}_Y$ of the components of the separation. The letters $\mathbb{A}$ and $\mathbb{B}$ illustrate the reflection of the field.

In this section, we relax the assumption of Sec. 5.2 that the two sources are of equal strength.

The scheme consists of two stages as illustrated in Fig. 5.3. The first stage involves the separation of the input field $\hat{E}(x,y)$ into antisymmetric ($\hat{E}_1$) and symmetric ($\hat{E}_S$) parts with respect to reflection about the y-axis. The symmetric and antisymmetric field components can be obtained by splitting the input field $\hat{E}(x,y)$ using a 50-50 beamsplitter, inverting the x-coordinates of the field (i.e., reflecting the field about the y-axis) from one output and recombining the two beams at a second 50-50 beamsplitter. The optics of the first stage thus consists of a balanced Mach-Zehnder

interferometer with an extra reflection at an appropriately aligned plane mirror in one arm. Consider the quantum treatment of beamsplitter input-output relations, the output field operators are given by

$$\hat{E}_1(x,y) = \frac{1}{2}\{[\hat{E}(x,y) - \hat{E}(-x,y)] + [\hat{V}_1(x,y) + \hat{V}_1(-x,y)]\}, \quad (5.72)$$

$$\hat{E}_S(x,y) = \frac{1}{2}\{[\hat{E}(x,y) + \hat{E}(-x,y)] + [\hat{V}_1(x,y) - \hat{V}_1(-x,y)]\}, \quad (5.73)$$

where $\hat{V}_1(x,y)$ is the input vacuum field operator at the open port of the first beamsplitter. At the antisymmetric output port with the field pattern of Eq. (5.72), an on-off non-spatially-resolving (bucket) detector is placed to distinguish between no photon and one photon. The measurement outcome $g^{(1)}$ is binary – zero if the detector does not click and one if it does.

In the second stage, the output beam $\hat{E}_S(x,y)$ of the symmetric port is used as input to a second interferometer which similarly splits the field into antisymmetric ($\hat{E}_2$) and symmetric ($\hat{E}_3$) components with respect to reflection about the x-axis. The output field operators are given by

$$\hat{E}_2(x,y) = \frac{1}{2}\{[\hat{E}_S(x,y) - \hat{E}_S(x,-y)] + [\hat{V}_2(x,y) + \hat{V}_2(x,-y)]\}, \quad (5.74)$$

$$\hat{E}_3(x,y) = \frac{1}{2}\{[\hat{E}_S(x,y) + \hat{E}_S(x,-y)] + [\hat{V}_2(x,y) - \hat{V}_2(x,-y)]\}, \quad (5.75)$$

where $\hat{V}_2(x,y)$ is the input vacuum field operator at the open port of the first beamsplitter in this stage. The output fields $\hat{E}_2$ and $\hat{E}_3$ of this stage impinge upon two individual on-off bucket detectors to give measurement outcomes $g^{(2)}$ and $g^{(3)}$, respectively. Similar to the previous stage, binary outcomes $g^{(2)}$ and $g^{(3)}$ – 0 if the corresponding detector does not click and 1 if it does – are recorded.

5.4.1 Detection probabilities

The expected photon number at the rth detector is $\mathrm{tr}(\hat{\rho}\hat{N}_r)$ where the total-photon-number operator

$$\hat{N}_r = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \hat{E}_r^{\dagger}(x,y)\hat{E}_r(x,y), \quad r = 1,2,3. \tag{5.76}$$

Our assumption of negligible probability for multiple photon emission events ensures that at most one photon will impinge upon any of the three photodetectors. There are no valid cases of having more than one detector clicks; in the event of a detector clicks, only one photon impinges upon it. Therefore, there are only four mutually exclusive measurement outcomes: Probability $P(0)$ corresponds to the case where no photon detected in any of the three detectors and probability $P(r)$ corresponds to the case where only the rth detector clicks, viz.,

$$\begin{aligned}
P(0) &= \Pr[g^{(1)} = 0, g^{(2)} = 0, g^{(3)} = 0], \\
P(1) &= \Pr[g^{(1)} = 1, g^{(2)} = 0, g^{(3)} = 0], \\
P(2) &= \Pr[g^{(1)} = 0, g^{(2)} = 1, g^{(3)} = 0], \\
P(3) &= \Pr[g^{(1)} = 0, g^{(2)} = 0, g^{(3)} = 1].
\end{aligned} \tag{5.77}$$

Since either zero or one photon arrives at each detector, the probability of only the r-th detector clicks is equal to the expected photon number at the r-th detector,

$$P(r) = \mathrm{tr}(\hat{\rho}\hat{N}_r) \tag{5.78}$$

$$= \sum_{s=1,2} \epsilon_s \langle \Psi_s | \hat{N}_r | \Psi_s \rangle, \tag{5.79}$$

where ϵ_s is given by Eq. (5.69) and the states

$$|\Psi_s\rangle = |\psi_s\rangle|0\rangle_1|0\rangle_2, \qquad s = 1, 2. \tag{5.80}$$

as the single-photon source states defined in Eq. (5.69) augmented with vacuum states in the beam-splitter open port modes $\hat{V}_1(x, y)$ and $\hat{V}_2(x, y)$. To evaluate $\mathrm{tr}(\hat{\rho}\hat{N}_r)$, from Eqs. (5.72)–(5.76) we need

$$\langle\Psi_s|\hat{E}^\dagger(x, y)\hat{E}(x', y')|\Psi_s\rangle, \tag{5.81}$$

$$\langle\Psi_s|\hat{E}^\dagger(x, y)\hat{V}_1(x', y')|\Psi_s\rangle, \tag{5.82}$$

$$\langle\Psi_s|\hat{E}^\dagger(x, y)\hat{V}_2(x', y')|\Psi_s\rangle, \tag{5.83}$$

$$\langle\Psi_s|\hat{V}_1^\dagger(x, y)\hat{V}_1(x', y')|\Psi_s\rangle, \tag{5.84}$$

$$\langle\Psi_s|\hat{V}_1^\dagger(x, y)\hat{V}_2(x', y')|\Psi_s\rangle, \tag{5.85}$$

$$\langle\Psi_s|\hat{V}_2^\dagger(x, y)\hat{V}_2(x', y')|\Psi_s\rangle, \qquad s = 1, 2, \tag{5.86}$$

for arbitrary (x, y) and (x', y'). Eq. (5.81) is the only nonzero term as $\hat{V}_1(x, y)$ and $\hat{V}_2(x, y)$ are in vacuum.

To evaluate Eq. (5.81), we first consider an arbitrary orthogonal basis $\{|\varphi_q\rangle\}_{q=0}^{\infty}$ and define operator $\hat{a}_q$ as the annihilation operator that reduces the occupation number of the state $|\varphi_q\rangle$ by 1. The image field operator $\hat{E}(x, y)$ can be expanded as

$$\hat{E}(x, y) = \sum_{q=0}^{\infty} \hat{a}_q \varphi_q(x, y), \tag{5.87}$$

where

$$\varphi_q(x, y) = \langle x, y|\varphi_q\rangle, \tag{5.88}$$

and state $|x, y\rangle$ is defined in Sec. 5.2. Using the following expression for $\hat{a}_q$,

$$\hat{a}_q = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \hat{E}(x, y) \varphi_q^*(x, y), \tag{5.89}$$

and the commutation relation $[\hat{E}(x, y), \hat{E}^\dagger(x', y')] = \delta(x - x')\delta(y - y')$, we have commutation relation

$$\begin{aligned}
[\hat{a}_p, \hat{a}_q^\dagger] &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \int_{-\infty}^{\infty} \mathrm{d}x' \int_{-\infty}^{\infty} \mathrm{d}y' \varphi_p^*(x, y) \varphi_q(x', y') \\
&\quad \times [E(x, y), E^\dagger(x', y')] \\
&= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \varphi_p^*(x, y) \varphi_q(x, y) \\
&= \delta_{pq},
\end{aligned} \tag{5.90}$$

where δ_{pq} is the Kronecker delta.

We can choose an orthogonal basis $\{|\varphi_q\rangle\}$ such that

$$|\varphi_0\rangle = |\psi_1\rangle, \tag{5.91}$$

where the state $|\psi_1\rangle$ is given by Eq. (5.3). Hence, the field operator becomes

$$\hat{E}(x, y) = \hat{a}_0 \psi_1(x, y) + \sum_{q=1}^{\infty} \hat{a}_q \varphi_q(x, y), \tag{5.92}$$

where wavefunction $\psi_1(x, y)$ is given by Eq. (5.4). Similarly, we can have another orthogonal basis $\{|\varphi_q'\rangle\}$ such that

$$|\varphi_0'\rangle = |\psi_2\rangle, \tag{5.93}$$

and the field operator is given by

$$\hat{E}(x, y) = \hat{a}_0' \psi_2(x, y) + \sum_{q=1}^{\infty} \hat{a}_q' \varphi_q'(x, y), \tag{5.94}$$

where state $|\psi_2\rangle$ is defined in Eq. (5.3) and wavefunction $\psi_2(x,y)$ is given by Eq. (5.4). The operators $\hat{a}_p$ and $\hat{a}'_q$ do not necessarily commute with one another and the state $|\varphi_p\rangle$ might not be orthogonal to $|\varphi_q\rangle$, for all p,q. Using the expansion of $\hat{E}(x,y)$ in Eqs. (5.92) and (5.94) and the commutation relation in Eq. (5.90), we can show that

$$\langle\Psi_s|\hat{E}^\dagger(x,y)\hat{E}(x',y')|\Psi_s\rangle = \psi_s^*(x,y)\psi_s(x',y'), \qquad s=1,2. \tag{5.95}$$

Using this along with Eqs. (5.72)–(5.76) to evaluate Eq. (5.78) results in

$$\begin{aligned} P(1) &= \frac{\epsilon}{2}(1-\delta_x), \\ P(2) &= \frac{\epsilon}{4}(1+\delta_x-\delta_y-\delta), \\ P(3) &= \frac{\epsilon}{4}(1+\delta_x+\delta_y+\delta), \end{aligned} \tag{5.96}$$

and no-click probability $P(0) = 1-\epsilon$. Here, $\epsilon = \epsilon_1 + \epsilon_2$ is the total source intensity,

$$\delta_x = \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\,\psi^*(x,y)\psi(x-d_X,y), \tag{5.97}$$

$$\delta_y = \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\,\psi^*(x,y)\psi(x,y-d_Y), \tag{5.98}$$

and δ is defined Eq. (5.5).

5.4.2 Fisher information matrix for SLIVER

The Fisher information matrix $j^{(\mathrm{SLI})}$ for the separation vector $\eta = (d_X, d_Y)^\top$ using SLIVER is given by

$$j_{\mu\nu}^{(\mathrm{SLI})} = \sum_{r=1}^{3} P(r)\frac{\partial \ln P(r)}{\partial\eta_\mu}\frac{\partial \ln P(r)}{\partial\eta_\nu}, \quad \mu,\nu=1,2. \tag{5.99}$$

Using Eqs. (5.99) and (5.96), we evaluate the Fisher information matrix and it has elements

$$
\begin{aligned}
j_{11}^{(\mathrm{SLI})} &= \frac{\epsilon}{2}\left[\frac{1}{1-\delta_x}\left(\frac{\partial\delta_x}{\partial d_X}\right)^2 + \frac{1}{2(1+\delta_x-\delta_y-\delta)}\left(\frac{\partial\delta_x}{\partial d_X}-\frac{\partial\delta}{\partial d_X}\right)^2\right. \\
&\qquad \left. + \frac{1}{2(1+\delta_x+\delta_y+\delta)}\left(\frac{\partial\delta_x}{\partial d_X}+\frac{\partial\delta}{\partial d_X}\right)^2\right], \\
j_{22}^{(\mathrm{SLI})} &= \frac{\epsilon}{4}\left[\frac{1}{1+\delta_x-\delta_y-\delta}\left(\frac{\partial\delta_y}{\partial d_Y}+\frac{\partial\delta}{\partial d_Y}\right)^2\right. \\
&\qquad \left. + \frac{1}{1+\delta_x+\delta_y+\delta}\left(\frac{\partial\delta_y}{\partial d_Y}+\frac{\partial\delta}{\partial d_Y}\right)^2\right], \\
j_{12}^{(\mathrm{SLI})} &= j_{21}^{(\mathrm{SLI})} \\
&= \frac{\epsilon}{4}\left[\frac{1}{1+\delta_x+\delta_y+\delta}\left(\frac{\partial\delta_x}{\partial d_X}+\frac{\partial\delta}{\partial d_X}\right)\left(\frac{\partial\delta_y}{\partial d_Y}+\frac{\partial\delta}{\partial d_Y}\right)\right. \\
&\qquad \left. - \frac{1}{1+\delta_x-\delta_y-\delta}\left(\frac{\partial\delta_x}{\partial d_X}-\frac{\partial\delta}{\partial d_X}\right)\left(\frac{\partial\delta_y}{\partial d_Y}+\frac{\partial\delta}{\partial d_Y}\right)\right]. \qquad (5.100)
\end{aligned}
$$

We illustrate the above results for a circular Gaussian PSF as in Eq. (5.66). The PSF-dependent quantities appearing in the Fisher information are

$$
\begin{aligned}
\delta &= \delta_x\delta_y, \\
\delta_x &= \exp\left(-\frac{d_X^2}{8\sigma^2}\right), \qquad \delta_y = \exp\left(-\frac{d_Y^2}{8\sigma^2}\right), \\
\frac{\partial\delta}{\partial d_X} &= \delta_y\frac{\partial\delta_x}{\partial d_X}, \qquad \frac{\partial\delta}{\partial d_Y} = \delta_x\frac{\partial\delta_y}{\partial d_Y}, \\
\frac{\partial\delta_x}{\partial d_X} &= -\frac{d_X}{4\sigma^2}\exp\left(-\frac{d_X^2}{8\sigma^2}\right), \\
\frac{\partial\delta_y}{\partial d_Y} &= -\frac{d_Y}{4\sigma^2}\exp\left(-\frac{d_Y^2}{8\sigma^2}\right). \qquad (5.101)
\end{aligned}
$$

In this example, the Fisher information matrix $j^{(\mathrm{SLI})}$ has elements

$$
\begin{aligned}
j_{11}^{(\mathrm{SLI})} &= \frac{\epsilon}{1-\exp(-d_X^2/4\sigma^2)}\frac{d_X^2}{16\sigma^4}\exp\left(-\frac{d_X^2}{4\sigma^2}\right), \\
j_{22}^{(\mathrm{SLI})} &= \frac{\epsilon[1+\exp(-d_X^2/8\sigma^2)]}{2[1-\exp(-d_Y^2/4\sigma^2)]}\frac{d_Y^2}{16\sigma^4}\exp\left(-\frac{d_Y^2}{4\sigma^2}\right), \\
j_{12}^{(\mathrm{SLI})} &= j_{21}^{(\mathrm{SLI})} = 0. \qquad (5.102)
\end{aligned}
$$

As $d_X, d_Y \to 0$, the matrix elements approach

$$\begin{aligned} j_{11}^{(\mathrm{SLI})} &\to \epsilon \Delta k^2 = J_{33}, \\ j_{22}^{(\mathrm{SLI})} &\to \epsilon \Delta k^2 = J_{44}, \end{aligned} \tag{5.103}$$

where $\Delta k^2 = (4\sigma^2)^{-1}$. The Fisher information elements $j_{11}^{(\mathrm{SLI})}$ and $j_{22}^{(\mathrm{SLI})}$ of Eq. (5.102) are plotted as a function of separation parameters d_X and d_Y in Fig. 5.4. The total source strength $\epsilon = 2 \times 10^{-3}$ photons. The plots are normalized to $\epsilon \Delta k^2$, the values of J_{33} and J_{44}. We see that the maximum values of $j_{11}^{(\mathrm{SLI})}$ and $j_{22}^{(\mathrm{SLI})}$, attained at $d_X = d_Y = 0$, are equal to the value of QCR bound as shown in Eq. (5.103). Fig. 5.4(a) suggests that the Fisher information on d_X – $j_{11}^{(\mathrm{SLI})}$ – remains unchanged despite variation in d_Y, while the Fisher information on d_Y – $j_{22}^{(\mathrm{SLI})}$ – depends on values of both d_X and d_Y as depicted in Fig. 5.4(b). This asymmetry is a consequence of us estimating d_X in the first stage of the scheme and d_Y in the second.

A simpler non-cascaded version of the scheme may be envisaged in which the input field $\hat{E}(x, y)$ is split using a 50:50 beamsplitter, and the two outputs are used to separately estimate d_X and d_Y. Though it treats the separation components symmetrically, such a setup can only approach half of the QCR bound for each component due to the energy splitting. On the other hand, in the cascaded scheme given here, if $d_X \approx 0$, $\hat{E}_X(x, y) \approx 0$, and $\hat{E}_S(x, y) \approx \hat{E}(x, y)$, so that the first stage taps only a small fraction of the available energy (the energy loss is zero if $d_X = 0$). This enables d_Y to be estimated in the second stage with little loss of energy to the first stage, and allows approaching the QCR bound for sub-Rayleigh separations.

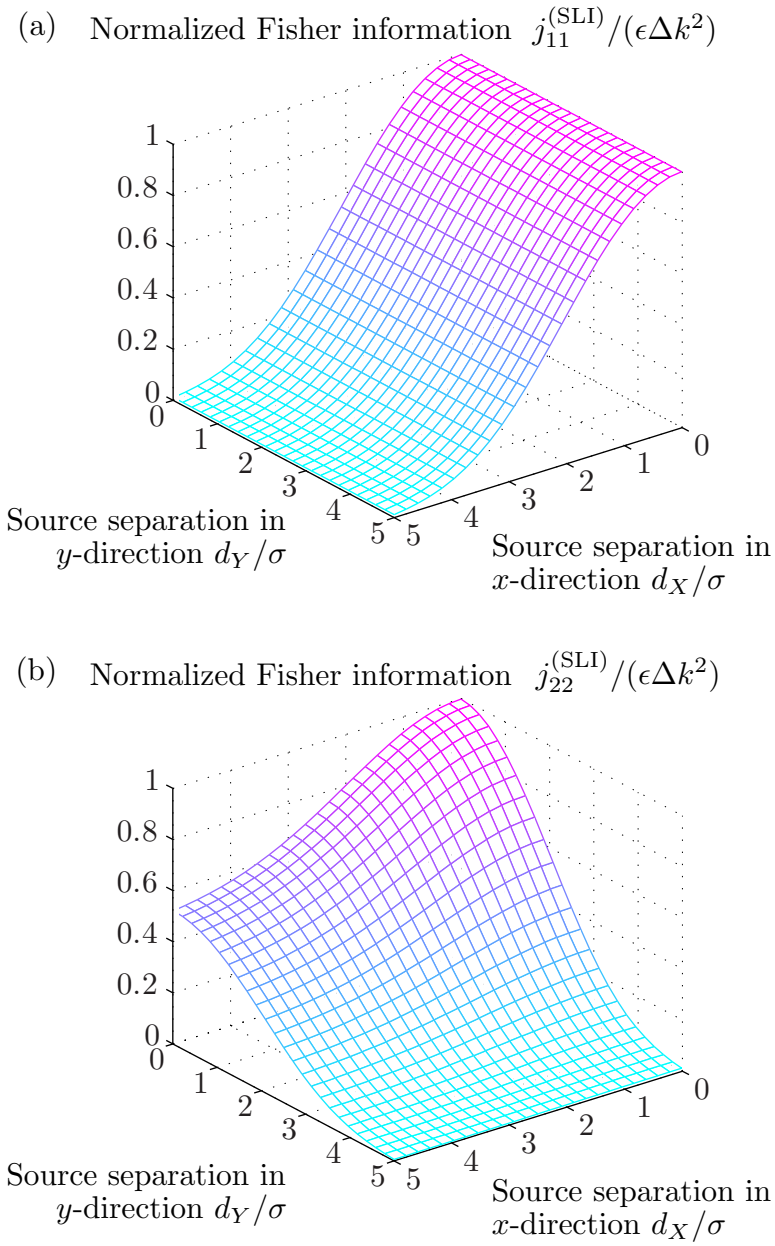


Figure 5.4: The (classical) Fisher information matrix $j^{(\mathrm{SLI})}$ for the SLIVER scheme as a function of the source separation (d_X, d_Y). (a) Fisher information for x-separation $j_{11}^{(\mathrm{SLI})}$. (b) Fisher information for y-separation $j_{22}^{(\mathrm{SLI})}$. The plots are normalized with respect to the value $\epsilon\Delta k^2$ of J_{33} and J_{44}. The quantum bound is attained at $d_X = d_Y = 0$ as illustrated in (a) and (b). The circular Gaussian PSF of Eq. (5.66) is assumed, with the total source strength $\epsilon = 2 \times 10^{-3}$ photons.

5.5 Two-dimensional Spatial-mode Demultiplexing (SPADE)

We now generalize the SPADE scheme of Ref. [101] to the estimation of the vector separation. In the derivation of the QFI matrix J in Sec. 5.3, we worked in the orthonormal basis $\{|e_n\rangle; n = 0, \ldots, 6\}$ given by Eqs. (5.19)–(5.25).

Now consider the discrete Hermite-Gaussian (HG) basis $\{|\phi_{qr}\rangle; q, r = 0, 1, \ldots\}$ for the one-photon subspace, where

$$|\phi_{qr}\rangle = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y\, \phi_{qr}(x, y)|x, y\rangle, \tag{5.104}$$

$$\begin{aligned}\phi_{qr}(x, y) = &\left(\frac{1}{2\pi\sigma_X\sigma_Y}\right)^{1/2} \frac{1}{\sqrt{2^{q+r}q!r!}} H_q\left(\frac{x}{\sqrt{2}\sigma_X}\right) \\ &\times H_r\left(\frac{y}{\sqrt{2}\sigma_Y}\right) \exp\left(-\frac{x^2}{4\sigma_X^2} - \frac{y^2}{4\sigma_Y^2}\right),\end{aligned} \tag{5.105}$$

H_q and H_r are the Hermite polynomials [107], and $q, r = 0, 1, \ldots$. Consider the POVM consisting of the projections

$$\hat{W}_0 = |\mathrm{vac}\rangle\langle\mathrm{vac}|, \tag{5.106}$$

$$\hat{W}_1(q, r) = |\phi_{qr}\rangle\langle\phi_{qr}|, \quad q, r = 0, 1, \ldots. \tag{5.107}$$

We further assume that the PSF is circular Gaussian as in Eq. (5.66). The image-plane field in Eq. (5.2) is scaled in one direction with a series of mirrors such that the PSF is now of the elliptical Gaussian form

$$\psi'(x, y) = \left(\frac{1}{2\pi\sigma_X\sigma_Y}\right)^{1/2} \exp\left(-\frac{x^2}{4\sigma_X^2} - \frac{y^2}{4\sigma_Y^2}\right) \tag{5.108}$$

$$= \phi_{00}(x, y), \tag{5.109}$$

where $\sigma_X = \sigma$, $\sigma_Y = s\sigma$ and s is the scaling factor. The above POVM, if performed on the resulting density operator ρ', has the outcome probabili-

ties

$$P_0 \equiv \mathrm{tr}(\hat{W}_0\hat{\rho}') = 1-\epsilon, \tag{5.110}$$

$$P_1(q,r) \equiv \mathrm{tr}[\hat{W}_1(q,r)\hat{\rho}'] \tag{5.111}$$

$$= \frac{\epsilon}{2}\left(|\langle\phi_{qr}|\psi_1'\rangle|^2 + |\langle\phi_{qr}|\psi_2'\rangle|^2\right). \tag{5.112}$$

If $(\bar{X},\bar{Y}) = (0,0)$, the wavefunctions become

$$\psi_1'(x,y) = \psi'\left(x+\frac{d_X}{2}, y+\frac{sd_Y}{2}\right), \tag{5.113}$$

$$\psi_2'(x,y) = \psi'\left(x-\frac{d_X}{2}, y-\frac{sd_Y}{2}\right), \tag{5.114}$$

and the overlaps in Eq. (5.112) are

$$|\langle\phi_{qr}|\psi_1'\rangle|^2 = \left|\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\,\phi_{qr}^*(x,y)\phi_{00}\left(x+\frac{d_X}{2}, y+\frac{sd_Y}{2}\right)\right|^2, \tag{5.115}$$

$$|\langle\phi_{qr}|\psi_2'\rangle|^2 = \left|\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}y\,\phi_{qr}^*(x,y)\phi_{00}\left(x-\frac{d_X}{2}, y-\frac{sd_Y}{2}\right)\right|^2. \tag{5.116}$$

Evaluate the integrals using properties of Hermite polynomials, we have

$$|\langle\phi_{qr}|\psi_1'\rangle|^2 = |\langle\phi_{qr}|\psi_2'\rangle|^2 = \exp(-Q-R)\frac{Q^qR^r}{q!r!}, \tag{5.117}$$

where

$$Q = \frac{d_X^2}{16\sigma^2}, \qquad R = \frac{d_Y^2}{16\sigma^2}. \tag{5.118}$$

As the two integrals are equal, the probability

$$P_1(q,r) = \epsilon\exp(-Q-R)\frac{Q^qR^r}{q!r!} \tag{5.119}$$

is valid even if the two sources have unequal intensities. The Fisher information matrix for the HG-basis measurement on $\eta = (d_X, d_Y)^\top$ can be

found using Eq. (2.60), with elements

$$j_{11}^{(\mathrm{HG})} = \sum_{q,r=0}^{\infty} P_1(q,r) \left[\frac{\partial}{\partial d_X} \ln P_1(q,r) \right]^2 \tag{5.120}$$

$$= \frac{\epsilon}{Q} \left(\frac{\partial Q}{\partial d_X} \right)^2 = \frac{\epsilon}{4\sigma^2} = J_{33}, \tag{5.121}$$

$$j_{22}^{(\mathrm{HG})} = \sum_{q,r=0}^{\infty} P_1(q,r) \left[\frac{\partial}{\partial d_Y} \ln P_1(q,r) \right]^2 \tag{5.122}$$

$$= \frac{\epsilon}{R} \left(\frac{\partial R}{\partial d_Y} \right)^2 = \frac{\epsilon}{4\sigma^2} = J_{44}, \tag{5.123}$$

$$j_{12}^{(\mathrm{HG})} = j_{21}^{(\mathrm{HG})} = 0, \tag{5.124}$$

which equals the QFI matrix given by Eq. (5.53). This proves that the above measurement scheme using the HG basis is optimal for a Gaussian PSF.

A quadratic-index fiber can support the HG mode profiles [108]. To measure in the HG basis, the optical field is coupled into an elliptical multi-mode fiber supporting the modes in Eq. (5.105). Choosing the scaling factor s carefully (for example, an irrational number), each mode will have a distinct propagation constant β_{qr} along propagation direction z. On the other hand, a cylindrical fiber will result in degenerate propagation constant for modes of same order $(q + r)$. Each mode in the elliptic fiber is then coupled to different single-mode waveguides of specific propagation constant via evanescent coupling as illustrated in Fig. 5.5. The phase matching condition ensures that only one mode from the elliptic fiber is coupled to each waveguide, which are then detected using individual photon counter in the far-field.

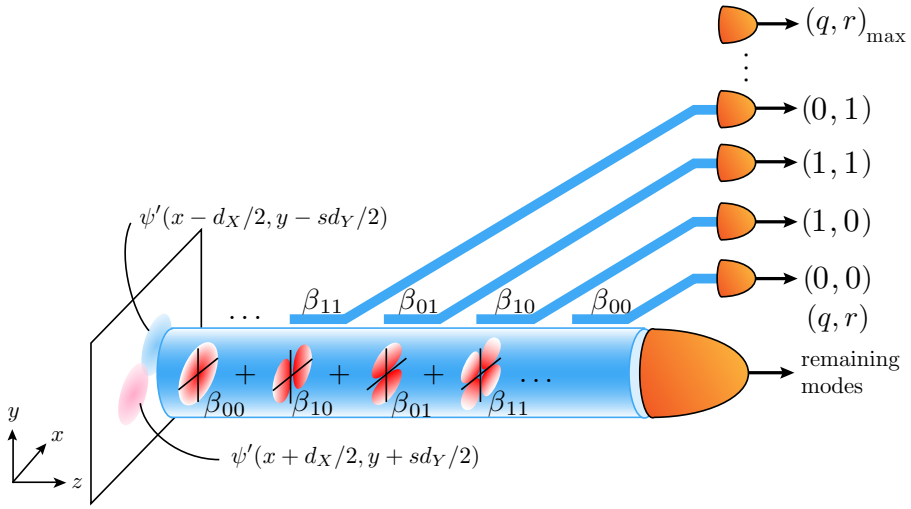


Figure 5.5: A proposed schematic of a multimode-fiber SPADE. The scaled field of Eq. (5.108) is coupled into an elliptical multimode fiber. With evanescent coupling, each mode is coupled to a single-mode waveguide of specific propagation constant and led to a photon counter. The photon counter at the end of the multimode fiber captures any remaining photon in the higher-order or leaky modes.

5.6 Monte-Carlo Analysis of SLIVER and SPADE

To demonstrate that the two schemes can perform as predicted by their CR bounds, we study Monte-Carlo simulations of the mean-square error (MSE) for SLIVER and SPADE. The circular Gaussian PSF of Eq. (5.66) is assumed. Each MSE is computed by averaging over 10^5 Monte-Carlo runs.

5.6.1 Monte-Carlo analysis of SLIVER

In M trials, consider direct detection of $\hat{E}_1(x,y)$, $\hat{E}_2(x,y)$ and $\hat{E}_3(x,y)$ using three on-off bucket detectors as in Fig. 5.3. The measurement record consists of the bitstrings $(g_1^{(1)}, g_2^{(1)}, \ldots, g_M^{(1)})$, $(g_1^{(2)}, g_2^{(2)}, \ldots, g_M^{(2)})$, and $(g_1^{(3)}, g_2^{(3)}, \ldots, g_M^{(3)})$, where $g_m^{(1)}$, $g_m^{(2)}$ and $g_m^{(3)}$ are zero (one) if the corresponding detector did not click (clicked) in the m-th trial.

The total numbers of clicks observed in the three detectors over M

measurements are, respectively,

$$G^{(1)} = \sum_{m=1}^{M} g_m^{(1)}, \quad G^{(2)} = \sum_{m=1}^{M} g_m^{(2)}, \quad G^{(3)} = \sum_{m=1}^{M} g_m^{(3)}, \tag{5.125}$$

and the total number of detected photons $L = G^{(1)} + G^{(2)} + G^{(3)}$. For estimation error analysis we can condition on L, which is obtained after an experiment, instead of the average photon number N; the QCR bounds for the estimation of d_X and d_Y become $4\sigma^2/L$. The ML estimators for d_X and d_Y can be shown to be

$$\begin{aligned}
\check{d}_X^{(\mathrm{SLI})} &= \begin{cases} 2\sigma\sqrt{-2\ln\left(1-\frac{2G^{(1)}}{L}\right)} & \text{if } \frac{2G^{(1)}}{L} < 1, \\ 2\sigma & \text{otherwise,} \end{cases} \\
\check{d}_Y^{(\mathrm{SLI})} &= \begin{cases} 2\sigma\sqrt{-2\ln\left(1-\frac{2G^{(2)}}{L-G^{(1)}}\right)} & \text{if } \frac{2G^{(2)}}{L-G^{(1)}} < 1, \\ 2\sigma & \text{otherwise.} \end{cases}
\end{aligned} \tag{5.126}$$

The second case for both $\check{d}_X^{(\mathrm{SLI})}$ and $\check{d}_Y^{(\mathrm{SLI})}$ is necessary because the logarithm function $\ln(z)$ in the equations for the estimators is undefined for $z \leq 0$. The estimators are set to an arbitrary value if the argument of the logarithm goes negative, which happens with a probability tending to zero as L increases.

Figures 5.6 and 5.7 show the simulated MSEs of the ML estimators in Eq. (5.126). The simulated MSEs are plotted relative to the minimum value of the CR bound for that L at $d_X = d_Y = 0$. Fig. 5.6 plots the MSE of estimator $\check{d}_X^{(\mathrm{SLI})}$ as a function of x-separation for $d_Y = 0$, the ML estimator beats the CR bound for $d_X/\sigma < 0.1$ due to the biasedness of the estimator. The MSE of estimator $\check{d}_Y^{(\mathrm{SLI})}$ as a function of y-separation for $d_X = 0$ is virtually identical to that of Fig. 5.6 and is not shown. This behavior is expected as $\hat{E}_S(x, y) = \hat{E}(x, y)$ if $d_X = 0$.

Fig. 5.7 explores the dependence of the MSE of the ML estimators on

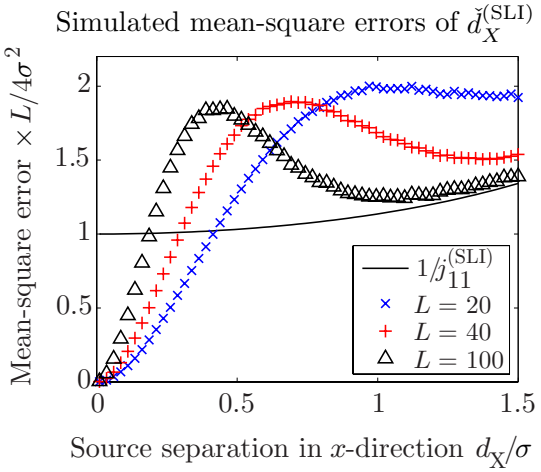


Figure 5.6: Simulated mean-square errors of SLIVER with ML estimator of Eq. (5.126). MSE of estimator $\check{d}_X^{(\mathrm{SLI})}$ as a function of the separation in x-direction for $L = 20, 40, 100$ photons and $d_Y = 0$.

the separations d_X and d_Y for $L = 100$ photons, with the corresponding CR bounds for the relevant separations are shown. Fig. 5.7(a) shows the simulated MSE of estimator $\check{d}_X^{(\mathrm{SLI})}$ as a function of d_X, for $d_Y = 0, 0.74\sigma$, and 1.5σ. We see that both the estimator and the CR bound show little dependence on the separation d_Y. Fig. 5.7(b) plots the simulated MSE of estimator $\check{d}_Y^{(\mathrm{SLI})}$ against d_Y for the case of $d_X = 0, 0.74\sigma$, and 1.5σ. As d_X increases, the CR bound increases along with the MSE of the estimator.

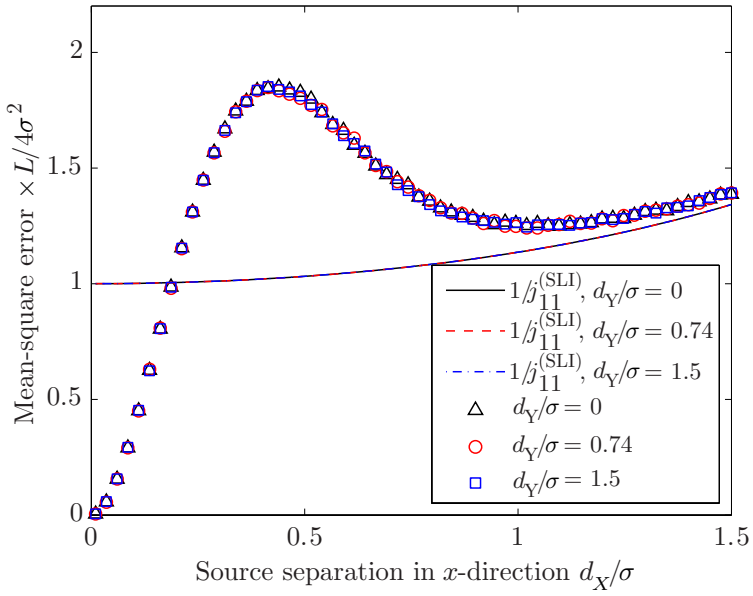
5.6.2 Monte-Carlo analysis of SPADE

After M trials of the SPADE measurement of Sec. 5.5, the total number L of detected photons is known, and for estimation error analysis we can condition on L instead of the average photon number N. Then the QCR bounds for estimation of d_X and d_Y become $4\sigma^2/L$.

Each photon collected by the imaging system triggers precisely one photon detector indexed by the HG mode index (q, r). Let the mode excited by the l-th photon be denoted (q_l, r_l), for $1 \leq l \leq L$. The maximum likelihood estimators for d_X and d_Y can be shown to be

$$\check{d}_X^{(\mathrm{SPA})} = 4\sigma\sqrt{\frac{H_X}{L}}, \qquad \check{d}_Y^{(\mathrm{SPA})} = 4\sigma\sqrt{\frac{H_Y}{L}}. \tag{5.127}$$

(a) Simulated mean-square errors of $\check{d}_X^{(\mathrm{SLI})}$ as a function of d_X for $L = 100$



(b) Simulated mean-square errors of $\check{d}_Y^{(\mathrm{SLI})}$ as a function of d_Y for $L = 100$

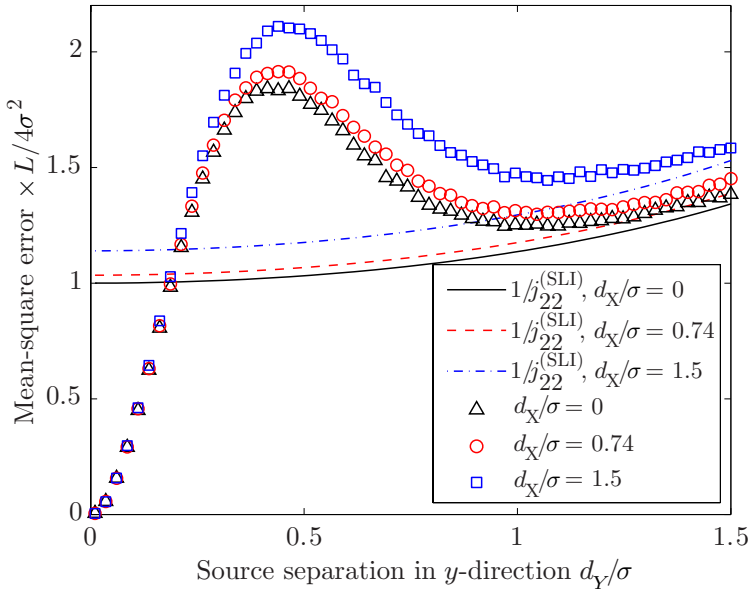


Figure 5.7: Simulated mean-square errors of SLIVER with ML estimator of Eq. (5.126) for different values of d_X and d_Y for $L = 100$ photons. The corresponding CR bounds are included in the plots for comparison. (a) MSE of estimator $\check{d}_X^{(\mathrm{SLI})}$ as a function of the separation in x-direction for $d_Y/\sigma = 0, 0.74, 1.5$. (b) MSE of estimator $\check{d}_Y^{(\mathrm{SLI})}$ as a function of the separation in y-direction for $d_X/\sigma = 0, 0.74, 1.5$.

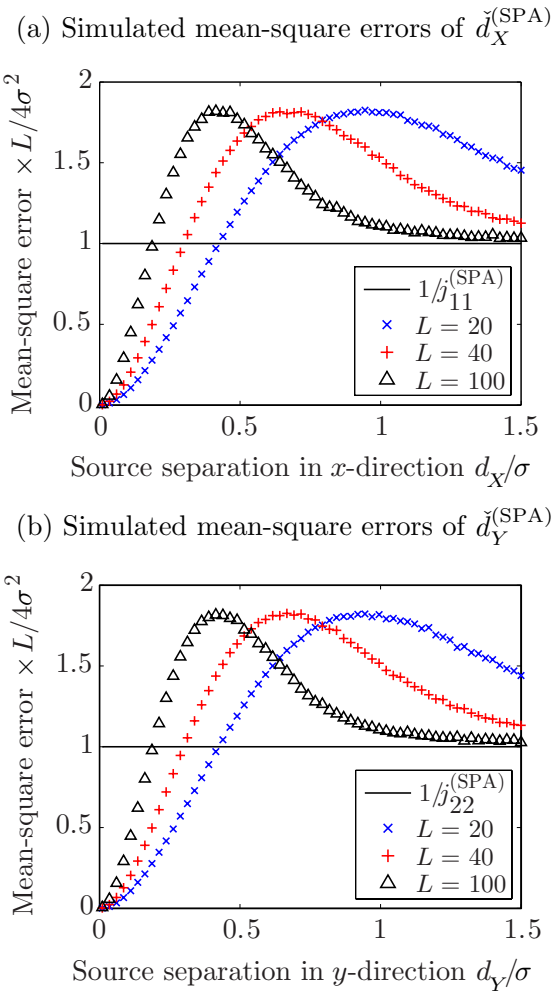


Figure 5.8: Simulated mean-square errors of SPADE with ML estimator of Eq. (5.127). (a) MSE of estimator $\check{d}_X^{(\mathrm{SPA})}$ as a function of the separation in x-direction for $L = 20, 40, 100$ photons. (b) MSE of estimator $\check{d}_Y^{(\mathrm{SPA})}$ as a function of the separation in y-direction for $L = 20, 40, 100$ photons.

where $H_X = \sum_{l=0}^{L} q_l$ and $H_Y = \sum_{l=0}^{L} r_l$. Fig. 5.8(a) and (b) show the results for the MSE of the ML estimators for d_X and d_Y in Eq. (5.127) for $L = 20, 40, 100$ photons. The ML estimators beat the CR bounds in the estimation of d_X and d_Y for small separations. The errors remain less than twice the CR bounds for any separations.

5.7 Discussion

In this study, we have calculated the QCR bound for locating two weak incoherent optical point sources on a two-dimensional plane using imaging

systems with inversion-symmetric point-spread functions. The key result is that, in stark contrast to direct imaging [99, 101], the bounds on the mean-square error of the x- and y-separations are independent of the vector separation between the two sources.

We have also proposed and analyzed two measurement schemes – the extended SLIVER and SPADE schemes – for simultaneously estimating the components of the separation, whose classical CR bounds approach the quantum bounds for sub-Rayleigh separations (SLIVER) or all separations if the PSF is Gaussian (SPADE). Monte-Carlo simulations show that the two schemes have MSEs no larger than twice predicted by the quantum limits for small source separation.

The extended SLIVER scheme given here does not employ an image inversion device (see, e.g., Refs. [109, 110]), which was required in the scheme of Ref. [103]. Thus, each interferometer stage may be technically simpler to implement than the original SLIVER. However, the original scheme is tailored to estimation of the magnitude $d = \sqrt{d_X^2 + d_Y^2}$ of the separation and is likely to be superior for that purpose than the two-stage scheme, as supported by the simulations in Sec. 5.6.

In both measurement schemes, we have assumed that the centroid position is known. If that knowledge is unavailable, a portion of the light can be used for image-plane photon counting to determine the centroid position before performing either of the schemes as detailed in Ref. [101].

Our analysis here can be extended in various directions. On the theoretical side, our quantum Fisher information calculations can be extended to sources of arbitrary strength employing the Gaussian-state model [111, 112]. The study of sub-Rayleigh imaging can also be generalized to sources emitting light in more general quantum states [112]. In principle, it can also be extended to multiple sources, although finding near-optimal measurement schemes is likely to be challenging. On the practical side, it is

important to study the performance of SLIVER and SPADE in non-ideal cases, e.g., if the centroid is not aligned to the optical axis in both schemes, or in the presence of non-unity coupling efficiency in SPADE, and for unequal detection efficiencies of the detectors.

Chapter 6

Concluding Remarks

6.1 Conclusion

In this thesis, we have obtained a series of results on the application of parameter estimation theory to optical spectroscopy and imaging problems. In Chapter 3, we have analyzed in detail the problem of parameter estimation from a noisy optomechanical system. We have found analytically the CR lower bound on the estimation errors in the long-time limit, and compared it with the errors of averaging, radiometer, and EM estimators in the estimation of the force noise power. When applied to experimental data, we have found that EM algorithm has the lowest error and follows the CR bound most closely. We hope that our analytic results will be valuable to optomechanical experiment design, and EM algorithm will be useful to system identification problems.

In Chapter 4, we have found that a property of the QFI matrix, extended convexity, is useful in deriving analytical expressions for the quantum limit to the estimation error for dynamical systems. The result is a simple expression in terms of basic PSDs. We have applied the analysis to a continuous optical phase estimation experiment and demonstrated that homodyne detection performed in the experiment, when the SNR is high,

is close to our quantum limit. We have further proposed a spectral photon counting method that can attain the quantum limit when the SNR is low.

In the last chapter, we have shown that ideas from parameter estimation theory can be extended to the resolution problem in optical imaging. We have obtained the QCR bound for the estimation errors of the Cartesian components of the centroid and separation of two incoherent point sources. We have found that the quantum limit is independent of separation, in contrast with conventional wisdom. We have proposed two measurement schemes that asymptotically attain the quantum bound for both components of the separation, assuming a known centroid position.

6.2 Future Works

During the investigation of parameter estimation of optomechanical system in Chapter 3, we have assumed that the parameters are constant in time. For applications of optomechanics in sensing and control, there is a need to develop estimation theory for time-varying parameters, and Bayesian estimators may be a suitable candidate to explore. Recently, optomechanical devices have found exciting application in many fields in physics and engineering. Given demonstrated success of our estimation of the mechanical resonance frequency, dissipation rate and effective mass, we could apply our statistical techniques to improve optomechanical design in applications, such as ultra-sensitive, high bandwidth accelerometers and force sensors, wavelength converters and tunable optical filters.

In the study of parameter estimation error for dynamical systems, we have modeled our bath classically. A possible future work would be to account for nonclassical baths and make our theory applicable for a wider range of applications. Another possible direction is to study the proposed spectral photon counting method as it offers a large potential improvement

compared to homodyne detection in the low SNR regime. More analyses are required to study its performance in actual experimental settings. One of the key limitations to its experimental adaption will be the implementation of counting in each spectral mode. The experimental performance will be dependent on the spectral resolution of chosen dispersive optical element and losses in every spectral mode. As analytical results of our theory are in terms of basic PSDs, they can be applied to optics and optomechanics experiments, with applications in optical sensing and gravitational-wave astronomy.

Although extensive study has been done to understand the fundamental limit in resolving two incoherent point sources, this study has assumed a weak-source condition. While such approximation is valid, e.g., for astronomical observations at optical and infrared wavelengths, there is a need to derive the quantum limits to the source separation problem for intense sources, e.g., in the microwave and far-infrared regimes, and for laser sources. As the quantum theory for resolution is in early stage, for future applications in imaging, we need to build on the success of this study and extend its scope. One important research direction would be the extension of this work to resolving $n \geq 3$ optical point sources. There are a few interesting cases, for example, the clustering of optical point sources into different regions and uneven photon emission of those point sources. Also, we have not considered the effect of noises on the quantum Cramér-Rao bound for the estimation of the separation between two sources, a more detailed analysis of such practical but important concern is needed. There are two possible sources of noise, the first kind is background noise and the second kind is from noisy imaging system. For experimental works, there are currently a few proof-of-concept demonstrations. As we have started to understand the origin of the additional Fisher information, we can better design measurement schemes to extract the whole information from the

light. The designs should take into account of implementation with only linear-optical elements and limit introduction of noise during measurement. I would foresee that it has applications in astronomy, notably binary-star observation, and single-molecule fluorescence microscopy.

Bibliography

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part. I* (John Wiley & Sons, New York, 2001).

[2] B. C. Levy, *Principles of Signal Detection and Parameter Estimation* (Springer, 2008).

[3] M. Aspelmeyer, T. J. Kippenberg, and F. Marquardt, "Cavity optomechanics," Reviews of Modern Physics **86**, 1391 (2014).

[4] D. Griffiths, *Introduction to Quantum Mechanics*, Pearson international edition (Pearson Prentice Hall, 2005).

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge university press, 2010).

[6] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics* (Addison-Wesley, 1994).

[7] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering* (Pearson/Prentice Hall, 2008).

[8] E. Jaynes and G. Bretthorst, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).

[9] H. Wiseman and G. Milburn, *Quantum Measurement and Control* (Cambridge University Press, 2010).

[10] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society. Series B (Methodological) **39**, 1 (1977).

[12] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications* (Springer, 2000).

[13] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part. III* (John Wiley & Sons, 2001).

[14] P. Whittle, "The analysis of multiple stationary time series," Journal of the Royal Statistical Society. Series B (Methodological) **15**, 125 (1953).

[15] A. S. Holevo, *Statistical Structure of Quantum Theory* (Springer Berlin Heidelberg, 2011).

[16] M. G. A. Paris, "Quantum estimation for quantum technology," International Journal of Quantum Information **7**, 125 (2009).

[17] S. L. Braunstein and C. M. Caves, "Statistical distance and the geometry of quantum states," Phys. Rev. Lett. **72**, 3439 (1994).

[18] M. Hayashi, *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (World Scientific, 2005).

[19] A. Fujiwara, "Strong consistency and asymptotic efficiency for adaptive quantum estimation problems," Journal of Physics A: Mathematical and General **39**, 12489 (2006).

[20] H. Yuen and M. Lax, "Multiple-parameter quantum estimation and measurement of nonselfadjoint observables," IEEE Transactions on Information Theory **19**, 740 (2006).

[21] E. Gavartin, P. Verlot, and T. J. Kippenberg, "A hybrid on-chip optomechanical transducer for ultrasensitive force measurements," Nature Nanotechnology **7**, 509 (2012).

[22] G. I. Harris, D. L. McAuslan, T. M. Stace, A. C. Doherty, and W. P. Bowen, "Minimum requirements for feedback enhanced force sensing," Phys. Rev. Lett. **111**, 103603 (2013).

[23] S. Z. Ang, G. I. Harris, W. P. Bowen, and M. Tsang, "Optomechanical parameter estimation," New Journal of Physics **15**, 103028 (2013).

[24] A. N. Cleland and M. L. Roukes, "A nanometre-scale mechanical electrometer," Nature **392**, 160 (1998).

[25] D. Rugar, R. Budakian, H. J. Mamin, and B. W. Chui, "Single spin detection by magnetic resonance force microscopy," Nature **430**, 329 (2004).

[26] K. Jensen, K. Kim, and A. Zettl, "An atomic-resolution nanomechanical mass sensor," Nature Nanotechnology **3**, 533 (2008).

[27] A. G. Krause, M. Winger, T. D. Blasius, Q. Lin, and O. Painter, "A high-resolution microchip optomechanical accelerometer," Nature Photonics **6**, 768 (2012).

[28] S. Forstner, S. Prams, J. Knittel, E. D. van Ooijen, J. D. Swaim, G. I. Harris, A. Szorkovszky, W. P. Bowen, and H. Rubinsztein-Dunlop, "Cavity optomechanical magnetometer," Phys. Rev. Lett. **108**, 120801 (2012).

[29] T. P. Purdy, R. W. Peterson, and C. A. Regal, "Observation of radiation pressure shot noise on a macroscopic object," Science **339**, 801 (2013).

[30] E. Verhagen, S. Deléglise, S. Weis, A. Schliesser, and T. J. Kippenberg, "Quantum-coherent coupling of a mechanical oscillator to an optical cavity mode," Nature **482**, 63 (2012).

[31] S. Mancini, D. Vitali, and P. Tombesi, "Scheme for teleportation of quantum states onto a mechanical resonator," Phys. Rev. Lett. **90**, 137901 (2003).

[32] D. W. C. Brooks, T. Botter, S. Schreppler, T. P. Purdy, N. Brahms, and D. M. Stamper-Kurn, "Non-classical light generated by quantum-noise-driven cavity optomechanics," Nature **488**, 476 (2012).

[33] T. P. Purdy, P.-L. Yu, R. W. Peterson, N. S. Kampel, and C. A. Regal, "Strong optomechanical squeezing of light," Phys. Rev. X **3**, 031012 (2013).

[34] M. R. Vanner, I. Pikovski, and M. S. Kim, "Towards optomechanical quantum state reconstruction of mechanical motion," Annalen der Physik **527**, 15 (2015).

[35] K. H. Lee, T. G. McRae, G. I. Harris, J. Knittel, and W. P. Bowen, "Cooling and control of a cavity optoelectromechanical system," Phys. Rev. Lett. **104**, 123604 (2010).

[36] A. Schliesser, G. Anetsberger, R. Rivière, O. Arcizet, and T. J. Kippenberg, "High-sensitivity monitoring of micromechanical vibration using optical whispering gallery mode resonators," New Journal of Physics **10**, 095015 (2008).

[37] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE Transactions on Communication Technology **15**, 52 (1967).

[38] D. Simon, *Optimal State Estimation* (John Wiley & Sons, Inc., 2006).

[39] M. Guta and N. Yamamoto, "System identification for passive linear quantum systems," IEEE Transactions on Automatic Control **61**, 921 (2016).

[40] V. Petersen and K. Mølmer, "Estimation of fluctuating magnetic fields by an atomic magnetometer," Phys. Rev. A **74**, 043802 (2006).

[41] M. Tsang, "Time-symmetric quantum theory of smoothing," Phys. Rev. Lett. **102**, 250403 (2009).

[42] M. Tsang, "Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing," Phys. Rev. A **80**, 033840 (2009).

[43] S. Gammelmark, B. Julsgaard, and K. Mølmer, "Past quantum states of a monitored system," Phys. Rev. Lett. **111**, 160401 (2013).

[44] M. Ting, A. O. Hero, D. Rugar, C.-Y. Yip, and J. A. Fessler, "Near-optimal signal detection for finite-state Markov signals with application to magnetic resonance force microscopy," IEEE Transactions on Signal Processing **54**, 2049 (2006).

[45] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software* (Wiley, 2004).

[46] B. R. S. Arulampalam, *Beyond the Kalman Filter: Particle Filters for Tracking Applications* (Artech House, 2004).

[47] Y. Chen, "Macroscopic quantum mechanics: theory and experimental concepts of optomechanics," Journal of Physics B: Atomic, Molecular and Optical Physics **46**, 104001 (2013).

[48] T. A. Wheatley, D. W. Berry, H. Yonezawa, D. Nakane, H. Arao, D. T. Pope, T. C. Ralph, H. M. Wiseman, A. Furusawa, and

E. H. Huntington, "Adaptive optical phase estimation using time-symmetric quantum smoothing," Phys. Rev. Lett. **104**, 093601 (2010).

[49] H. Yonezawa, D. Nakane, T. A. Wheatley, K. Iwasawa, S. Takeda, H. Arao, K. Ohki, K. Tsumura, D. W. Berry, T. C. Ralph, H. M. Wiseman, E. H. Huntington, and A. Furusawa, "Quantum-enhanced optical-phase tracking," Science **337**, 1514 (2012).

[50] K. Iwasawa, K. Makino, H. Yonezawa, M. Tsang, A. Davidovic, E. Huntington, and A. Furusawa, "Quantum-limited mirror-motion estimation," Phys. Rev. Lett. **111**, 163602 (2013).

[51] J. K. Stockton, J. M. Geremia, A. C. Doherty, and H. Mabuchi, "Robust quantum parameter estimation: Coherent magnetometry with feedback," Phys. Rev. A **69**, 032109 (2004).

[52] M. R. James, H. I. Nurdin, and I. R. Petersen, "H^∞ control of linear quantum stochastic systems," IEEE Transactions on Automatic Control **53**, 1787 (2008).

[53] M. Tsang, H. M. Wiseman, and C. M. Caves, "Fundamental quantum limit to waveform estimation," Phys. Rev. Lett. **106**, 090401 (2011).

[54] M. Tsang, "Quantum metrology with open dynamical systems," New Journal of Physics **15**, 073005 (2013).

[55] M. Bagheri, M. Poot, M. Li, W. P. H. Pernice, and H. X. Tang, "Dynamic manipulation of nanomechanical resonators in the high-amplitude regime and non-volatile mechanical memory operation," Nature Nanotechnology **6**, 726 (2011).

[56] P. Rabl, S. J. Kolkowitz, F. H. L. Koppens, J. G. E. Harris, P. Zoller, and M. D. Lukin, "A quantum spin transducer based on nanoelectromechanical resonator arrays," Nature Physics **6**, 602 (2010).

[57] M. J. Griffin, "Bolometers for far-infrared and submillimetre astronomy," Nucl. Instrum. Meth. A **444**, 397 (2000).

[58] A. Alessandrello, V. Bashkirov, C. Brofferio, C. Bucci, D. V. Camin, O. Cremonesi, E. Fiorini, G. Gervasio, A. Giuliani, A. Nucciotti, M. Pavan, G. Pessina, E. Previtali, and L. Zanotti, "A scintillating bolometer for experiments on double beta decay," Phys Lett B **420**, 109 (1998).

[59] S. Ng, S. Z. Ang, T. A. Wheatley, H. Yonezawa, A. Furusawa, E. H. Huntington, and M. Tsang, "Spectrum analysis with quantum dynamical systems," Phys. Rev. A **93**, 042121 (2016).

[60] V. B. Braginsky, F. Y. Khalili, and K. S. Thorne, *Quantum Measurement* (Cambridge University Press, 1995).

[61] B. P. Abbott *et al.*, "An upper limit on the stochastic gravitational-wave background of cosmological origin," Nature **460**, 990 (2009).

[62] S. Nimmrichter, K. Hornberger, and K. Hammerer, "Optomechanical sensing of spontaneous wave-function collapse," Phys. Rev. Lett. **113**, 020405 (2014).

[63] L. Diósi, "Testing spontaneous wave-function collapse models on classical mechanical oscillators," Phys. Rev. Lett. **114**, 050403 (2015).

[64] T. M. Stace, "Quantum limits of thermometry," Phys. Rev. A **82**, 011611 (2010).

[65] U. Marzolino and D. Braun, "Precision measurements of temperature

and chemical potential of quantum gases," Phys. Rev. A **88**, 063609 (2013).

[66] M. Jarzyna and M. Zwierz, "Quantum interferometric measurements of temperature," Phys. Rev. A **92**, 032112 (2015).

[67] L. A. Correa, M. Mehboudi, G. Adesso, and A. Sanpera, "Individual quantum probes for optimal thermometry," Phys. Rev. Lett. **114**, 220405 (2015).

[68] R. Nair and M. Tsang, "Quantum optimality of photon counting for temperature measurement of thermal astronomical sources," The Astrophysical Journal **808**, 125 (2015).

[69] M. Tsang and R. Nair, "Fundamental quantum limits to waveform detection," Phys. Rev. A **86**, 042115 (2012).

[70] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, "Robust online hamiltonian learning," New Journal of Physics **14**, 103013 (2012).

[71] M. Hayashi, *Quantum Information: An Introduction* (Springer Berlin Heidelberg, 2006).

[72] M. Brunelli, S. Olivares, and M. G. A. Paris, "Qubit thermometry for micromechanical resonators," Phys. Rev. A **84**, 032105 (2011).

[73] C. Benedetti, F. Buscemi, P. Bordone, and M. G. A. Paris, "Quantum probes for the spectral properties of a classical environment," Phys. Rev. A **89**, 032114 (2014).

[74] S. Gammelmark and K. Mølmer, "Fisher information and the quantum Cramér-Rao sensitivity limit of continuous measurements," Phys. Rev. Lett. **112**, 170401 (2014).

[75] S. Alipour and A. T. Rezakhani, "Extended convexity of quantum Fisher information in quantum metrology," Phys. Rev. A **91**, 042104 (2015).

[76] O. E. Barndorff-Nielsen and R. D. Gill, "Fisher information in quantum statistics," Journal of Physics A: Mathematical and General **33**, 4481 (2000).

[77] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, "The elusive Heisenberg limit in quantum-enhanced metrology," Nature Communications **3**, 1063 (2012).

[78] A. De Pasquale, D. Rossini, P. Facchi, and V. Giovannetti, "Quantum parameter estimation affected by unitary disturbance," Phys. Rev. A **88**, 052117 (2013).

[79] D. F. Walls and G. J. Milburn, *Quantum Optics* (Springer Berlin Heidelberg, 2008).

[80] J. H. Shapiro, "Quantum measurement eigenkets for continuous-time direct detection," Quantum and Semiclassical Optics: Journal of the European Optical Society Part B **10**, 567 (1998).

[81] D. Brady, *Optical Imaging and Spectroscopy* (Wiley, 2009).

[82] S. T. Chu, B. E. Little, W. Pan, T. Kaneko, S. Sato, and Y. Kokubun, "An eight-channel add-drop filter using vertically coupled microring resonators over a cross grid," IEEE Photonics Technology Letters **11**, 691 (1999).

[83] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Observation of gravitational waves from a binary black hole merger," Phys. Rev. Lett. **116**, 061102 (2016).

[84] M. Tsang, "Quantum transition-edge detectors," Phys. Rev. A **88**, 021801 (2013).

[85] B. Escher, R. de Matos Filho, and L. Davidovich, "General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology," Nature Physics **7**, 406 (2011).

[86] M. Tsang, "Ziv-Zakai error bounds for quantum parameter estimation," Phys. Rev. Lett. **108**, 230401 (2012).

[87] D. W. Berry, M. Tsang, M. J. W. Hall, and H. M. Wiseman, "Quantum Bell-Ziv-Zakai bounds and Heisenberg limits for waveform estimation," Phys. Rev. X **5**, 031018 (2015).

[88] L. T. Hall, J. H. Cole, C. D. Hill, and L. C. L. Hollenberg, "Sensing of fluctuating nanoscale magnetic fields using nitrogen-vacancy centers in diamond," Phys. Rev. Lett. **103**, 220802 (2009).

[89] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. F. Matthews, T. Hashimoto, J. L. O'Brien, and A. Laing, "Universal linear optics," Science **349**, 711 (2015).

[90] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "An optimal design for universal multiport interferometers," arXiv:1603.08788 [physics.optics] .

[91] L. Rayleigh, "XXXI. Investigations in optics, with special reference to the spectroscope," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **8**, 261 (1879).

[92] E. Betzig, "Proposed method for molecular optical imaging," Opt. Lett. **20**, 237 (1995).

[93] W. E. Moerner and L. Kador, "Optical detection and spectroscopy of single molecules in a solid," Phys. Rev. Lett. **62**, 2535 (1989).

[94] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," Opt. Lett. **19**, 780 (1994).

[95] S. Weisenburger and V. Sandoghdar, "Light microscopy: an ongoing contemporary revolution," Contemporary Physics **56**, 123 (2015).

[96] S. Z. Ang, R. Nair, and M. Tsang, "Quantum limit for two-dimensional resolution of two incoherent optical point sources," arXiv:1606.00603 [quant-ph] .

[97] E. Bettens, D. V. Dyck, A. den Dekker, J. Sijbers, and A. van den Bos, "Model-based two-object resolution from observations having counting statistics," Ultramicroscopy **77**, 37 (1999).

[98] S. V. Aert, A. den Dekker, D. V. Dyck, and A. van den Bos, "High-resolution electron microscopy and electron tomography: resolution versus precision," Journal of Structural Biology **138**, 21 (2002).

[99] S. Ram, E. S. Ward, and R. J. Ober, "Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy," Proceedings of the National Academy of Sciences of the United States of America **103**, 4457 (2006).

[100] M. Tsang, "Quantum limits to optical point-source localization," Optica **2**, 646 (2015).

[101] M. Tsang, R. Nair, and X.-M. Lu, "Quantum theory of superresolution for two incoherent optical point sources," Phys. Rev. X **6**, 031033 (2016).

[102] M. Tsang, R. Nair, and X.-M. Lu, "Semiclassical theory of superresolution for two incoherent optical point sources," arXiv:1602.04655 [quant-ph] .

[103] R. Nair and M. Tsang, "Interferometric superlocalization of two incoherent optical point sources," Opt. Express **24**, 3684 (2016).

[104] J. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts & Company Publishers, 2005).

[105] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, 1995).

[106] J. Goodman, *Statistical Optics*, Wiley Series in Pure and Applied Optics (Wiley, 2000).

[107] A. Yariv, *Quantum Electronics* (Wiley, New York, 1989).

[108] K. Zhang and D. Li, *Electromagnetic Theory for Microwaves and Optoelectronics* (Springer Berlin Heidelberg, 2013).

[109] K. Wicker, S. Sindbert, and R. Heintzmann, "Characterisation of a resolution enhancing image inversion interferometer," Opt. Express **17**, 15491 (2009).

[110] D. Weigel, H. Babovsky, A. Kiessling, and R. Kowarschik, "Investigation of the resolution ability of an image inversion interferometer," Optics Communications **284**, 2273 (2011).

[111] R. Nair and M. Tsang, "Far-field superresolution of thermal electromagnetic sources at the quantum limit," Phys. Rev. Lett. **117**, 190801 (2016).

[112] C. Lupo and S. Pirandola, "Ultimate precision bound of quantum and subwavelength imaging," Phys. Rev. Lett. **117**, 190802 (2016).

[113] C. Hettich, C. Schmitt, J. Zitzmann, S. Kühn, I. Gerhardt, and V. Sandoghdar, "Nanometer resolution and coherent optical dipole coupling of two individual molecules," Science **298**, 385 (2002).

[114] N. Fang, H. Lee, C. Sun, and X. Zhang, "Sub–diffraction-limited optical imaging with a silver superlens," Science **308**, 534 (2005).

[115] J. Gelles, B. J. Schnapp, and M. P. Sheetz, "Tracking kinesin-driven movements with nanometre-scale precision," Nature **331**, 450 (1988).

[116] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," Science **313**, 1642 (2006).

[117] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," Nature Methods **3**, 793 (2006).

[118] Z. S. Tang, K. Durak, and A. Ling, "Fault-tolerant and finite-error localization for point emitters within the diffraction limit," Opt. Express **24**, 22004 (2016).

[119] F. Yang, A. Tashchilina, E. S. Moiseev, C. Simon, and A. I. Lvovsky, "Far-field linear optical superresolution via heterodyne detection in a higher-order local oscillator mode," Optica **3**, 1148 (2016).

[120] W.-K. Tham, H. Ferretti, and A. M. Steinberg, "Beating Rayleigh's curse by imaging using phase information," arXiv:1606.02666 [physics.optics] .

[121] M. Paúr, B. Stoklasa, Z. Hradil, L. L. Sánchez-Soto, and J. Rehacek, "Achieving the ultimate optical resolution," Optica **3**, 1144 (2016).

Appendix A

Super-resolution Microscopy

We have studied the quantum limit for two-dimensional resolution of two incoherent optical point sources and found that the limit is well below conventional Rayleigh resolution limit in Chapter 5. It seems fair that we discuss the recent advances in super-resolution techniques in optical microscopy, marked by the award of the 2014 Nobel Prize in Chemistry to Eric Betzig, Stefan Hell and William E. Moerner 'for the development of super-resolved fluorescence microscopy'. In general, these super-resolution microscopy techniques circumvent the Rayleigh resolution limit of optical microscopy, and do not challenge the limit directly. We will explain this in the next few paragraphs.

Techniques such as scanning near-field optical microscope (SNOM) [113] and hyperlens [114] realized with multilayer metamaterial are the examples of near-field super-resolution microscopy. SNOM works by raster scanning the image surface at a distance much smaller than optical wavelength and should be considered as a type of scanning probe microscopy as its resolution is limited by the probe/aperture size. Hyperlens aims to project the near field into the far field, hence both propagating and evanescent fields could be image. Currently, hyperlens is implemented with metamaterial which result in high loss. Another disadvantage of hyperlens is the place-

ment of detector close to the sample, which is not always possible in every application. Our proposal has the advantage that only far-field information is needed.

For far-field super-resolution microscopy, we have stimulated emission depletion (STED) microscopy [94] which depletes fluorescent molecules in specific regions while keeping a sub-Rayleigh-size center spot active. Image forming is again by raster scan technique which is time consuming.

Single-molecule based fluorescent microscopy is another family of far field super-resolution microscopy. Given that a single fluorescent molecule could be detected [93] and with high localization [115], Betzig proposed to resolve two closely-spaced molecules if they could be activated individually [92]. The idea was extended into large number of fluorophores by stochastically activate sparse subset of fluorophores, such that the PSFs of activated fluorophores do not overlap. This method is known as photoactivated localization microscopy (PALM) [116] or stochastic image reconstruction microscopy (STORM) [117], and it sidesteps Rayleigh criterion by not imaging point sources with overlapping PSFs at the same moment.

To challenge Rayleigh criterion, an imaging technique should resolve point sources with overlapping PSFs and separation less than full width at half maximum of the PSF. Although the above techniques are ingenious, they require careful control of the source and do not challenging Rayleigh's criterion fundamentally. In comparison, our study focuses on the resolution of two optical point sources of sub-Rayleigh separations which the current state-of-the-art optical imaging techniques could not resolve and shows that the separation could be estimated with constant error. In Appendix B.2, we discuss recent experimental demonstrations for our proposed imaging methods.

Appendix B

Recent Advances in Quantum Theory of Two-source Localization

In Chapter 5 of this thesis, we have made the assumption of weak-source model, given by Eq. (5.2), such that the probability of both sources emitting one photon is negligible for the two incoherent point sources localization problem. While this thesis remains focus on theoretical treatment based on the assumption, we would like to discuss about a recent theoretical paper [111] on the estimation of separation between two thermal point sources and experimental demonstrations reported in the past few months.

B.1 Thermal Point Sources

Recently, the incoherent two-source resolution problem was extended for estimating the one-dimensional separation of two thermal point sources of arbitrary strength [111]. The model considered consists of two incoherent thermal point sources imaged by a spatially-invariant unit-magnification imaging system. The QCR bound for estimating the separation d of the two sources was derived. Assuming the PSF $\psi(x)$ of the imaging system is

an even function, the QFI

$$J_d = N\Delta k^2 - \gamma^2 \frac{N^2(N+2)}{(N+2)^2 - \delta^2 N^2} \tag{B.1}$$

where N is the average photon number collected,

$$\Delta k^2 = \int_{-\infty}^{\infty} \mathrm{d}x \left| \frac{\partial \psi(x)}{\partial x} \right|^2, \tag{B.2}$$

$$\delta = \int_{-\infty}^{\infty} \mathrm{d}x \psi^*(x-d)\psi(x), \tag{B.3}$$

$$\gamma = \int_{-\infty}^{\infty} \mathrm{d}x \psi^*(x-d)\frac{\partial \psi(x)}{\partial x}. \tag{B.4}$$

We notice that the first term in Eq. B.1 is independent of the separation parameter and is identical to the one-dimensional QCR bound in Ref. [101]. It is also similar to the quantum limit we have derived earlier, and could be extended for two-dimensional separation parameter (d_X, d_Y). The key result is that for a small separation, J_d of Eq. (B.1) approaches the quantum bound of weak-source model:

$$\lim_{d \to 0} J_d = N\Delta k^2. \tag{B.5}$$

Hence, the Rayleigh resolution limit is similarly evaded for localization of two thermal point sources.

B.2 Review of Recent Experiments

We have proposed two potential measurement methods in Chapter 5 – two-dimensional SPADE and extended SLIVER – that could achieve the quantum resolution limit. In the past few months, there have been four proof-of-concept experiments [118, 119, 120, 121] reported.

1. The experiment reported in Ref. [118] was based on SLIVER scheme. The test sources were created from a laser-illuminated single-mode

optical fiber whose output is separated by a polarizing beamsplitter.

2. For sub-Rayleigh precision in separation measurement, Ref. [119] proposed detecting light in higher-order transverse electromagnetic modes (TEMs). The experiment was performed by means of heterodyne detection with the local oscillator in the desired TEM. The incoherent sources were created by placing a white paper library card before double slits illuminated by a single laser, and the card was moved in transverse plane to achieve averaging over the incoherent light statistics.

3. The experiment reported in Ref. [120] was based on binary SPADE [101] which has full performance of SPADE for small separation distance. They proposed a different experimental implementation, Super-resolved Position Localization by Inversion of Coherence along an Edge (SPLICE). The image filed was projected into a mode orthogonal to the fundamental TEM, the separation distance was deduced from the ratio between projection and total field. The light source was a heralded single-photon source, one photon was used for heralding while the other was split at 50-50 fiber-splitter and out coupled to free space.

4. Ref. [121] reported a quantum-optimal measurements on two incoherent point sources generated with a high-frequency-switched digital micromirror chip illuminated by a laser source. The experiment scheme was based on SPADE where light is measured in different HG basis. The projections onto zeroth- and first-order modes were performed with a spatial light modulator in the amplitude mode.

All four experiments report separation estimation error lower than predicted by Rayleigh criterion.

Appendix C

List of Publications

1. Shan Zheng Ang, Glen I. Harris, Warwick P. Bowen, and Mankei Tsang, "Optomechanical parameter estimation," New Journal of Physics **15**,103028 (2013).

   We tackled the problem of estimating system parameters from a noisy optomechanical system using a statistical framework. The estimation error of the proposed expectation-maximization algorithm follows fundamental error limit given by Cramér-Rao bound closely, and is lower than the commonly used averaging estimator.

2. Shilin Ng, Shan Zheng Ang, Trevor A. Wheatley, Hidehiro Yonezawa, Akira Furusawa, Elanor H. Huntington, and Mankei Tsang, "Spectrum analysis with quantum dynamical systems," Phys. Rev. A **93**, 042121 (2016).

   We tackled the problem of estimating spectrum parameters of a classical stochastic process coupled to a quantum dynamical system. As quantum noise is becoming a major limiting factor in this problem, a quantum limit to the accuracy of parameter estimation is required to understand such systems.

3. Shan Zheng Ang, Ranjith Nair, and Mankei Tsang, "Quantum limit for two-dimensional resolution of two incoherent optical point sources,"

arXiv:1606.00603 [quant-ph].

In this paper, we studied the problem of two-dimensional optical resolution in the framework of parameter estimation theory. We found that the quantum Cramér-Rao bound for estimation error of separation between two incoherent point sources is well below the conventional Rayleigh resolution limit. We proposed two linear-optics measurement schemes that approach the quantum limit at small separation.