QUANTUM DETECTION AND PARAMETER ESTIMATION FOR CONTINUOUSLY MEASURED SYSTEMS

NG SHI LIN

(B.Sc. (Hons) in Physics, NUS)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

NATIONAL UNIVERSITY OF SINGAPORE

2017

Supervisor:

Assistant Professor Mankei Tsang

Examiners:

Professor Berthold-Georg Englert

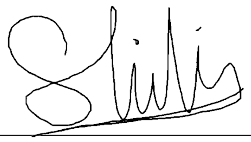
Professor Gong Jiangbin

Associate Professor Mădălin Guţă, University of Nottingham

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.



Ng Shi Lin

16 January 2017

Dedicated to my family.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Assistant Professor Dr. Mankei Tsang for his advice, patience, and moral support throughout my Ph.D. study. The completion of this thesis would not have been possible if not for his continuous guidance and support. I would like to extend my gratitude towards my colleagues from the Quantum Measurement Group. I thank Shan Zheng for his friendship and our constant discussions and/or arguments on various topics; Soham and Dan for their friendship; Ranjith and Xiao-Ming for numerous discussions and advice. Discussions with Shibdas, Cong Son, HL Dao, Alex, Sumei, Deng Jun, Davide and Adam are also greatly appreciated, and in many cases, their presentations during our group meetings have exposed me to different fields of research. I would like to take this opportunity to thank our collaborators from the University of New South Wales Canberra, the University of Tokyo, and the Australian National University: Dr. Trevor Wheatley, Dr. Hidehiro Yonezawa, Prof. Akira Furusawa and Prof. Elanor Huntington. Their experiment from six years ago is still telling us something.

Many thanks to Professor Gong Jiangbin for recommending me to Mankei's group four years ago, as well as his advice on thesis writing. I thank Professor Berthold-Georg Englert also for his useful recommendations during my qualifying examination and thesis writing. I would like to also thank Professor Kuok Meng Hau, Professor Ng Ser Choon, Associate

Professor Tok Eng Soon and Associate Professor Ramanathan Mahendiran; and lab staff members Mrs. Lee, Dicky, Mr. Chen and Mrs. Ng from the module PC3193 Experimental Physics II, which I tutored for two and a half years. My appreciation is also extended to the admin staff from both Department of Physics and Department of Electrical and Computer Engineering for administrative matters regarding my Ph.D. study. I am also grateful to the National University of Singapore and the National Research Foundation for offering a research scholarship.

I thank all of my friends for their company and the good time we have been through, many of whom also shaped how I view the world around me: Boon Hao for his sense of humour; Chuan Wai and Hong Tat for showing that goodness comes from every little thing in life; Qin Jin, Zheng Zhi and Sos Fuf for their wonderful friendships; Sheng Feng and Xian Zhi for being my oldest partners in crime; Valerio for the many small talks and showing me that religion is not always at odds with science; And of course the great people whom I know from the Association of Malaysian Chinese Independent Schools Alumni, for being tolerable towards my various antics.

Finally, I would like to thank my family members for their support. Thanks to my parents, Ng Choi and Oon Huat for all the hardships they endured in raising us up. Thanks to Su Qiong for always being there for us; Su Yi for being the fun sister; Shi Qi as the only one who would listen to my Science talks; Jeffrey for his support and two little additions to the family. Last but not least, I thank Xin Rui for her love and encouragement, and for making the world a better place.

Contents

Abstract **xiii**

List of publications **xv**

List of Figures **xvii**

1 Introduction **1**

1.1 Probability theory 2

1.2 The quantum regime and continuous measurement 3

1.3 Thesis outline 5

2 Theoretical background **7**

2.1 Stochastic calculus 7

2.2 Detection Theory 11

2.2.1 Binary hypothesis testing and likelihood ratio tests . 11

2.2.2 Performance of likelihood ratio tests 15

2.3 Estimation Theory 17

2.3.1 Random parameter estimation 17

2.3.2 Non-random parameter estimation and the Cramér-Rao lower bound 19

2.4 Quantum detection and estimation 23

2.5 Conclusion 27

3 Optimal signal processing for continuous qubit readout **29**

3.1 M-ary hypothesis testing 31
3.2 Gaussian noise model 32
3.2.1 Observation process 32
3.2.2 Estimator-correlator formula 35
3.2.3 Qubit dynamics 36
3.2.4 Estimator 37
3.2.5 Deterministic-signal detection 38
3.2.6 No spontaneous excitation ($L_m^+ = 0$) 40
3.2.7 No spontaneous decay ($L_m^- = 0$) 41
3.3 Poissonian noise model 42
3.3.1 Observation process 42
3.3.2 Estimator-correlator formula 43
3.3.3 Estimator 44
3.3.4 No spontaneous excitation ($L_m^+ = 0$) 45
3.3.5 No spontaneous decay ($L_m^- = 0$) 46
3.4 Conclusion 46

4 Spectrum analysis with quantum dynamical systems 49
4.1 Quantum metrology 50
4.1.1 Parameter estimation 50
4.1.2 Extended convexity 52
4.1.3 Dynamical systems 53
4.1.4 Variational bound 55
4.2 Continuous optical phase modulation 60
4.2.1 Error bounds 60
4.2.2 Spectral photon counting 63
4.2.3 Ornstein-Uhlenbeck spectrum analysis 66
4.2.4 Experimental data analysis 68
4.3 Conclusion 73

5 Spectral analysis for quantum stochastic processes 77

5.1 The modified extended convexity of quantum Fisher information 78

5.2 Gaussian bosonic systems 82

5.3 Upper bounds on the Quantum Fisher information 86

5.3.1 A family of bath states ρ_{tot} such that $\rho_B = \mathrm{tr}_{an}(\rho_{tot})$ 88

5.3.2 Upper bound on the quantum Fisher information . . 91

5.3.3 Tightening the upper bound for the symmetric case . 93

5.3.4 Tightening the upper bound for the antisymmetric case 96

5.4 Discussion and Conclusion 98

6 Thesis conclusion 101

6.1 Summary 101

6.2 Future outlook 102

Bibliography 105

A Quantum formalism of continuous QND for qubits 117

B Experimental data recalibration for Chap. 4 119

C A special class of thermal states for Chap. 5 121

D Proofs for the upper bounds for Chap. 5 125

D.1 The symmetric case and $R_2' = R_1'$ 125

D.2 The symmetric case and $R_2' = \frac{1}{2}$ 127

D.3 The antisymmetric case 128

Abstract

Many measurement schemes of quantum systems involve coupling of the system to an electromagnetic field, which is then measured continuously. This includes qubit systems, such as nitrogen-vacancy centers and quantum dots, as well as mechanical systems, such as trapped ions and cavity optomechanics. In this thesis, we study the problem of detection and estimation for quantum systems that are measured continuously. A signal processing architecture for qubit readout is proposed to determine the initial state of a qubit. We consider both Gaussian and Poissonian noise models and derive analytical solutions to our protocol, which should be useful for fast feedback control and error correction purposes. We also propose a framework of spectrum-parameter estimation for stochastic processes using quantum dynamical systems, proving fundamental limits and investigating measurement and data analysis techniques that approach the limits. Lastly, we discuss our attempt at generalizing the framework of spectrum-parameter estimation to include quantum stochastic processes.

List of publications

- Shilin Ng and Mankei Tsang. *Optimal signal processing for continuous qubit readout*, Phys. Rev. A **90**, 022325 (2014).

- 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).

List of Figures

3.1 Realizations of the signal component $S_m(t)x_m(t)$ 33

3.2 The Gaussian observation model 34

3.3 Implementation of the estimator-correlator formula 35

3.4 Duncan-Mortensen-Zakai (DMZ) equation 38

3.5 Solution to the DMZ equation 41

3.6 The Poissonian observation model 43

3.7 Estimator-correlator structure for the Poissonian observation model 44

3.8 Solution to the Poissonian DMZ equation 45

4.1 Examples of the hidden stochastic process $X(t)$ and generator $\hat{Q}$ 56

4.2 Adaptive homodyne detection and spectral photon counting 62

4.3 Plots of the quantum limit $\tilde{\mathcal{J}}^{-1}$ and homodyne limit $\tilde{j}^{-1}$ . . 69

4.4 Comparisons of the experimental mean-square estimation errors to the limits 74

Chapter 1

Introduction

The study of measurements in quantum mechanics is a fascinating subject. For example, the Copenhagen interpretation postulates a partition between the observer and the system being observed. The observer is not described by wave mechanics, rather his job is to 'collapse' the state of the system into one of the eigenstates of a Hermitian observable that is measured. As an apparatus is commonly used to measure a system, we can also consider the apparatus as part of the quantum system. The wavefunction collapse can happen in the larger Hilbert space consisting of both the apparatus and the original system. This leads to the so-called 'measurement problem' [1], where research in this direction has a few interesting implications [2, 3].

Ultimately, the purpose of measuring a system is to obtain information about the system. The problems we study in this thesis are detection and estimation. Detection theory is the study of hypothesis testing, where the objective is to decide the best hypothesis among several possibilities. One classic scenario of the detection problem is the detection of targets using sonar, where there are only two choices: whether a specified target exists, or not. In estimation theory, it is often more common to study continuous parameters such as temperature. As we will see in the next chapter, we can view the processes of choosing a hypothesis and estimating a parameter as

data processing protocols aimed at extracting useful information.

Decision and estimation strategies are often designed to perform with respect to a measure of quality, which is chosen according to the context. For example, we can focus on minimizing the probability of making a wrong decision for detection problems; for parameter estimation problems, we can use the variance of estimation error as a measure of accuracy. Owing to measurement errors or intrinsic probabilistic behaviors such as the quantum uncertainty, the process of measuring a system does not always give a deterministic outcome. Therefore, it is often impossible to devise a strategy where there would be no error, as it would imply a perfect measurement process.

Thus, given a noisy measurement outcome, there should be a limit to the performance of decision and estimation strategies. It is then interesting to study this limit, and ways of achieving it, both as a benchmark to evaluate the performance of a certain data processing strategy, and to discover fundamental limitations for a problem. In order to study this problem, we will need the formalism of probability theory.

1.1 Probability theory

A probability space is given by the triplet $\{\Omega, \mathfrak{F}, P\}$. Ω is the *sample space* containing all the possible *outcomes* each labeled by ω, and the σ-*field* $\mathfrak{F}$ contains subsets of Ω which satisfy the following conditions:

- $\Omega \in \mathfrak{F}$;
- $\mathfrak{F}$ is closed under complementation;
- $\mathfrak{F}$ is closed under countable union.

The elements of $\mathfrak{F}$ are to be interpreted as *events*, and probabilities are assigned to these events by a *probability measure* P. Defined on $\mathfrak{F}$, P

satisfies the usual properties of probability:

- $P(\Omega) = 1$;
- $0 \leq P(A) \leq 1$ for $A \in \mathcal{F}$;
- $P(\bigcup A_n) = \sum_n P(A_n)$ if $A_n \cap A_m = \emptyset, \quad m \neq n$.

A *random variable* is a function $Z : \Omega \to \Omega'$ which is measurable:

$$Z^{-1}(A') = \{\omega : Z(\omega) \in A'\} \in \mathfrak{F}, \quad \forall A' \in \mathfrak{F}'. \tag{1.1}$$

We call $z = Z(\omega)$ the *realization* of the random variable Z, and the *state space* Ω' is the collection of all z. Roughly speaking, the measurable function Z preserves the structures of the σ-field $\mathfrak{F}$, such that probabilities can be assigned to the σ-field $\mathfrak{F}'$ on the state space through the relation $P_Z(A') = P(Z^{-1}(A'))$. Therefore, we often ignore the underlying probability space $(\Omega, \mathfrak{F}, P)$ and study the probability space $(\Omega', \mathfrak{F}', P_Z)$ instead. For more information on probability theory, the readers are referred to [4, 5].

1.2 The quantum regime and continuous measurement

To generalize probability theory to the quantum regime, we have to reformulate the theory in terms of the density matrix $\hat{\rho}$ and the positive operator-valued measure (POVM) $\hat{E}$, both of which are operators on a Hilbert space. Similar to the previous section, let us assume that the experimental outcomes are labeled as $z \in \Omega'$, and $\mathfrak{F}'$ is a σ-field on the state space Ω'. The POVM $\hat{E}$ satisfies the following conditions [6–8]

- $\hat{E}(\Omega') = \hat{I}$;
- $\hat{E}(A) \geq 0$, for $A \in \mathfrak{F}'$;
- $\hat{E}\left(\bigcup A_n\right) = \sum_n \hat{E}(A_n)$, if $A_n \cap A_m = \emptyset, \quad m \neq n$.

POVM represents the effect of a particular measurement process, where the probability measure of the measurement process is given by the trace formula $P_{\hat{E}}(A) = \mathrm{tr}(\hat{E}(A)\hat{\rho})$. For discrete countable state space, we have $\hat{E}(z_i) \equiv \hat{E}_i$ and $\sum_i \hat{E}_i = \hat{I}$ [9, 10].

There are a few reasons to study detection and estimation theory in the quantum regime. First, while measurements in classical physics are described by probability measures, we need both the density matrix and the POVM to describe the statistics of a quantum measurement process. Thus, it is possible to extract more information by performing a different measurement on the quantum system, a freedom that is not present in classical systems. An example would be the recently discovered limit of resolution of incoherent point sources [11–13], which overturned long-held belief that the ability to resolve two point sources is constrained by Rayleigh's criterion.

Second, there are phenomena in quantum mechanics that are not explainable by classical physics [14]. One of the most well-known examples is entanglement [15], which is present in highly nonclassical states. These states provide nonclassical correlations and are known to be useful for tasks such as quantum cryptography and quantum teleportation. In terms of metrology, researchers have proposed to utilize these correlations as a resource to improve the accuracy of phase estimation and the resolution of imaging [16–18].

On a more pragmatic note, quantum mechanics is needed to take quantum noise into account. In particular, the interaction of light and matter has received considerable attention as means of measuring and controlling quantum systems [19–22]. For example, in cavity quantum electrodynamics

or cavity optomechanics, the electromagnetic field inside the cavity interacts with the system, either in the form of electromagnetic interaction or through radiation pressure acting on the system. Information about the system can then be obtained by measuring the electromagnetic field. In a lot of these experiments, the electromagnetic field is measured continuously. Continuous measurements allow the monitoring of time-dependent quantities of a system, such as the position of a mechanical oscillator [23, 24] and the force acting on it [25, 26]. Continuous measurements of a qubit are also often performed to infer the state of the qubit, which is the fundamental building blocks of quantum computers [9]. For example, the readout of a cavity quantum electrodynamics system [27] and the resonance fluorescence photon counting of qubits [28, 29] both involve continuous measurement of a qubit.

Fundamentally, all measurement processes take time to complete. It is then natural to consider continuous measurements and study the extraction of information by this type of measurements. For example, continuous measurements of a signal are useful for inferring the spectral density of the signal and studying the evolution of feedback-controlled systems [30, 31]. Motivated by these developments, in this thesis, we explore the problem of detection and estimation for quantum systems that are measured continuously.

1.3 Thesis outline

To end this chapter, we give an outline of this thesis. We start by reviewing concepts that are relevant to this thesis in Chap. 2, which include stochastic calculus, detection, and estimation theory for both classical and quantum statistics. In particular, we emphasize detection and estimation theory and highlight similarities as well as differences between the quantum and

classical cases.

In Chap. 3, we study the problem of distinguishing the initial state of a qubit which is subjected to continuous measurements. We propose an optimal signal processing protocol that can infer the qubit state from the measurement in the presence of noise and qubit dynamics. Assuming continuous quantum nondemolition measurements with Gaussian or Poissonian noise and a classical Markov model for the qubit, we derive analytic solutions to the protocol in some special cases of interest using stochastic calculus.

In Chap. 4, we prove a measurement-independent quantum limit to the accuracy of estimating the spectrum parameters of a classical stochastic process coupled to a quantum dynamical system. We demonstrate our results by analyzing the data from a continuous optical phase estimation experiment and showing that the experimental performance with homodyne detection is close to the quantum limit. We further propose a spectral photon counting method that can attain quantum-optimal performance for weak modulation and a coherent-state input, with an error scaling superior to that of homodyne detection at low signal-to-noise ratios.

We generalize the results of Chap. 4 to include the case of quantum stochastic processes in Chap. 5. Inspired by the techniques used in Chap. 4, we provide modifications to the variational method for quantum Fisher information in the case of a quantum stochastic process. We then use the modified variational method to find quantum limits to the accuracy of estimating the spectrum parameters of a quantum stochastic process.

Lastly, we give an overall summary of the thesis in the conclusion chapter and discuss some possible future directions and outlooks.

Chapter 2

Theoretical background

In this chapter, we give a review of various concepts that are crucial to the understanding of later chapters. We present stochastic calculus in the first section as a framework to describe stochastic processes and highlight differences between ordinary calculus and stochastic calculus. In order to study diffusive processes and jump processes, we introduce stochastic differential equation driven by the Wiener process and the Poisson process. These equations will allow us to describe measurement processes such as homodyne detection and photon counting. Concepts central to classical detection and estimation theory such as likelihood ratio tests and Fisher information, as well as their generalizations to the quantum regime, will be introduced in the following sections. For a more detailed account of detection and estimation theory, the readers are referred to [32–35] for classical statistics and [7, 36] for quantum statistics.

2.1 Stochastic calculus

Let us consider a *stochastic differential equation* given by

$$dx(t) = a[x(t), t]dt + b[x(t), t]dW(t), \tag{2.1}$$

where $a[x(t), t]$ and $b[x(t), t]$ are functions of x and t. $W(t)$ is the *Wiener process*, which is a Gaussian process that satisfies the following conditions

$$\begin{aligned} \mathbb{E}[W(t)] &= 0, \\ \mathbb{E}[W(t)^2] &= t, \\ \mathbb{E}[W(t) - W(s)|W(t')] &= 0, \quad t > s \geq t'. \end{aligned} \tag{2.2}$$

Eq. (2.1) should be interpreted as an integral equation given by

$$x(t) - x(0) = \int_0^t a[x(t'), t']dt' + \int_0^t b[x(t'), t']dW(t'), \tag{2.3}$$

where the integration with respect to the Wiener process is called a *stochastic integration*. There are mainly two ways to define the stochastic integration, the first one is the *Itô integral*, where the integral is defined by a forward pointing Riemann sum[1]

$$\int_0^t b[x(t'), t']dW(t') \equiv \lim_{n\to\infty} \sum_{j=1}^{n} b[x(t_{j-1}), t_{j-1}][W(t_j) - W(t_{j-1})], \tag{2.4}$$

The second one is the *Stratonovich integral* where we use the mid point Riemann sum:

$$\int_0^t f(t')dW(t') \equiv \lim_{n\to\infty} \sum_{j=1}^{n} b[\frac{x(t_j) + x(t_{j-1})}{2}, t_{j-1}][W(t_j) - W(t_{j-1})]. \tag{2.5}$$

The solution $x(t)$ to the stochastic differential equation Eq. (2.1) given in Eq. (2.3) depends on the definition of the stochastic integration. Therefore, given the stochastic differential equation Eq. (2.1), one has to specify whether to interpret the differential equation in the Itô sense or the Stratonovich sense. This is in contrast to ordinary calculus, where the different Riemann sums converges to the same limit regardless the choice of

[1]The limit is in the mean square sense, where $\lim_{n\to\infty} X_n = X$ means $\lim_{n\to\infty} \mathbb{E}\{(X_n - X)^2\} = 0$. See [37].

forward pointing sum or mid point sum. In fact, given Eq. (2.1) in the Itô sense and let $g \equiv g(x,t)$ be a function of x and t, the differential of g is given by *Itô's lemma* [37, 38]:

$$dg = \left[\frac{\partial g}{\partial t} + \frac{b^2}{2}\frac{\partial^2 g}{\partial x^2}\right] dt + \frac{\partial g}{\partial x} dx. \tag{2.6}$$

Compared to the ordinary differential rule, there is an additional second partial derivative term. On the other hand, the differential rule for stochastic differential equations in the Stratonovich sense follows the usual calculus rules. While Itô calculus follows a different differential rule, it is often much easier to manipulate mathematically due to the fact that $b(x,t)$ and $dW(t)$ are statistically independent of each other [37]. The choice of the appropriate calculus depends on the particular application [39]. In general, one can argue that the Stratonovich definition is appropriate for systems that are influenced by external sources of noise. Itô calculus can find applications in systems driven by intrinsic noise, for example in radioactive decay where the probability of decay depends on the number of particles before the decay.

We can understand Itô's lemma by expanding $g(x,t)$ in a Taylor series. Denoting the increments as δg, δx, and δt,

$$\delta g \approx \frac{\partial g}{\partial t}\delta t + \frac{\partial g}{\partial x}\delta x + \frac{1}{2}\left[\frac{\partial^2 g}{\partial x^2}\delta x^2 + \frac{\partial^2 g}{\partial t^2}\delta t^2 + 2\frac{\partial^2 g}{\partial x \partial t}\delta t \delta x\right], \tag{2.7}$$

where we ignored terms higher than second order in both δx and δt. Keeping only terms that are first order in δx and δt and compare δg to Eq. (2.6), we see that $\delta x \delta t = 0$ and $(\delta x)^2 = b^2 \delta t$. Expanding $\delta x = a\delta t + b\delta W$, we can summarize the Itô's lemma as:

$$\begin{aligned} dt dW &= 0, \\ dW^2 &= dt, \end{aligned} \tag{2.8}$$

which roughly means that dW is on the order of $\sqrt{dt}$. The significance of Itô's lemma is that it allows computation of stochastic integrals without needing to resort to the basic definition, much like how chain rule and the fundamental theorem of calculus allow us to compute ordinary integrals without having to start with Riemann sums.

The stochastic processes we described so far have continuous paths. However, for particle counting experiments, the particle count as a function of time is not continuous. The paths of these stochastic processes must have discontinuities at the time when a particle is registered. A good description of these processes is given by the *Poisson process* [40], where $n(t) \in \{0, 1, 2, ...\}$ is the number of particles counted up until time t. The probability distribution of a Poisson process is characterized by an intensity function $\lambda(t)$:

$$P\left(n(t) - n(s)\right) = \exp\left[-\int_s^t d\tau\lambda(\tau)\right] \frac{\left[\int_s^t d\tau\lambda(\tau)\right]^{n(t)-n(s)}}{(n(t) - n(s))!}. \tag{2.9}$$

A stochastic differential equation of the form

$$dx(t) = a[x(t), t]dt + b[x(t), t]dn(t) \tag{2.10}$$

can also be interpreted in the Itô sense. Under a change of variable $g = g(x, t)$, the differential dg is given by [40]

$$dg = \left(\frac{\partial g}{\partial t} + a\frac{\partial g}{\partial x}\right) dt + [g(x + b, t) - g(x, t)]dn. \tag{2.11}$$

To understand the Itô's lemma for the Poisson process, let us define the increment by $\delta n(t) \equiv n(t + \delta t) - n(t)$. The probabilities of $\delta n(t) = \{0, 1\}$ are given by

$$\begin{aligned} P(\delta n(t) = 0) &= 1 - \lambda(t)\delta t + \mathcal{O}(\delta t), \\ P(\delta n(t) = 1) &= \lambda(t)\delta t + \mathcal{O}(\delta t), \end{aligned} \tag{2.12}$$

while for $\delta n(t) \geq 2$, the probability is at least on the second order of δt. Therefore for small intervals, there can only be unit jump or no jump at all, thus $[\delta n(t)]^2 = \delta n(t)$. Assume that $\delta n(t)\delta t = 0$ and expand $g(x+\delta x, t+\delta t)$ in a Taylor series, one would then obtain Itô's lemma Eq. (2.11). Hence, we can also summarize Itô's lemma for Poisson process as

$$\begin{aligned} dtdn &= 0, \\ dn^2 &= dn. \end{aligned} \tag{2.13}$$

2.2 Detection Theory

2.2.1 Binary hypothesis testing and likelihood ratio tests

The study of hypothesis testing is also called detection theory in the literature, where the simplest scenario is binary hypothesis testing. Suppose that we are given two hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$. We define $P(.|\mathcal{H}_m)$ as the probability measure conditioned on hypothesis $\mathcal{H}_m$. The task is to decide on a hypothesis given an experiment outcome. This separates the state space into two regions, labeled as $\mathcal{Z}_m$ which corresponds to the hypothesis $\mathcal{H}_m$. We call $\mathcal{Z}_m$ the decision region for the hypothesis $\mathcal{H}_m$: when an observation falls in the region $\mathcal{Z}_m$, we would decide on $\mathcal{H}_m$.

In the Bayesian setting, we assume that we have prior knowledge about these hypotheses before the observation is made. This knowledge is captured by the prior probability distribution $P(\mathcal{H}_m)$. When we make a choice between the hypotheses, there would be four scenarios: choose $\mathcal{H}_0$, $\mathcal{H}_0$ is true; choose $\mathcal{H}_0$, $\mathcal{H}_1$ is true; and so on. Denoting the positive number C_{nm} as the *cost* that corresponds to each scenario, the expected value of the cost

can be used as a figure of merit for the performance of a detection strategy:

$$\mathfrak{R} = \sum_{n,m=0,1} C_{nm} P(\mathcal{H}_m) \int_{\mathcal{Z}_n} dP(z|\mathcal{H}_m), \tag{2.14}$$

where we call $\mathfrak{R}$ the *risk* of a detection strategy. In Bayes tests, we aim to minimize the risk by changing the decision regions. A simple change of the domain of integration yields

$$\mathfrak{R} = \sum_{m=0,1} C_{1m} P(\mathcal{H}_m) + \sum_{m=1,2} (C_{0m} - C_{1m}) P(\mathcal{H}_m) \int_{\mathcal{Z}_0} dP(z|\mathcal{H}_m), \tag{2.15}$$

where only the second term depends on $\mathcal{Z}_0$. By changing the probability measure, the second term can be rewritten as

$$\int_{\mathcal{Z}_0} \left[(C_{00} - C_{10}) P(\mathcal{H}_0) + (C_{01} - C_{11}) P(\mathcal{H}_1) \frac{dP(z|\mathcal{H}_1)}{dP(z|\mathcal{H}_0)} \right] dP(z|\mathcal{H}_0), \tag{2.16}$$

where $\Lambda(z) = \frac{dP(z|\mathcal{H}_1)}{dP(z|\mathcal{H}_0)}$ is called the *likelihood ratio*.

Hence, to minimize $\mathfrak{R}$ one simply chooses $\mathcal{Z}_0$ such that the integrand in Eq. (2.16) is always negative, which gives us the decision rule in terms of the likelihood ratio:

$$\Lambda(z) < \eta \equiv \frac{P(\mathcal{H}_0)(C_{10} - C_{00})}{P(\mathcal{H}_1)(C_{01} - C_{11})}, \quad z \in \mathcal{Z}_0. \tag{2.17}$$

Note that we have assumed that $C_{10} - C_{00}$ and $C_{01} - C_{11}$ are positive. In other words, we assume that the cost of making the wrong decision is always larger than the cost of making the right decision. The decision rule given by Eq. (2.17) involves a data processing step to calculate a positive function $\Lambda(z)$ and test it against a threshold η. Such tests are called *likelihood ratio tests*, and we will see the same decision-making process in a frequentist framework. Note that the region where $\Lambda(z) = \eta$ does not contribute to the risk, thus can be freely assigned to any of the decision regions.

Before we continue our discussion on likelihood ratio tests, let us examine the likelihood ratio in more detail. Usually, instead of working directly with a probability measure, the probability measure is given in terms of a reference measure M:

$$\int_A dP(z|\mathcal{H}_m) = \int_A \frac{dP(z|\mathcal{H}_m)}{dM(z)} dM(z). \tag{2.18}$$

The quantity

$$p(z|\mathcal{H}_m) = \frac{dP(z|\mathcal{H}_m)}{dM(z)} \tag{2.19}$$

as a function of $\mathcal{H}_m$ is called the *likelihood function*. In particular, for continuous random variable, the likelihood function is just the probability density function conditioned on $\mathcal{H}_m$; for discrete random variable, the likelihood function is the probability mass function conditioned on $\mathcal{H}_m$. Using the chain rule $\frac{dP(z|\mathcal{H}_1)}{dP(z|\mathcal{H}_0)} = \frac{dP(z|\mathcal{H}_1)}{dM(z)} / \frac{dP(z|\mathcal{H}_0)}{dM(z)}$, the likelihood ratio can also be expressed in terms of the ratio of likelihood functions:

$$\Lambda(z) = \frac{p(z|\mathcal{H}_1)}{p(z|\mathcal{H}_0)}, \tag{2.20}$$

which explains the nomenclature of $\Lambda(z)$.

For cases where the prior probabilities or the costs are not available, we may focus on the false alarm probability P_F and the miss probability P_M defined by

$$P_F = \int_{\mathcal{Z}_1} dP(z|\mathcal{H}_0), \tag{2.21}$$

$$P_M = \int_{\mathcal{Z}_0} dP(z|\mathcal{H}_1). \tag{2.22}$$

The Neyman-Pearson test seeks to constrain one of these error probabilities and minimize the other. This choice is mainly because P_F and P_M both depend on the decision regions and in general, can not be minimized

simultaneously. Suppose that the experimenter decides to constrain the false alarm probability to be less than a certain threshold α:

$$P_F = \alpha' \leq \alpha \tag{2.23}$$

and minimize the miss probability[2]. We start with a given $\alpha' \leq \alpha$ and employ the method of Lagrange multiplier to minimize P_M, where the following function is minimized in place of P_M:

$$P = P_M + \eta[P_F - \alpha'], \tag{2.24}$$

and η is the Lagrange multiplier yet to be determined. Again we use the same technique as in the Bayesian case, namely, a change of integration region and probability measure, to obtain the decision rule

$$\Lambda(z) < \eta, \quad z \in \mathcal{Z}_0. \tag{2.25}$$

η is then chosen to satisfy the constraint $P_F = \alpha'$. Finally, the decision rule is given by

$$\begin{aligned} \Lambda(z) < \eta, \quad z \in \mathcal{Z}_0, \\ P_F = \int_{\mathcal{Z}_1(\eta)} dP(z|\mathcal{H}_0) = \alpha', \end{aligned} \tag{2.26}$$

which is a likelihood ratio test as in the Bayes test, but with a different threshold η. To obtain a test for $P_F \leq \alpha$, α' is then increased to the largest $\alpha' \leq \alpha$, and P_M would be minimized given the constraint $P_F \leq \alpha$.

In practice, the same experiment would often be repeated under similar circumstances in the hope of acquiring better performance. Such N repeated experiments would then produce N independent, identically distributed (IID) random variables Z_j. The joint probability measure of $\{Z_j\}_{j=1}^N$ is just the product of the individual probability measures, and

[2]The inequality is included to describe cases where the exact value of $P_F = \alpha$ is not achievable by any decision rule, for example when Z is a discrete random variable.

the likelihood ratio is a multiplication $\Lambda(z_1...z_N) = \prod_j \Lambda(z_j)$. Thus to implement the likelihood test, it is sometimes easier to use the *log-likelihood ratio* $\ln \Lambda(z)$ as it becomes the sum of the individual log-likelihood ratios under the IID condition. Since the logarithmic function is a monotonically increasing function, the likelihood ratio test can be formulated in terms of the log-likelihood ratio, where the threshold is $\ln \eta$.

2.2.2 Performance of likelihood ratio tests

The performance of a Bayes test or a Neyman-Pearson test can be evaluated by computing explicitly the Bayesian risk or the error probabilities. However, analytical formulae are not always available, and bounds on the performance are often derived in order to provide insight. Let us first examine the Bayesian case, and assign the Bayesian cost

$$C_{01} = C_{10} = 1,$$
$$C_{11} = C_{00} = 0.$$

The risk is then given by the average error probability

$$\begin{aligned} \mathfrak{R} = P_e &= P(\mathcal{H}_0)P_F + P(\mathcal{H}_1)P_M \\ &= \int \min\left[P(\mathcal{H}_0), P(\mathcal{H}_1)\Lambda(z)\right] dP(z|\mathcal{H}_0), \end{aligned} \tag{2.27}$$

where the integration is over all the state space of Z. The function $\min(a, b)$ outputs the smaller number between the pair $\{a, b\}$.

Since exponentiation x^s is a monotonic function in terms of s, where it is monotonically increasing (decreasing) when $x > 1$ ($x < 1$), we have the following identity for $a, b \geq 0$

$$\min(a, b) \leq a^{1-s}b^s \leq \max(a, b), \quad 0 \leq s \leq 1. \tag{2.28}$$

Introducing the conditional expectation notation $\mathbb{E}_{Z|\mathcal{H}_0}(\Lambda^s) = \int \Lambda^s(z) dP(z|\mathcal{H}_0)$, an upper bound on P_e is given by:

$$P_e \leq c(s) \equiv \mathbb{E}_{Z|\mathcal{H}_0}(\Lambda^s), \quad 0 \leq s \leq 1. \tag{2.29}$$

The Chernoff coefficient $c(s)$ equals unity when $s = 0$ or $s = 1$, and is a convex function in terms of s. Its derivative $\frac{dc(s)}{ds}$ is negative at $s = 0$ but is positive at $s = 1$, meaning that a local minimum exists within the interval $0 \leq s \leq 1$. Therefore, the Chernoff coefficient is always smaller than unity for $s \in [0, 1]$, making it a meaningful upper bound on the error probabilities. Aside from the minimum Chernoff coefficient, some particular cases of the Chernoff coefficient are also of interest, such as when $s = \frac{1}{2}$, where it is also known as the Bhattacharyya coefficient or the fidelity. We note that the error bound given by the Chernoff coefficient is also often given in terms of a positive exponent, termed *Chernoff exponent*:

$$C(s) = -\ln \mathbb{E}_{Z|\mathcal{H}_0}(\Lambda^s). \tag{2.30}$$

Tighter upper bounds on both the error probabilities P_F and P_M can be found in terms of the Chernoff exponent, see [32]. These bounds can then be applied to bound the performances of both the Bayes test and the Neyman-Pearson test.

To conclude this section, let us consider the performance of a repeated experiment. Under the IID condition, the maximum Chernoff exponent is the asymptotic error rate for Bayes tests:

$$\lim_{N\to\infty} \frac{1}{N} \ln P_e = -\max_s C(s), \tag{2.31}$$

where the maximum Chernoff exponent is termed *Chernoff information* in the literature. This relation means that for sufficiently large number

of experiments N, the average error probability of a Bayes test decreases exponentially to zero at a rate given by the Chernoff information. For Neyman-Pearson tests in IID experiments, similar relations hold for the error probabilities. However, the corresponding error exponents are given by the relative entropy instead. See [41] for more information on asymptotic hypothesis testing and the relation between Chernoff information and relative entropy. These results show that repeated experiments can improve our ability to test hypotheses, as common wisdom dictates.

2.3 Estimation Theory

2.3.1 Random parameter estimation

As in detection theory, let us categorize parameter estimation into two paradigms, depending on whether a prior probability measure is assigned to the parameter of interest. For Bayesian estimation, where we assume prior knowledge about the parameter, we treat the unknown parameter as a random variable Θ, and the realization is denoted as θ. For multiparameter problems, Θ is a multivariate random variable which can be represented by a column vector.

Analogous to the detection theory, we define a cost function $C(\theta, \check{\theta})$ where $\check{\theta} \equiv \check{\theta}(z)$ is an estimator of θ given the outcome z. Given a cost function, the Bayesian procedure is to find the estimator that minimizes the Bayesian risk:

$$\mathfrak{R} = \mathbb{E}_{Z,\Theta}(C), \tag{2.32}$$

where the expectation is over the joint probability measure of both the random variables Z and Θ. For scalar parameter problems, the quantity $\theta_e = \check{\theta} - \theta$ represents the error of the estimator $\check{\theta}$. Thus, it can be used to define the cost function such that $C(\theta, \check{\theta}) = C(\theta_e)$. This way, the Bayesian

Risk $\mathfrak{R}$ can be interpreted as the expected value of estimation error. However, the definition $C = \check{\theta} - \theta$ admits negative values of error, which may be undesirable as the Bayesian Risk can be low even when the estimation error θ_e has a high variability. A natural alternative is to use the absolute error $C = |\check{\theta} - \theta|$ so that there would be no negative error. Another alternative is the square error $C = (\check{\theta} - \theta)^2$ where large values of errors are penalized.

To find the estimator which minimizes the Bayesian risk, we write $\mathbb{E}_{Z,\Theta}(C) = \mathbb{E}_Z[\mathbb{E}_{\Theta|z}(C)]$, where $\mathbb{E}_{\Theta|z}(.)$ denotes the expectation conditioned on the experiment outcome z. Hence, minimizing $\mathfrak{R}$ is equivalent to minimizing $\mathbb{E}_{\Theta|z}(C)$. A particular elegant solution is given by the *minimum mean-square error estimator*, where the mean-square error is minimized. The minimum mean-square error estimator is the *posterior mean* of Θ:

$$\check{\theta}_{MMSE}(z) = \mathbb{E}_{\Theta|z}(\Theta). \tag{2.33}$$

For minimum mean absolute error estimator, the estimator is given by the median of the posterior probability density of Θ given an outcome z.

We note that many applications of hypothesis testing require the estimation of a random parameter. For instance, consider the problem of detecting Gaussian signals in white Gaussian noise:

$$\begin{aligned} \mathcal{H}_0 : &\quad z(t) = \xi(t), \\ \mathcal{H}_1 : &\quad z(t) = x(t) + \xi(t), \end{aligned}$$

where $\mathbb{E}[\xi(t)\xi(t')] = \delta(t - t')$ and $x(t)$ is a zero mean Gaussian signal with covariance function $\mathbb{E}[x(t)x(t')] = K(t, t')$. If we define the time integral of

the $z(t)$ as $y(t) = \int_0^t z(s)ds$, the hypotheses become

$$\begin{aligned} \mathcal{H}_0 : &\quad y(t) = W(t), \\ \mathcal{H}_1 : &\quad y(t) = \int_0^T x(t')dt' + W(t), \end{aligned}$$

where $W(t)$ is the Wiener process. The log-likelihood ratio can be expressed in the Itô sense as [42, 43]

$$\ln \Lambda(T) = \int_0^T \check{x}_1(t)dy(t) - \frac{1}{2}\int_0^T \check{x}_1^2(t)dt, \tag{2.34}$$

where $\check{x}_1(t)$ is the minimum mean-square error estimator given past measurement outcome $\{y(s), 0 \leq s \leq t\}$ and conditioned on $\mathcal{H}_1$. This expression for the likelihood ratio where the estimator is correlated with the observation is called the *estimator-correlator formula*. It is valid even if the signal $x(t)$ is not Gaussian [42, 43].

2.3.2 Non-random parameter estimation and the Cramér-Rao lower bound

In a lot of situations, it is undesirable to treat the unknown parameter as random. For these situations, we assume that the parameter θ is fixed but hidden from the observers. What is available to the observers is then the conditional probability measure $P_{Z|\theta}$, where θ is the true value of the parameter of interest. Similar to the case of random parameter estimation, let us first consider a single parameter problem. It is preferred to have an estimator which is accurate in the sense of being distributed close to the true value of the unknown parameter. As the moments of a distribution provide a sense of the 'shape' of the distribution, we can study the moments of the estimator to quantify the quality of an estimator. The first moment is the mean of the estimator, conditioned on the true value of the parameter.

An accurate estimator is expected to have a mean which is close to the true value. Such an estimator is called an *unbiased estimator*, where $\mathbb{E}_{Z|\theta}(\check{\theta} - \theta) = 0$. The second moment is the variance

$$Var(\check{\theta} - \theta) = \mathbb{E}_{Z|\theta}[(\check{\theta} - \theta)^2], \tag{2.35}$$

which tells us the 'spread' of the estimator. Thus, we expect that the more accurate an unbiased estimator is, the lower the variance of its estimation error.

Our goal is to investigate the limit of estimation accuracy in terms of the variance of estimation error, which is also the mean-square error of the estimator. To proceed, let us generalize to vectorial parameter and introduce the mean-square estimation error matrix,

$$\Sigma = \mathbb{E}_{Z|\theta}\Big[(\check{\theta} - \theta)(\check{\theta} - \theta)^\top\Big]. \tag{2.36}$$

A lower bound on the estimation error matrix for unbiased estimators is given by the *Cramér-Rao bound*:

$$\begin{gathered} \Sigma \geq j^{-1}(p_Z), \\ j_{\mu\nu}(p_Z) = \mathbb{E}_{Z|\theta}\Big(\frac{\partial \ln p_Z}{\partial \theta_\mu}\frac{\partial \ln p_Z}{\partial \theta_\nu}\Big), \end{gathered} \tag{2.37}$$

where $j(p_Z)$ is the *Fisher information matrix*, and $p_Z \equiv p_Z(z|\theta)$ is the likelihood function, which is a function of the parameter θ. The matrix inequality $\Sigma \geq j^{-1}$ means that the matrix $\Sigma - j^{-1}$ is a positive-semidefinite matrix. The proof is fairly easy to reproduce [7, 32, 34]: we assume *regularity conditions*, where differentiation $\frac{\partial}{\partial \theta_\mu}$ and integration (of the expectation operation) are swappable, and the log-likelihood function is differentiable. One then starts with taking derivative of $\mathbb{E}_{Z|\theta}(\check{\theta}_\mu - \theta_\mu) = 0$ with respect to θ_ν to obtain $\mathbb{E}_{Z|\theta}\Big[(\check{\theta}_\mu - \theta_\mu)\frac{\partial \ln p_Z}{\partial \theta_\nu}\Big] = \delta_{\mu\nu}$. Multiplying vectors a_μ and b_ν

and sum over all indices, we get

$$\begin{aligned} E_{Z|\theta}(AB) &= a^\top b, \\ A &= \sum_\mu a_\mu(\check{\theta}_\mu - \theta_\mu), \\ B &= \sum_\mu b_\mu \frac{\partial \ln p_Z}{\partial \theta_\mu}. \end{aligned} \tag{2.38}$$

As the expectation $E_{Z|\theta}(AB)$ is an inner product, we can apply Schwarz inequality to get

$$\mathbb{E}_{Z|\theta}(A^2)\mathbb{E}_{Z|\theta}(B^2) = (a^\top \Sigma a)(b^\top j b) \geq |a^\top b|^2. \tag{2.39}$$

Finally, choosing $b = j^{-1}a$ and after some algebraic manipulations, one obtains $\Sigma \geq j^{-1}$.

An estimator is called *efficient* when it saturates the Cramér-Rao bound. This happens when the functions A and B are linearly dependent, where the condition is

$$\sum_\mu a_\mu(\check{\theta}_\mu - \theta_\mu) = \sum_\mu b_\mu \frac{\partial \ln p_Z}{\partial \theta_\mu}. \tag{2.40}$$

Since this condition must be fulfilled for all values of z and θ, there is no guarantee of the existence of such an estimator. However, whenever this condition is satisfied, the efficient estimator is given by the maximum likelihood estimator (MLE). The MLE is an estimator which is chosen to be the value of θ that maximizes the likelihood function given an outcome z:

$$\check{\theta}_{MLE}(z) = \arg\max_\theta p_Z(z|\theta), \tag{2.41}$$

where $\max_\theta$ denotes maximizing over all possible values of θ. Assuming that the likelihood function is differentiable, the MLE can be found by solving the simultaneous equations $\frac{\partial p_Z}{\partial \theta_\mu}|_{\theta=\check{\theta}_{MLE}} = 0$, or equivalently in terms of the *log-likelihood function* $\frac{\partial \ln p_Z}{\partial \theta_\mu}|_{\theta=\check{\theta}_{MLE}} = 0$. From the condition Eq. (2.40), we

have

$$0 = \sum_{\mu} b_{\mu} \frac{\partial \ln p_Z}{\partial \theta_{\mu}}|_{\theta=\check{\theta}_{MLE}} = \sum_{\mu} a_{\mu}(\check{\theta}_{\mu} - \theta_{\mu})_{\theta=\check{\theta}_{MLE}}, \tag{2.42}$$

and $\check{\theta}$ equals the MLE estimator. Thus, whenever an efficient estimate exists, it is given by the MLE.

The idea that repeated measurements should improve the accuracy of an estimator is captured by the fact that, for independent experiments, the total Fisher information is the sum of Fisher information of individual experiments. Hence, under the IID condition, the mean-square error matrix is lower bounded by

$$\Sigma \geq \frac{1}{Nj}, \tag{2.43}$$

where N is the number of repetitions. Under certain conditions the MLE is asymptotically efficient [32], meaning that the lower bound in Eq. (2.43) is achievable as $N \to \infty$. This shows that as more experiments are performed, estimators which are more accurate can be formed from the experiment outcomes. In addition, the MLE is also asymptotically consistent, where it converges in probability to the true value of the parameter[3] as more experiments are performed.

Finally, we can draw connections between detection theory and estimation theory by considering the hypothesis testing problem where the two different hypotheses are given by different true values of a parameter. Suppose that $\mathcal{H}_0$ assumes the value θ while $\mathcal{H}_1$ assumes θ', the Chernoff coefficient is a function of both θ and θ'. The derivative of the Bhattacharyya coefficient $c(0.5)$ with respect to θ_{μ} equals zero when $\theta' = \theta$. Its second derivative is proportional to the Fisher information [44]:

$$j_{\mu\nu} = -4\Big[\frac{\partial^2}{\partial\theta_{\mu}\partial\theta_{\nu}} c(0.5;\theta,\theta')\Big]_{\theta'=\theta}. \tag{2.44}$$

[3]It means that the probability that the estimator is finitely different from the true parameter, is zero

These results show that for two hypotheses which are infinitesimally close,

$$c(0.5;\theta,\theta+\delta\theta) \approx 1 - \frac{1}{8}\delta\theta^\top J \delta\theta, \tag{2.45}$$

see also [45]. We note a relation similar to Eq. (2.44) can also be given in terms of the Bhattacharyya distance $C(0.5)$ [46], where

$$j_{\mu\nu} = 4\Big[\frac{\partial^2}{\partial\theta_\mu \partial\theta_\nu} C(0.5;\theta,\theta')\Big]_{\theta'=\theta}. \tag{2.46}$$

These relations allow one to calculate the Fisher information whenever the Chernoff coefficient or Chernoff exponent is available.

2.4 Quantum detection and estimation

In the context of quantum mechanics, the concepts introduced in the previous sections can be regarded as a theory of signal processing after a given measurement has been performed. For example, the Cramér-Rao bound is specific to a given POVM. It is not clear whether better performance can be acquired given another POVM. Hence, to study the optimal performance permissible by quantum mechanics, one also has to work on the level of POVM and find the POVM which provides the best performance. Working on POVM amounts to experimental design. Thus, in contrast to the classical case, we have the freedom to choose the type of measurement to perform in the quantum case.

For binary hypothesis testing, each hypothesis $\mathcal{H}_m$ is now associated with a density matrix $\hat{\rho}_m$. The decision region $\mathcal{Z}_m$ of classical detection theory corresponds to a POVM $\hat{E}_m$ in the quantum case, where $\hat{E}_0 + \hat{E}_1 = \hat{I}$. The probability of choosing $\mathcal{H}_n$ while $\mathcal{H}_m$ is correct is then given by $\Pr(n|m) = \mathrm{tr}(\hat{E}_n \hat{\rho}_m)$, which corresponds to the quantity $\int_{\mathcal{Z}_n} dP(Z|\mathcal{H}_m)$ in Sec. (2.2). In particular, we have $P_F = \Pr(1|0)$ and $P_M = \Pr(0|1)$ for

the false alarm probability and the miss probability. The Bayesian risk of detection in Eq. (2.14) is also similarly defined:

$$\mathfrak{R} = \sum_{n,m=1,2} C_{nm} P(\mathcal{H}_m) \Pr(n|m). \tag{2.47}$$

The goals of Bayes test and the Neyman-Pearson test for quantum hypothesis testing are the same as in classical statistics. The difference here is that in order to minimize the risk (miss probability) for Bayes test (Neyman-Pearson test), one needs to solve a series of operator optimization equations for the POVM [7]. The result is somewhat similar to the likelihood ratio test, where the POVM $\hat{E}_1$ ($\hat{E}_0$) is a projection into the positive (negative) eigenspace of the operator $\hat{\rho}_1 - \eta\hat{\rho}_0$, and η is a threshold that depends on the problem. Again, exact performance is often difficult to evaluate but bounds on the average error probability are available, see [47]. A quantum analogue of the Chernoff coefficient or Chernoff exponent can also be defined [48, 49]; it is the best error exponent achievable asymptotically in a quantum IID setting, analogous to the classical case.

The case of quantum estimation is very similar to quantum detection, where the conditional probability measure $P_{Z|\theta}$ is replaced by a density matrix $\hat{\rho}_\theta$. For Bayesian estimation, a prior probability measure P_Θ is assumed for the parameter and the unconditional density matrix is given by $\hat{\rho} = \mathbb{E}_\Theta(\hat{\rho}_\theta)$. We will also need to formulate the quantum estimation process in terms of a POVM $\hat{E}$. Suppose that the POVM can be given by an integral of infinitesimal operators

$$\hat{E}(A) = \int_A d\hat{E}(\check{\theta}), \tag{2.48}$$

where $\check{\theta}$ is the estimator. The Bayesian risk is defined as

$$\mathfrak{R} = \mathbb{E}_\theta \left(\int C(\theta, \check{\theta}) \operatorname{tr} \left[\hat{\rho}_\theta d\hat{E}(\check{\theta}) \right] \right). \tag{2.49}$$

In order to minimize $\mathfrak{R}$, we need to choose a specific cost function $C(\theta, \check{\theta})$ and solve a continuous version of the optimization equations as in the quantum detection problem [7].

Similar to the classical estimation theory, we will focus on the estimation of nonrandom parameters. We are interested in the optimal accuracy allowed by quantum mechanics and seek quantum generalizations of the Cramér-Rao bound for the mean-square error. The first quantum generalization of the Cramér-Rao bound is given in terms of Hermitian operators L_μ, called the symmetrized logarithmic derivatives (SLD) of $\hat{\rho}_\theta$ [7]. These operators satisfy

$$\begin{aligned} \frac{\partial \hat{\rho}_\theta}{\partial \theta_\mu} &= \frac{1}{2}(\hat{L}_\mu \hat{\rho}_\theta + \hat{\rho}_\theta \hat{L}_\mu), \\ \mathrm{tr}(\hat{\rho}_\theta \hat{L}_\mu) &= 0. \end{aligned} \tag{2.50}$$

The SLD quantum Cramér-Rao bound is given by

$$\begin{aligned} \Sigma &\geq J^{-1}(\hat{\rho}_\theta), \\ J_{\mu\nu}(\hat{\rho}_\theta) &= \frac{1}{2}\,\mathrm{tr}\left[\hat{\rho}_\theta(\hat{L}_\mu \hat{L}_\nu + \hat{L}_\nu \hat{L}_\mu)\right], \end{aligned} \tag{2.51}$$

where

$$\Sigma = \int (\check{\theta} - \theta)(\check{\theta} - \theta)^\top \,\mathrm{tr}[\hat{\rho}_\theta d\hat{E}(\check{\theta})] \tag{2.52}$$

is the mean-square estimation error matrix and $J(\hat{\rho}_\theta)$ is called the SLD quantum Fisher information matrix. The proof of the SLD quantum Cramér-Rao bound is analogous to the classical one, where we use the following Schwarz inequality for inner product of operators $\hat{A}$ and $\hat{B}$:

$$\Big(\int \mathrm{tr}\,\hat{\rho}_\theta \hat{A}^\dagger d\hat{E}(\check{\theta})\hat{A}\Big)\Big(\int \mathrm{tr}\,\hat{\rho}_\theta \hat{B}^\dagger d\hat{E}(\check{\theta})\hat{B}\Big) \geq |\int \mathrm{tr}\,\rho_\theta \hat{B}^\dagger d\hat{E}(\check{\theta})\hat{A}|^2. \tag{2.53}$$

To derive the quantum Cramér-Rao bound, we take derivatives on the unbiased estimator condition $\int \mathrm{tr}\left[\hat{\rho}_\theta(\check{\theta}_\nu - \theta_\nu)d\hat{E}(\check{\theta})\right] = 0$ with respect to the parameters of interest. Repeating the procedure in Sec. (2.3.2), the

result is

$$\begin{aligned} \Re\left(\int \mathrm{tr}\,\rho_\theta \hat{B}^\dagger d\hat{E}(\check{\theta})\hat{A}\right) &= a^\top b, \\ \hat{A} &= \sum_\mu a_\mu(\check{\theta_\mu} - \theta_\mu), \\ \hat{B} &= \sum_\mu b_\mu L_\mu, \end{aligned} \tag{2.54}$$

where $\Re$ denotes the real part of a complex number and a, b are real vectors. One then obtains the following inequality by applying the Schwarz inequality:

$$\Big(\int \mathrm{tr}\,\hat{\rho}_\theta \hat{A}^\dagger d\hat{E}(\check{\theta})\hat{A}\Big)\Big(\int \mathrm{tr}\,\rho_\theta \hat{B}^\dagger d\hat{E}(\check{\theta})\hat{B}\Big) = (a^\top \Sigma a)(b^\top J b) \geq (a^\top b)^2, \tag{2.55}$$

which is formally the same as Eq. (2.39). Eq. (2.51) follows by choosing $b = J^{-1}a$.

Instead of the SLD, we can also use the right logarithmic derivative (RLD) to define the RLD quantum Fisher information matrix [50, 51]:

$$J^{RLD}_{\mu\nu} = \mathrm{tr}(\hat{\rho}_\theta \hat{L}'_\mu \hat{L}'^\dagger_\nu). \tag{2.56}$$

where the RLD $\hat{L}'_\mu$ satisfies the equation

$$\frac{\partial \hat{\rho}_\theta}{\partial \theta_\mu} = \hat{\rho}_\theta \hat{L}'_\mu. \tag{2.57}$$

This will give us another quantum Cramér-Rao bound in terms of the RLD. In fact, we note that other generalizations of quantum Fisher information exist and each of them defines a quantum Cramér-Rao bound [52–54].

The SLD version of the Cramér-Rao bound is of particular interest to us. For single parameter problems, one can apply an adaptive quantum estimation scheme [52, 55]. The adaptive quantum estimation scheme is asymptotically efficient and consistent in the sense of saturating the SLD quantum Cramér-Rao bound. On the other hand, each POVM $\hat{E}$ defines

a probability measure $P_{\hat{E}}$ where the classical Fisher information is $j_{\hat{E}}$. A quantum Cramér-Rao bound states that [36, 53, 54]

$$J(\hat{\rho}_\theta) \geq j_{\hat{E}}, \tag{2.58}$$

which means that the SLD quantum Fisher information is the maximum amount of information extractable by any measurement. In this sense, the SLD quantum Fisher information represents a measurement-independent quantum limit of estimation for any unbiased estimator. Hence, another approach to saturate the SLD quantum Cramér-Rao bound is to find a POVM such that $j_{\hat{E}} = J(\hat{\rho}_\theta)$. Given that such a POVM exists, the MLE estimator can then attain the quantum Cramér-Rao bound asymptotically.

An equation similar to Eq. (2.45) exists in the quantum regime, where the Bures distance can be expanded infinitesimally in terms of the SLD Fisher information matrix [36, 56]:

$$D_B^2(\hat{\rho}_\theta, \hat{\rho}_{\theta+\epsilon u}) \equiv 2[1 - F(\hat{\rho}_\theta, \hat{\rho}_{\theta+\epsilon u})] = \frac{1}{4} u_\mu u_\nu J_{\mu\nu} \epsilon^2 + o(\epsilon), \tag{2.59}$$

where $F(\hat{\rho}_\theta, \hat{\rho}_{\theta'}) = \operatorname{tr} \sqrt{\sqrt{\hat{\rho}_\theta}\hat{\rho}_{\theta'}\sqrt{\hat{\rho}_\theta}}$ is the quantum fidelity and u is an arbitrary constant vector. Therefore, an equation analogous to Eq. (2.44) exists:

$$J_{\mu\nu} = -4\Big[\frac{\partial^2 F(\hat{\rho}_\theta, \hat{\rho}_{\theta'})}{\partial\theta_\mu \partial\theta_\nu}\Big]_{\theta'=\theta}. \tag{2.60}$$

This result is not too surprising considering that the quantum fidelity is equal to the Bhattacharyya coefficient when $\hat{\rho}_\theta$ and $\hat{\rho}_{\theta'}$ commutes.

2.5 Conclusion

In this chapter, we introduced the stochastic differential equation by starting from a discussion of the stochastic integration. The stochastic integration can be defined in terms of Riemann sums; however, unlike ordinary

calculus, different definitions converge differently. We discussed two different definitions of stochastic integration, namely the Stratonovich integral and the Itô integral. The Stratonovich calculus follows the same differentiation rules as the ordinary calculus, while the Itô calculus follows a distinct rule.

We also reviewed classical detection and estimation theory as a mathematical framework to reduce errors and to optimize decision processes. The likelihood ratio test is introduced as the optimal data processing strategy in both Bayesian and frequentist protocols. The quantum version of the likelihood ratio test is also discussed. Although the exact performance of likelihood ratio tests is difficult to calculate, bounds on the performance are available. For parameter estimation, we focused on estimation of nonrandom parameters and introduced the Fisher information as a figure of merit for the ultimate accuracy of unbiased estimators. We presented MLE as the estimator that is able to approach this ultimate accuracy. Connections between detection theory and estimation theory are also discussed. We provided a link between Fisher information and Bhattacharyya coefficient, and a similar relation in the quantum case. These relations will be useful in later chapters.

Chapter 3

Optimal signal processing for continuous qubit readout

Consider a quantum two-level system, or a qubit in modern terminology. According to von Neumann, measurement of a qubit can be instantaneous and perfectly accurate, with two possible outcomes and the qubit collapsing to a specific state depending on the outcome [10]. In practice, this measurement model, called a projective measurement, is an idealization. A qubit measurement in real physical systems, such as superconducting microwave circuits [57–59], trapped ions [60, 61], nitrogen-vacancy centers in diamond [62, 63], semiconductor quantum dots [64, 65], and phosphorus donors in silicon [66, 67], is often performed by coupling the qubit to an electromagnetic field, before the field is measured continuously. The qubit state can only be inferred with some degree of uncertainty from the noisy measurement. During the measurement, the qubit may also undergo spontaneous transitions, which further obscure the initial qubit state and complicate the inference procedure. This qubit readout problem is challenging but important for many quantum information processing applications, such as quantum computing [9], magnetometry [68], and atomic clocks [69, 70], which all require accurate measurements of qubits. The choice of a signal

processing method is crucial to the readout performance. Refs. [71, 72] in particular contain detailed theoretical studies of qubit-readout signal processing protocols.

In this chapter, we propose a new signal-processing architecture for optimal qubit readout by exploiting well known techniques in classical detection theory [32, 42, 43, 73, 74]. Following prior work [71, 72], we assume that the measurement is quantum nondemolition (QND) [10, 75], meaning that a classical stochastic theory is sufficient [10, 76, 77]. In addition to the Gaussian observation noise assumed in Refs. [71, 72], we also consider a Poissonian noise model [40], which is more suitable for photon-counting measurements [60–63, 65, 69]. We find that the likelihood ratio needed for optimal hypothesis testing can be determined from the celebrated estimator-correlator formulas [42, 43, 73, 74, 78], which break down the likelihood-ratio calculation into an estimator step and an easy correlator step. The estimator turns out to have analytic solutions in special cases of interest and simple numerical algorithms in general.

Although our protocols and the ones proposed in Refs. [71, 72] should result in the same end results for the likelihood ratio in the case of Gaussian noise, our analytic solutions involve elementary mathematical operations and may be implemented by low-latency electronics, such as analog or programmable logic devices [79], for fast feedback control and error correction purposes [10]. This is in contrast to the more complicated coupled stochastic differential equations recommended by the prior studies. Moreover, the prior studies never state whether their stochastic equations should be interpreted in the Itô sense or the Stratonovich sense, making it difficult for others to verify and correctly implement their protocols. As the equations are nonlinear with respect to the observation process, applying the wrong stochastic calculus is likely to give wrong results [37, 38, 40, 42]. Our work here, on the other hand, makes explicit and consistent use of Itô

calculus to ensure its correctness. Our estimator-correlator protocol is also inherently applicable to multi-hypothesis testing, which can be useful for online parameter estimation and making the readout robust against model uncertainties [80–84].

3.1 M-ary hypothesis testing

In this section we give a simple generalization of Bayesian binary hypothesis testing to Bayesian M-ary hypothesis testing. Let $\{\mathcal{H}_m; m = 0, 1, 2, \ldots, M-1\}$ be the hypotheses to be tested. Given a noisy observation record z, the decision region for hypothesis $\mathcal{H}_m$ is labeled by $\mathcal{Z}_m$. Define the observation probability measure as $P(.|\mathcal{H}_m)$, the average error probability is

$$P_e \equiv \sum_m P(\mathcal{H}_m) \int_{z \notin \mathcal{Z}_m} dP(z|\mathcal{H}_m). \tag{3.1}$$

Similarly to Sec. (2.2), to find the decision rule that minimizes P_e, we write

$$P_e = 1 - \sum_m \int_{\mathcal{Z}_m} P(\mathcal{H}_m)\Lambda(z|\mathcal{H}_m) dP(z|\mathcal{H}_0), \tag{3.2}$$

where we have defined

$$\Lambda(z|\mathcal{H}_m) \equiv \frac{dP(z|\mathcal{H}_m)}{dP(z|\mathcal{H}_0)} \tag{3.3}$$

as the likelihood ratio for $\mathcal{H}_m$ against $\mathcal{H}_0$, the null hypothesis. The decision rule that minimizes P_e is to select $z \in \mathcal{Z}_m$ whenever

$$P(\mathcal{H}_m)\Lambda(z|\mathcal{H}_m) > P(\mathcal{H}_n)\Lambda(z|\mathcal{H}_n), \quad \forall n \neq m. \tag{3.4}$$

This is equivalent to choosing the hypothesis that maximizes the posterior probability function [32, 85], which can be expressed as

$$P(\mathcal{H}_m|z) = \frac{P(\mathcal{H}_m)\Lambda(z|\mathcal{H}_m)}{\sum_m P(\mathcal{H}_m)\Lambda(z|\mathcal{H}_m)}, \tag{3.5}$$

The minimum-error decision strategy thus boils down to the computation of $\Lambda(z|\mathcal{H}_m)$ for all hypotheses of interest, and then finding the hypothesis that maximizes $P(\mathcal{H}_m|z)$, or equivalently in terms of the log-likelihood ratio (LLR)

$$\mathcal{Z}_m = \{z : \ln\Lambda(z|\mathcal{H}_m) + \ln P(\mathcal{H}_m) > \ln\Lambda(z|\mathcal{H}_n) + \ln P(\mathcal{H}_n), \quad \forall n \neq m\}. \tag{3.6}$$

3.2 Gaussian noise model

3.2.1 Observation process

Assume that the observation process $z(t)$ conditioned on a hypothesis is

$$\mathcal{H}_m : \; z(t) = S_m(t)x_m(t) + \xi(t), \tag{3.7}$$

where $S_m(t)$ is a deterministic signal amplitude assumed by the hypothesis, $x_m(t)$ is a hidden stochastic process, $\xi(t)$ is a zero-mean white Gaussian noise with covariance

$$\mathbb{E}\left[\xi(t)\xi(t')\right] = R(t)\delta(t-t'), \tag{3.8}$$

$\mathbb{E}$ denotes expectation, and $R(t)$ is the noise power, assumed here to be the same for all hypotheses. It is possible to test other values of noise power by prescaling the observation and redefining $S_m(t)$. For qubit readout, the hypothesis should determine $S_m(t)$ and the statistics of $x_m(t)$; Fig. 3.1 sketches a few example realizations of the signal component $S_m(t)x_m(t)$.

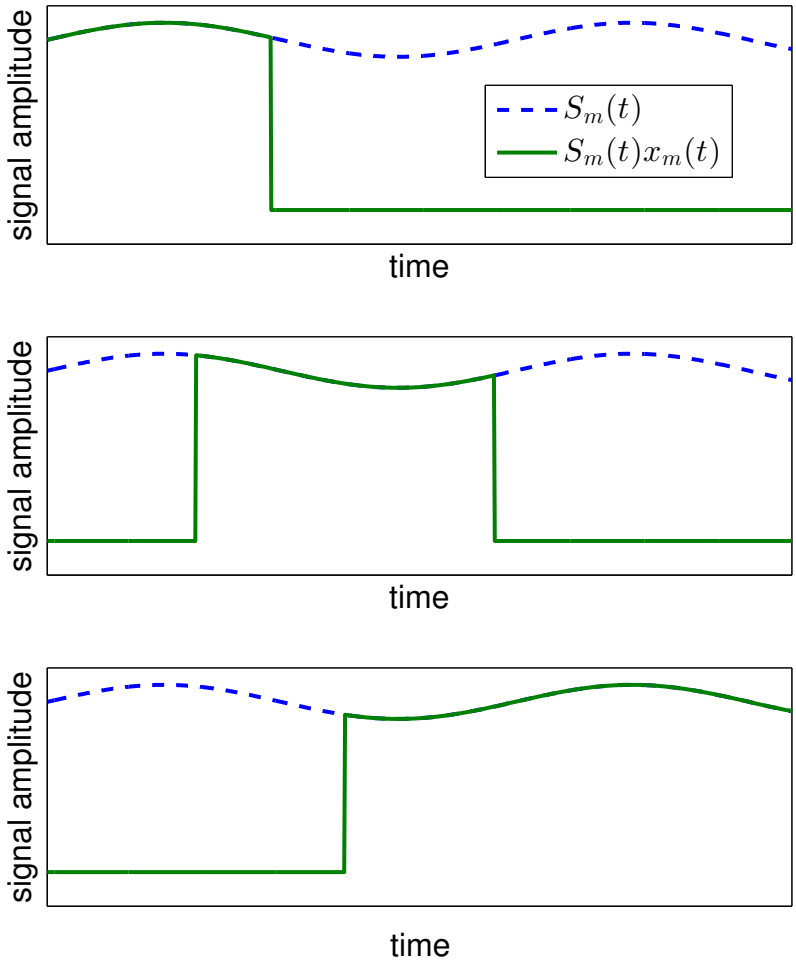


Figure 3.1: (Color online) Some example realizations of the signal component $S_m(t)x_m(t)$ of the observation process. Given a hypothesis $\mathcal{H}_m$, $S_m(t)$ is a deterministic signal amplitude and $x_m(t)$ is a binary stochastic process. The axes are in arbitrary units.

In stochastic detection theory, it is convenient to define a normalized observation process $y(t)$ as the time integral of $z(t)$:

$$y(t) \equiv \int_0^t d\tau \frac{z(\tau)}{\sqrt{R(\tau)}}, \tag{3.9}$$

and represent it using a stochastic differential equation:

$$\begin{aligned} \mathcal{H}_m : \; dy(t) &\equiv y(t+dt) - y(t) \\ &= dt\sigma_m(t)x_m(t) + dW(t), \qquad (3.10) \\ \sigma_m(t) &\equiv \frac{S_m(t)}{\sqrt{R(t)}}, \qquad (3.11) \end{aligned}$$

where $W(t)$ is the standard Wiener process with increment variance $dW^2(t) = dt$ and Itô calculus is assumed throughout this chapter. The null hypothesis, in particular, is taken to be

$$\mathcal{H}_0 : dy(t) = dW(t). \tag{3.12}$$

Fig. 3.2 depicts the observation model through a block diagram.

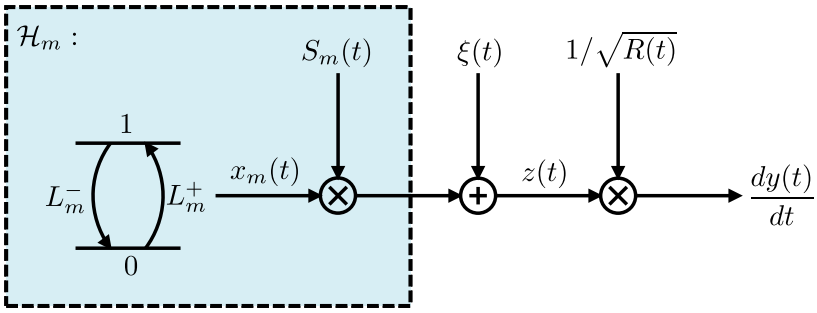


Figure 3.2: (Color online) A block diagram for the observation model. $\mathcal{H}_m$ is a hypothesis, $x_m(t)$ is the hidden signal, assumed here to be a two-state Markov process with transition rates L_m^- and L_m^+, $S_m(t)$ is the signal amplitude, $\xi(t)$ is an additive white Gaussian noise, and $z(t)$ is the observation process. The definition of observation processes $dy(t)/dt$ and $y(t)$, normalized with respect to the noise power $R(t)$, is for mathematical convenience.

3.2.2 Estimator-correlator formula

Define the observation record as

$$Y^T \equiv \{y(t); 0 \leq t \leq T\}\,. \tag{3.13}$$

Under rather general conditions about $x_m(t)$, the LLR $\ln \Lambda(Y^T|\mathcal{H}_m)$ can be expressed using the estimator-correlator formula [42, 43, 73, 78], which correlates the observation with an "assumptive" estimate $\mu_m(t)$:

$$\ln \Lambda(Y^T|\mathcal{H}_m) = \int_0^T dy(t)\mu_m(t) - \frac{1}{2}\int_0^T dt\mu_m^2(t), \tag{3.14}$$

where

$$\mu_m(t) \equiv \sigma_m(t)\mathbb{E}\left[x_m(t)|Y^t, \mathcal{H}_m\right] \tag{3.15}$$

is a causal estimator of the hidden signal conditioned on the observation record Y^t and the hypothesis $\mathcal{H}_m$. The integral with respect to $y(t)$ is an Itô integral, meaning that $dy(t)$ is the future increment ahead of time t and $\mu_m(t)$ in the integrand $dy(t)\mu_m(t)$ should not depend on $dy(t)$. This rule is important for consistent analytic and numerical calculations whenever one multiplies $dy(t)$ with a signal that depends on $y(t)$ [42]. Fig. 3.3 illustrates an implementation of the formula.

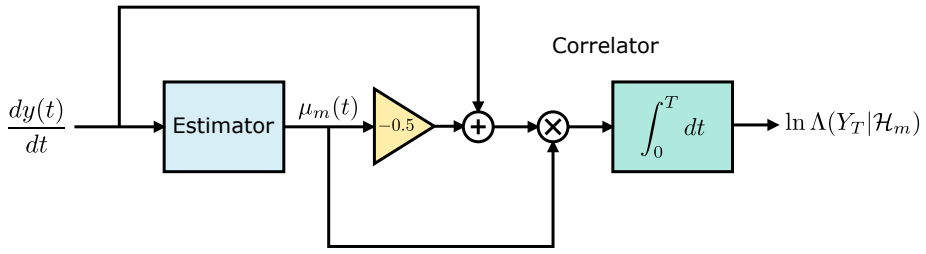


Figure 3.3: (Color online) An implementation of the estimator-correlator formula in Eq. (3.14), which can be written as $\ln\Lambda = \int_0^T [dy(t) - dt\mu_m(t)/2]\mu_m(t)$. $dy(t)$ in the integrand should be the future increment ahead of t in accordance with Itô calculus.

As each $\ln \Lambda(Y^T|\mathcal{H}_m)$ depends only on one hypothesis $\mathcal{H}_m$ (in addition to the fixed null hypothesis), once an algorithm for its computation is implemented, it can be re-used even if the other hypotheses are changed or new hypotheses are added. This makes the estimator-correlator protocol more flexible and extensible than the ones proposed in Refs. [71, 72], which are specific to the hypotheses considered there.

Despite its simple appearance, the formula does not in general reduce the complexity of the LLR calculation, as the estimator may still be difficult to implement. We shall, however, present a simple numerical method and some analytic solutions useful for the qubit readout problem in the following.

3.2.3 Qubit dynamics

For QND qubit readout, we assume that $x_m(t)$ is a classical two-state first-order Markov process; Appendix A shows explicitly how the classical theory can arise from the quantum formalism of continuous QND measurement. The possible values of $x_m(t)$ are assumed to be

$$x_m(t) \in \{0, 1\} \,. \tag{3.16}$$

Other possibilities can be modeled by subtracting a baseline value from the actual observation and defining an appropriate $\sigma_m(t)$ before the processing described here. In the absence of measurements, the probability function

of $x_m(t) = x$ obeys a forward Kolmogorov equation [37]:

$$\frac{d\boldsymbol{P}_m(t)}{dt} = \boldsymbol{L}_m(t)\boldsymbol{P}_m(t), \tag{3.17}$$

$$\boldsymbol{P}_m(t) \equiv \begin{pmatrix} P(x=0,t|\mathcal{H}_m) \\ P(x=1,t|\mathcal{H}_m) \end{pmatrix}, \tag{3.18}$$

$$\boldsymbol{L}_m(t) \equiv \begin{pmatrix} -L_m^+(t) & L_m^-(t) \\ L_m^+(t) & -L_m^-(t) \end{pmatrix}, \tag{3.19}$$

where L_m^- and L_m^+ are the spontaneous decay and excitation rates conditioned on the hypothesis and can be time-varying for generality. The decay time constant $1/L_m^-$ is commonly called T_1, and L_m^+ can be used to model a random turn-on time [72]. For example, we can model the problem studied by Gambetta and coworkers [71] by defining

- $\mathcal{H}_0$: the qubit is in the $x = 0$ state, and $x_0(t) = 0$.
- $\mathcal{H}_1$: the qubit is in the $x = 1$ state initially, $P(x = 1, t = 0|\mathcal{H}_1) = 1$, and the unconditional statistics of $x_1(t)$ obey Eqs. (3.17)–(3.19), with L_1^- being the decay rate and $L_1^+ = 0$.

3.2.4 Estimator

The estimator $\mu_m(t)$ can be computed using the Duncan-Mortensen-Zakai (DMZ) equation [86–89]:

$$d\boldsymbol{p}_m(t) = dt\boldsymbol{L}_m(t)\boldsymbol{p}_m(t) + dy(t)\sigma_m(t)\boldsymbol{x}\boldsymbol{p}_m(t), \tag{3.20}$$

$$\boldsymbol{p}_m(t) \equiv \begin{pmatrix} p_m(x=0,t) \\ p_m(x=1,t) \end{pmatrix}, \quad \boldsymbol{x} \equiv \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.21}$$

where

$$p_m(x,t) \propto P(x,t|Y^t,\mathcal{H}_m) \tag{3.22}$$

is the unnormalized posterior probability function of $x_m(t)$ conditioned on Y^t and $\mathcal{H}_m$, and the initial condition is determined by the initial prior probabilities:

$$p_m(x, t = 0) = P(x, t = 0|\mathcal{H}_m). \tag{3.23}$$

The estimator is then

$$\mu_m(t) = \frac{\sigma_m(t)p_m(1,t)}{p_m(0,t) + p_m(1,t)}, \tag{3.24}$$

as depicted by Fig. 3.4.

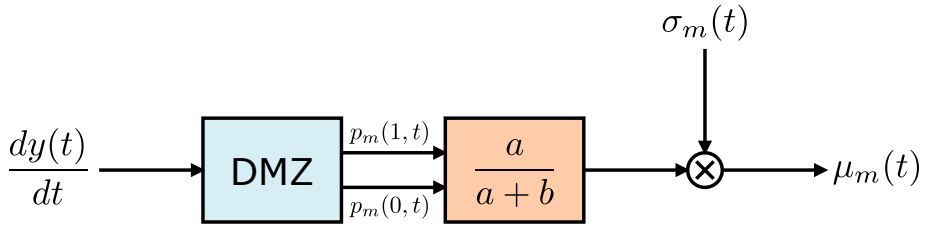


Figure 3.4: (Color online) A block diagram for the estimator using the Duncan-Mortensen-Zakai (DMZ) equation.

Although one can also use the Wonham equation [90] to perform the estimator, and the normalization step would not be needed in theory, the DMZ equation is linear with respect to $\boldsymbol{p}_m(t)$ and easier to solve analytically or numerically. In general, a numerical split-step method can be used [91]:

$$\boldsymbol{p}_m(t + dt) \approx \exp\left[dy(t)\sigma_m(t)\boldsymbol{x} - \frac{dt}{2}\sigma_m^2(t)\boldsymbol{x}^2 + dt\boldsymbol{L}_m(t)\right]\boldsymbol{p}_m(t). \tag{3.25}$$

Many other numerical methods are available [92]. Analytic solutions can be obtained in the following special cases.

3.2.5 Deterministic-signal detection

For a simple example, assume binary hypothesis testing ($M = 2$), no spontaneous transition ($L_m^- = L_m^+ = 0$), and deterministic initial conditions

given by

$$p_0(0,0) = P(x=0, t=0|\mathcal{H}_0) = 1, \tag{3.26}$$

$$p_0(1,0) = P(x=1, t=0|\mathcal{H}_0) = 0, \tag{3.27}$$

$$p_1(0,0) = P(x=0, t=0|\mathcal{H}_1) = 0, \tag{3.28}$$

$$p_1(1,0) = P(x=1, t=0|\mathcal{H}_1) = 1. \tag{3.29}$$

The estimator becomes independent of the observation:

$$\mu_0(t) = 0, \qquad \mu_1(t) = \sigma_1(t). \tag{3.30}$$

This is simply a case of deterministic-signal detection, when the estimator-correlator formula in Eq. (3.14) becomes a matched filter [32, 42]. The minimum error probability $P_{e,\min}$ has an analytic expression [32]:

$$P_{e,\min} = P_+ P(\mathcal{H}_0) + P_- P(\mathcal{H}_1), \tag{3.31}$$

$$P_\pm \equiv \frac{1}{2} \operatorname{erfc}\left[\sqrt{\frac{\mathrm{SNR}}{8}}\left(1 \pm \frac{2\lambda}{\mathrm{SNR}}\right)\right], \tag{3.32}$$

$$\operatorname{erfc} u \equiv \frac{2}{\sqrt{\pi}} \int_u^\infty dv \exp(-v^2), \tag{3.33}$$

$$\mathrm{SNR} \equiv \int_0^T dt \sigma_1^2(t), \quad \lambda \equiv \ln \frac{P(\mathcal{H}_1)}{P(\mathcal{H}_0)}. \tag{3.34}$$

For $\mathrm{SNR} \to \infty$, the error exponent has the asymptotic behavior $-\ln P_{e,\min} \to \mathrm{SNR}/8$.

Although this solution for $P_{e,\min}$ is not strictly valid when spontaneous transitions are present, it should be accurate when the observation time T is short relative to $1/L_1^-$ or $1/L_1^+$ and can serve as a rough guide for other cases.

3.2.6 No spontaneous excitation ($L_m^+ = 0$)

The case of $L_m^- > 0$ and $L_m^+ = 0$ corresponds to the model studied by Gambetta and coworkers [71]. Eq. (3.20) becomes

$$dp_m(0,t) = dtL_m^-(t)p_m(1,t), \tag{3.35}$$

$$dp_m(1,t) = -dtL_m^-(t)p_m(1,t) + dy(t)\sigma_m(t)p_m(1,t). \tag{3.36}$$

Eq. (3.36) describes the famous geometric Brownian motion [38]. Its well known solution can be obtained by applying Itô's lemma (2.8) to $d\ln p_m(1,t)$:

$$\begin{aligned} d\ln p_m(1,t) &= \frac{dp_m(1,t)}{p_m(1,t)} - \frac{[dp_m(1,t)]^2}{2p_m^2(1,t)} \\ &= -dt\Big[\frac{\sigma_m^2(t)}{2} + L\Big] + dy(t)\sigma_m(t). \end{aligned}$$

Integrating the above differential and upon exponentiation, the solution is given by

$$\begin{aligned} p_m(1,t) = p_m(1,0)\exp\Bigg\{&\int_0^t dy(\tau)\sigma_m(\tau) \\ &- \int_0^t d\tau\left[\frac{\sigma_m^2(\tau)}{2} + L_m^-(\tau)\right]\Bigg\}. \end{aligned} \tag{3.37}$$

A time integral of $p_m(1,t)$ then gives $p_m(0,t)$:

$$p_m(0,t) = p_m(0,0) + \int_0^t d\tau L_m^-(\tau)p_m(1,\tau). \tag{3.38}$$

For binary qubit state discrimination, we can assume that $\mu_0(t) = 0$, and $\mu_1(t)$ can be determined from Eqs. (3.37), (3.38), and (3.24), starting from the deterministic initial conditions given by Eqs. (3.28) and (3.29) if the measurement starts immediately after the qubit state is prepared, as shown in Fig. 3.5. If there is a finite arming time before the measurement starts [71, 72], the forward Kolmogorov equation (3.17) can be used to determine

the initial state probabilities.

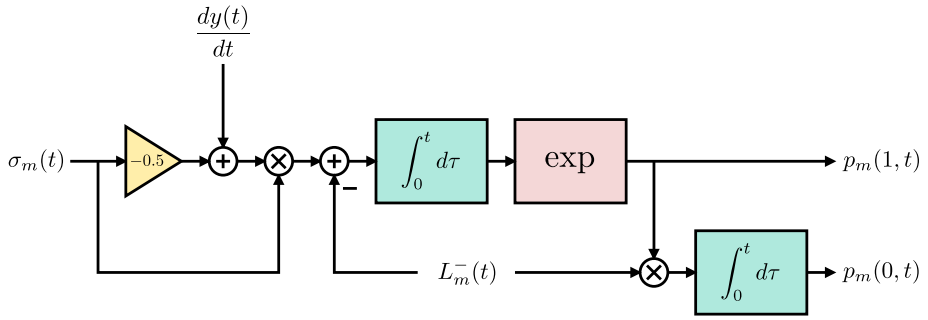


Figure 3.5: (Color online) Solution to the DMZ equation with spontaneous decay ($L_m^- > 0$), no spontaneous excitation ($L_m^+ = 0$), and an initial excited state ($p_m(1, t = 0) = 1$, $p_m(0, t = 0) = 0$).

3.2.7 No spontaneous decay ($L_m^- = 0$)

One can assume $L_m^+ > 0$ and $L_m^- = 0$ to model a random signal turn-on time [72] and negligible spontaneous decay ($T \ll 1/L_m^-$). The simplest way of computing $\mu_m(t)$ is to define a new observation process

$$\begin{aligned} dy'(t) &\equiv dy(t) - \sigma_m(t)dt \\ &= -dt\sigma_m(t)\left[1 - x_m(t)\right] + dW(t). \end{aligned} \tag{3.39}$$

A new DMZ equation can then be expressed in terms of $y'(t)$ and is given by

$$dp_m(0,t) = -dtL_m^+(t)p_m(0,t) - dy'(t)\sigma_m(t)p_m(0,t), \tag{3.40}$$

$$dp_m(1,t) = dtL_m^+(t)p_m(0,t), \tag{3.41}$$

which have the same form as Eqs. (3.35) and (3.36) and can be solved using the same method. The final solution is

$$p_m(0,t) = p_m(0,0)\exp\left\{-\int_0^t dy(\tau)\sigma_m(\tau) + \int_0^t d\tau\left[\frac{\sigma_m^2(\tau)}{2} - L_m^+(\tau)\right]\right\}, \tag{3.42}$$

$$p_m(1,t) = p_m(1,0) + \int_0^t d\tau L_m^+(\tau)p_m(0,\tau). \tag{3.43}$$

3.3 Poissonian noise model

3.3.1 Observation process

For photon-counting measurements, it is more appropriate to assume that the counting process $n(t) \in \{0,1,2,\dots\}$, conditioned on the hidden process $X_m^t \equiv \{x_m(\tau); 0 \le \tau \le t\}$, obeys Poissonian statistics [40]:

$$P(n(t)|X_m^t, \mathcal{H}_m) = \exp\left[-\int_0^t d\tau\lambda_m(\tau)\right]\frac{\left[\int_0^t d\tau\lambda_m(\tau)\right]^{n(t)}}{n(t)!}, \tag{3.44}$$

where

$$\lambda_m(t) \equiv \lambda_0(t)\left[1 + \alpha_m(t)x_m(t)\right] \tag{3.45}$$

is the intensity of the Poisson process and $\alpha_m(t)$ is a deterministic signal amplitude. $dn(t) \in \{0,1\}$ is then the detected photon number at time t. We assume $\mathcal{H}_0$ with known intensity $\lambda_0(t) > 0$ to be the null hypothesis.

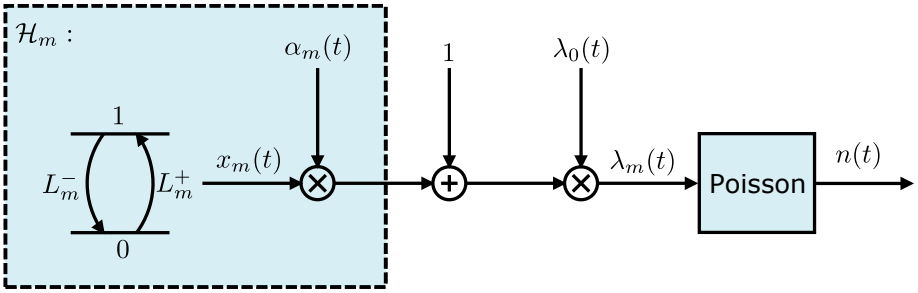


Figure 3.6: (Color online) The Poissonian observation model. The counting process $n(t)$ is driven by the stochastic intensity $\lambda_m(t)$.

3.3.2 Estimator-correlator formula

Define the observation record as

$$N^T \equiv \{n(t); t_0 \leq t \leq T\}\,. \tag{3.46}$$

Our goal is to calculate the LLR

$$\ln \Lambda(N^T|\mathcal{H}_m) = \ln \frac{dP(N^T|\mathcal{H}_m)}{dP(N^T|\mathcal{H}_0)}. \tag{3.47}$$

A formula analogous to the Gaussian case in Eq. (3.14) is given by [74, 78]

$$\begin{aligned} \ln \Lambda(N^T|\mathcal{H}_m) = &\int_0^T dn(t) \ln\left[1 + \nu_m(t)\right] \\ &- \int_0^T dt \lambda_0(t) \nu_m(t), \end{aligned} \tag{3.48}$$

$$\nu_m(t) \equiv \alpha_m(t) \mathbb{E}\left[x_m(t)|N^t, \mathcal{H}_m\right], \tag{3.49}$$

where the $dn(t)$ integral should again follow Itô's convention [40]. Fig. 3.7 illustrates the formula.

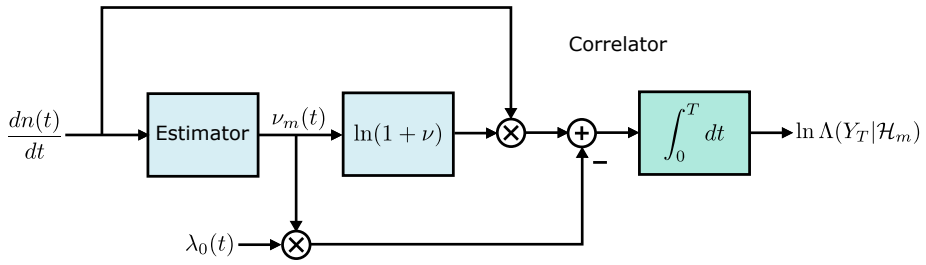


Figure 3.7: (Color online) The estimator-correlator structure for the Poissonian observation model. $dn(t)$ should be the future increment ahead of t when multiplied with $\ln[1+\nu_m(t)]$.

3.3.3 Estimator

We assume the same unconditional qubit dynamics described in Sec. 3.2.3.

The estimator can be computed from a DMZ-type equation [78, 89]:

$$\begin{aligned} d\boldsymbol{p}_m(t) = dt\boldsymbol{L}_m(t)\boldsymbol{p}_m(t) + [dn(t) - dt\kappa(t)] \\ \times \left\{ \frac{\lambda_0(t)}{\kappa(t)} [\boldsymbol{I} + \alpha_m(t)\boldsymbol{x}] - \boldsymbol{I} \right\} \boldsymbol{p}_m(t), \end{aligned} \tag{3.50}$$

where $\kappa(t) > 0$ is an arbitrary positive reference intensity and the estimator is

$$\nu_m(t) = \frac{\alpha_m(t) p_m(1,t)}{p_m(0,t) + p_m(1,t)}. \tag{3.51}$$

This procedure is identical to that depicted in Fig. 3.4. Assuming $\kappa(t) = \lambda_0(t)$, Eq. (3.50) can be solved using a numerical split-step method:

$$\begin{aligned} &\boldsymbol{p}_m(t+dt) \\ &\approx \exp\left\{ dn(t) \ln\left[\boldsymbol{I} + \alpha_m(t)\boldsymbol{x}\right] - dt\lambda_0(t)\alpha_m(t)\boldsymbol{x} \right\} \\ &\quad \times \exp\left[dt\boldsymbol{L}_m(t)\right] \boldsymbol{p}_m(t). \end{aligned} \tag{3.52}$$

Analytic solutions can be found in the following cases.

3.3.4 No spontaneous excitation ($L_m^+ = 0$)

Let $\kappa(t) = \lambda_0(t)$. Eq. (3.50) becomes

$$dp_m(0,t) = dtL_m^-(t)p_m(1,t), \tag{3.53}$$

$$dp_m(1,t) = -dtL_m^-(t)p_m(1,t) + [dn(t) - dt\lambda_0(t)]\,\alpha_m(t)p_m(1,t). \tag{3.54}$$

Following Chap. 5.3.1 in Ref. [40], we get

$$p_m(1,t) = p_m(1,0)\exp\left\{\int_0^t dn(\tau)\ln\left[1+\alpha_m(\tau)\right] - \int_0^t d\tau\left[\lambda_0(\tau)\alpha_m(\tau) + L_m^-(\tau)\right]\right\}, \tag{3.55}$$

$$p_m(0,t) = p_m(0,0) + \int_0^t d\tau L_m^-(\tau)p_m(1,\tau). \tag{3.56}$$

Fig. 3.8 depicts a block diagram for this solution.

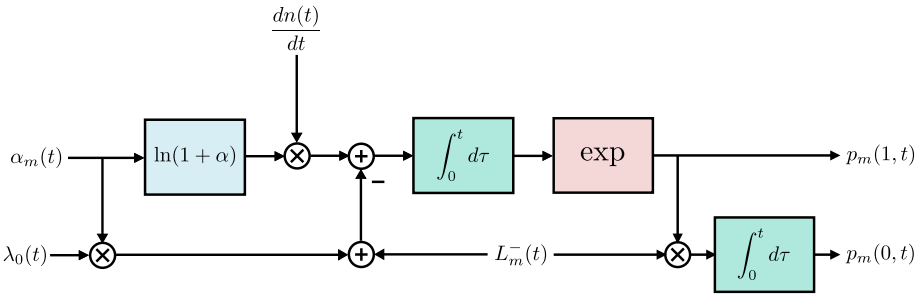


Figure 3.8: (Color online) A block diagram for Eqs. (3.55) and (3.56), a solution to the Poissonian DMZ equation. $L_m^+ = 0$, $p_m(1, t = 0) = 1$, and $p_m(0, t = 0) = 0$ are assumed.

3.3.5 No spontaneous decay ($L_m^- = 0$)

We now let $\kappa(t) = \lambda_0(t)\,[1 + \alpha_m(t)]$. Eq. (3.50) becomes

$$\begin{aligned} dp_m(0,t) &= -dtL_m^+(t)p_m(0,t) \\ &\quad - [dn(t) - dt\kappa(t)]\,\frac{\alpha_m(t)}{1+\alpha_m(t)}p_m(0,t), \end{aligned} \tag{3.57}$$

$$dp_m(1,t) = dtL_m^+(t)p_m(0,t). \tag{3.58}$$

Similar to the previous case, the solution is

$$\begin{aligned} p_m(0,t) = p_m(0,0)\exp\Bigg\{&-\int_0^t dn(\tau)\ln\left[1+\alpha_m(\tau)\right] \\ &+\int_0^t d\tau\left[\lambda_0(\tau)\alpha_m(\tau) - L_m^+(\tau)\right]\Bigg\}, \end{aligned} \tag{3.59}$$

$$p_m(1,t) = p_m(1,0) + \int_0^t d\tau L_m^+(\tau)p_m(0,\tau). \tag{3.60}$$

It is interesting to note that all the Poissonian results approach the Gaussian ones in Sec. 3.2 if we assume $dn = \sqrt{\lambda_0}dy + \lambda_0 dt$, $\alpha_m = \sigma_m/\sqrt{\lambda_0}$, and $\lambda_0 \to \infty$.

3.4 Conclusion

We have proposed an estimator-correlator architecture for optimal qubit-readout signal processing and found analytic solutions in some special cases of interest using Itô calculus. Although we have focused on a classical model, our formalism can potentially be extended to more general quantum dynamics [78, 93] and more realistic measurements, including artifacts such as dark counts and finite detector bandwidth [10]. An open problem of interest is the evaluation of readout performance beyond the case of deterministic-signal detection. Numerical Monte Carlo simulation is not difficult for two-level systems, but analytic solutions should bring additional insight and may be possible using tools in classical and quantum

detection theory [7, 32, 94–97]. Tsang [98] has found an upper bound on the minimum error probability based on the study of non-optimal Volterra filters, which should provide insight to better optimize the readout process.

Another open problem is the accuracy, speed, and practicality of our algorithms in reality, which will be subject to more specific experimental requirements and hardware limitations [79]. In terms of speed, D'Anjou *et al.* [99] proposed an adaptive decision scheme to speed up the readout process. Instead of a likelihood ratio test at the end of an experiment, their adaptive scheme involves setting an upper and lower threshold for the likelihood ratio, and stopping the experiment once the likelihood ratio reaches either one of the threshold. The decision then depends on whether the likelihood ratio reaches the upper or lower threshold. D'Anjou *et al.* have found that the average experimental duration can be as low as half of the duration for the non-adaptive case, while maintaining the same error probability.

Chapter 4

Spectrum analysis with quantum dynamical systems

Recent technological advances, especially in optomechanics [21], suggest that quantum noise will soon be the major limiting factor in many metrological applications [75]. Many tasks in optomechanics force sensing, including thermometry, estimation of stochastic gravitational-wave background [100, 101], and testing spontaneous wavefunction collapse [102, 103], involve the spectrum analysis of a stochastic force, and the effect of quantum noise on such tasks has been of recent interest [102, 103]. 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. [104–108]), their application to more complex dynamical systems with broadband measurements such as optomechanics remains unclear.

In this chapter, we propose a theoretical framework of spectrum-parameter estimation with quantum dynamical systems, proving fundamental limits and investigating measurement and data analysis techniques that approach the limits. An outstanding feature of our work is the simple analytic results

in terms of basic power spectral densities (PSDs) in the problem, 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 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 for gravitational-wave astronomy [26, 101, 109] and optical sensing in general.

4.1 Quantum metrology

4.1.1 Parameter estimation

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. Assume that the prior probability measure of $X(t)$ depends on a vector of unknown parameters θ. Let Y be the quantum measurement outcome and $\check{\theta}(Y)$ be an estimator of θ using Y. The central error figure of interest is the mean-square estimation error matrix, defined as

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

with $\mathbb{E}_Y$ denoting the expectation over the random variable Y. Our goal here is to compute analytic results concerning Σ and discover quantum measurement techniques that can accurately estimate θ.

For any unbiased estimator ($\mathbb{E}_Y(\hat{\theta}) = \theta$), the multiparameter Cramér-

Rao bound states that

$$\Sigma \geq j^{-1}(p_Y), \tag{4.2}$$

where $j(p_Y)$ is the classical Fisher information matrix for a given likelihood function p_Y [32]. For a quantum system, let $\hat{\rho}(\theta)$ be a θ-dependent density operator and $\hat{E}(y)$ be the POVM that models the measurement[1], such that

$$p_Y(y|\theta) = \mathrm{tr}\left[\hat{E}(y)\hat{\rho}(\theta)\right], \tag{4.3}$$

with tr being the operator trace. For dynamical systems, $\hat{\rho}(\theta)$ can be obtained using the principles of purification and deferred measurements [9, 95, 110, 111]. For the purpose of spectrum-parameter estimation, we model $\hat{\rho}$ as

$$\hat{\rho}(\theta) = \mathbb{E}_{X|\theta}\left\{\hat{U}[X,T]|\psi\rangle\langle\psi|\hat{U}^\dagger[X,T]\right\}, \tag{4.4}$$

where

$$\hat{U}[X,T] = \mathcal{T}\exp\left\{-\frac{i}{\hbar}\int_0^T dt\hat{H}[X,t]\right\} \tag{4.5}$$

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 hyperparameters in this context [112]. For any POVM, a quantum Cramér-Rao bound states that

$$j(p_Y) \leq J(\hat{\rho}), \tag{4.6}$$

where $J(\hat{\rho})$ is the quantum Fisher information matrix with respect to the

[1]The definition of POVM here is slightly different than in the introductory chapter. Here, we have $\int_{\Omega'}\hat{E}(y)dy = \hat{I}$

symmetric logarithmic derivatives of $\hat{\rho}$ [6, 7, 36].

4.1.2 Extended convexity

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. [113–115]), J is much more difficult to evaluate analytically for multimode high-dimensional dynamical systems under continuous measurements. To proceed, we exploit a recently discovered property of J known as the extended convexity [116], which states that

$$J(\hat{\rho}) \leq \mathcal{J}\left\{\hat{\sigma}, p_Z\right\} \equiv \mathbb{E}_{Z|\theta}\left[J\left(\hat{\sigma}\right)\right] + j(p_Z), \tag{4.7}$$

where $\{\hat{\sigma}, p_Z\}$ is any ensemble of $\hat{\rho}$ with elements $\hat{\sigma}$ and likelihood function p_Z such that $\hat{\rho}(\theta) = \mathbb{E}_{Z|\theta}[\hat{\sigma}(Z|\theta)]$.

The proof of extended convexity $J \leq \mathcal{J}$ for one parameter in Ref. [116] relies on the assumption that there exists an optimal POVM attaining $j = J$. Such an assumption is questionable however [117], and here we use instead the strong concavity of Uhlmann fidelity [9] to prove Eq. (4.7) for multiple parameters. Let $\{\hat{\sigma}, p_Z\}$ be an ensemble for $\hat{\rho}(\theta)$ such that

$$\hat{\rho}(\theta) = \int dz p_Z(z|\theta)\hat{\sigma}(z|\theta). \tag{4.8}$$

Define the Uhlmann fidelity as

$$F[\hat{\rho}, \hat{\rho}'] \equiv \mathrm{tr}\sqrt{\sqrt{\hat{\rho}}\hat{\rho}'\sqrt{\hat{\rho}}}. \tag{4.9}$$

The strong concavity states that [9]

$$F\left[\hat{\rho}(\theta), \hat{\rho}(\theta')\right] \geq \int dz \sqrt{p_Z(z|\theta)p_Z(z|\theta')} \times F[\hat{\sigma}(z|\theta), \hat{\sigma}(z|\theta')]. \tag{4.10}$$

To relate F to J, we use Eq. (2.59)

$$F[\hat{\rho}(\theta), \hat{\rho}(\theta + \epsilon u)] = 1 - \frac{\epsilon^2}{8} u_\mu J_{\mu\nu}(\hat{\rho}) u_\nu + o(\epsilon^2), \tag{4.11}$$

where ϵ is a scalar, u is any real vector with the same dimension as θ (Einstein summation is assumed throughout this chapter), and $o(\epsilon^2)$ denotes terms asymptotically smaller than ϵ^2. Next we use the equivalent result in classical statistics Eq. (2.45),

$$\int dz \sqrt{p_Z(z|\theta) p_Z(z|\theta + \epsilon u)} = 1 - \frac{\epsilon^2}{8} u_\mu j_{\mu\nu}(p_Z) u_\nu + o(\epsilon^2). \tag{4.12}$$

Expanding $F[\hat{\rho}(\theta), \hat{\rho}(\theta')]$ and $F[\hat{\sigma}(z|\theta), \hat{\sigma}(z|\theta')]$ in Eq. (4.10) using Eq. (4.11), applying Eq. (4.12) to the right-hand side of Eq. (4.10), and comparing the ϵ^2 terms on both sides, we obtain

$$u_\mu J_{\mu\nu}(\hat{\rho}) u_\nu \leq u_\mu \left\{ \mathbb{E}_{Z|\theta} \left[J_{\mu\nu}(\hat{\sigma}) \right] + j_{\mu\nu}(p_Z) \right\} u_\nu. \tag{4.13}$$

Since this holds for any u, we obtain the matrix inequality in Eq. (4.7). The classical simulation technique proposed in Ref. [118] can be regarded as a special case of extended convexity when $J(\hat{\sigma}) = 0$.

4.1.3 Dynamical systems

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

$$S_X(\omega|\theta) \equiv \int_{-\infty}^{\infty} d\tau \mathbb{E}_{X|\theta}[X(t) X(t + \tau)] \exp(i\omega\tau). \tag{4.14}$$

For the quantum system, we assume that the Hamiltonian is of the form

$$\hat{H} = \hat{H}_0 - \hat{Q}X(t), \tag{4.15}$$

where $\hat{Q}$ is the quantum generator and $\hat{H}_0$ is the rest of the Hamiltonian. For example, $X(t)$ can be the classical force on a mechanical oscillator and $\hat{Q}$ can be the quantum position operator, as depicted in Fig. 4.1(a).

A modified purification technique can transform the problem in the interaction picture and produce an alternative and possibly tighter bound in terms of the optical statistics alone [111]. For an optomechanical system, the Hamiltonian is of the form [21]

$$\hat{H}_{\mathrm{OM}} = \hat{H}_{\mathrm{M}} + \hat{H}_{\mathrm{O}} + \hat{h}, \tag{4.16}$$

where $\hat{H}_{\mathrm{M}}$ is the mechanical Hamiltonian, $\hat{H}_{\mathrm{O}}$ is the optical Hamiltonian, and $\hat{h}$ is the optomechanical interaction Hamiltonian. For example, if the mechanical oscillator with position operator $\hat{q}$ interacts with one cavity optical mode with photon-number operator $\hat{n}$, $\hat{h} = -\hbar g_0 \hat{n}\hat{q}$, where g_0 is a coupling constant. A classical force $f(t)$ on the mechanical oscillator leads to a term $-\hat{q}f(t)$ in $\hat{H}_{\mathrm{M}}$, and if we assume $\hat{U}$ to be the time-ordered exponential of $\hat{H}_{\mathrm{OM}}$, $f(t)$ can be regarded as the hidden process and $\hat{q}$ the generator.

In practice, measurements are made on the optics and not the mechanics directly, so one is free to modify the purification [119] by applying any mechanical unitary to the optomechanical one [111]. To be specific, let $\hat{U}_{\mathrm{OM}}$ be the time-ordered exponential of $\hat{H}_{\mathrm{OM}}$ and $\hat{U}_{\mathrm{M}}$ be the time-ordered exponential of $\hat{H}_{\mathrm{M}}$. Since the POVM is not applied to the mechanics, $\hat{U}|\psi\rangle\langle\psi|\hat{U}^\dagger$ with $\hat{U} = \hat{U}_{\mathrm{M}}^\dagger \hat{U}_{\mathrm{OM}}$ is also a valid purification for a given force [111]. $\hat{U}$ becomes the time-ordered exponential of the interaction-picture

Hamiltonian

$$\hat{H}(t) = \hat{H}_{\mathrm{O}} + \hat{h}_{\mathrm{M}}(t), \qquad \hat{h}_{\mathrm{M}}(t) \equiv \hat{U}_{\mathrm{M}}^{\dagger}(t)\hat{h}\hat{U}_{\mathrm{M}}(t). \tag{4.17}$$

For cavity optomechanics, $\hat{h}_{\mathrm{M}}(t) = -\hbar g_0 \hat{n}\hat{q}_{\mathrm{M}}(t)$, where $\hat{q}_{\mathrm{M}}(t)$ is the interaction-picture mechanical position. For a linear mechanical system, $\hat{q}_{\mathrm{M}}(t) = \hat{q}_0(t) + X(t)$, where $\hat{q}_0(t)$ is the operator-valued homogeneous component as a function of the initial position and momentum operators and $X(t)$ is the c-number inhomogeneous component of the displacement due to the classical force. We can hence take $X(t)$ to be the hidden process and $\hat{Q} = \hbar g_0 \hat{n}$ to be the generator, obtaining uncertainty relations between the displacement errors and the photon-number fluctuations, as depicted in Fig. 4.1(b).

In general, this interaction-picture purification method can be applied to any linear system with Hamiltonian of the form $\hat{H}_0 - \hat{Q}X(t)$, where $\hat{Q}$ is a canonical coordinate operator and $\hat{H}_0$ is quadratic with respect to canonical coordinates, as the effect of $X(t)$ remains a displacement operation in any interaction picture.

Figure 4.1(c) and (d) depict two other examples of Eq. (4.15) in the context of optical phase modulation, in which case $X(t)$ is the phase modulation on the optical beam and $\hat{Q}$ is proportional to the photon-flux operator. Other examples include the magnetometer, where $X(t)$ is an external magnetic field and $\hat{Q}$ is a spin operator [120], and the voltmeter, where $X(t)$ is an applied voltage and $\hat{Q}$ is a charge operator.

4.1.4 Variational bound

As the extended convexity holds for any ensemble of $\hat{\rho}$, tighter bounds can be obtained by choosing the ensemble judiciously [116]. Instead of the original ensemble given by Eq. (4.4), we define a new stochastic process

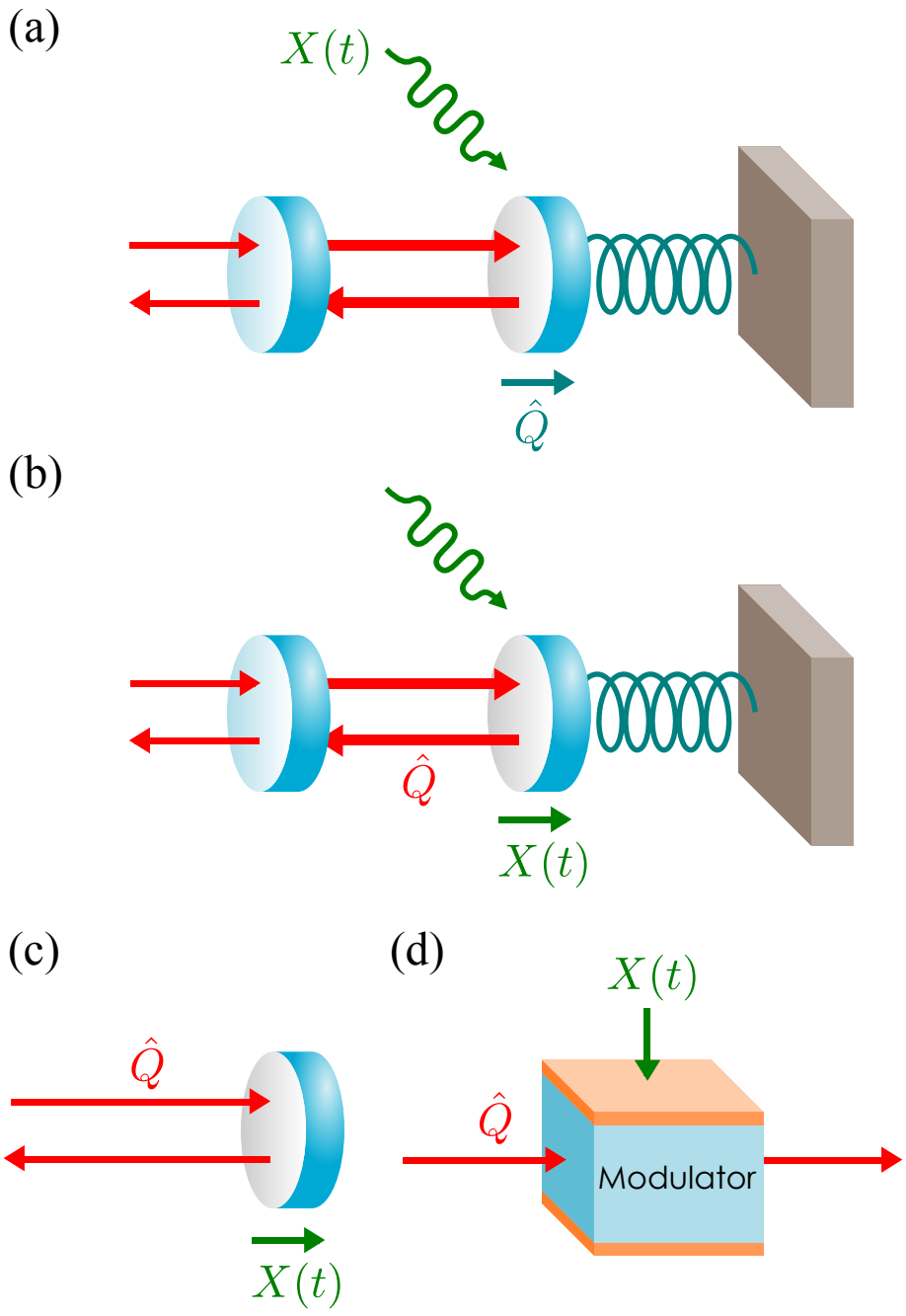


Figure 4.1: (Color online). Some examples of the hidden stochastic process $X(t)$ and generator $\hat{Q}$. (a) $X(t)$ is the classical force and $\hat{Q}$ is the mechanical position, (b) $X(t)$ is the c-number forced displacement and $\hat{Q}$ is proportional to the photon-number operator, (c) and (d) $X(t)$ is the phase modulation and $\hat{Q}$ is proportional to the photon-flux operator.

$Z(t)$ by

$$X(t) = \int_{-\infty}^{\infty} d\tau g(t-\tau|\theta) Z(\tau), \tag{4.18}$$

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

$$\hat{\rho}(\theta) = \mathbb{E}_{Z|\theta}\left\{\hat{U}[g*Z,T]|\psi\rangle\langle\psi|\hat{U}^{\dagger}[g*Z,T]\right\}, \tag{4.19}$$

where $*$ denotes convolution. With

$$\hat{\sigma} = \hat{U}[g*Z,T]|\psi\rangle\langle\psi|\hat{U}^{\dagger}[g*Z,T], \tag{4.20}$$

this results in a family of ensembles $\{\hat{\sigma}, p_Z\}$ parameterized by g for a given $\hat{\rho}$.

Assuming the Hamiltonian in Eq. (4.15), it can be shown that [110, 121]

$$\begin{aligned} J_{\mu\nu}(\hat{\sigma}) = \frac{4}{\hbar^2} \int_0^T dt \int_0^T dt' K_Q(t,t') \\ \times \int_{-\infty}^{\infty} d\tau \partial_\mu g(t-\tau|\theta) Z(\tau) \\ \times \int_{-\infty}^{\infty} d\tau' \partial_\nu g(t'-\tau'|\theta) Z(\tau'), \end{aligned} \tag{4.21}$$

where $\partial_\mu \equiv \partial/\partial\theta_\mu$ and $K_Q(t,t')$ is the quantum covariance of the generator in the Heisenberg picture, defined as

$$K_Q(t,t') \equiv \mathrm{Re}\left[\langle\psi|\Delta\hat{Q}(t)\Delta\hat{Q}(t')|\psi\rangle\right], \tag{4.22}$$

$$\Delta\hat{Q}(t) \equiv \hat{Q}(t) - \langle\psi|\hat{Q}(t)|\psi\rangle, \tag{4.23}$$

$$\hat{Q}(t) \equiv \hat{U}^{\dagger}(X,t)\hat{Q}\hat{U}(X,t). \tag{4.24}$$

We now assume that $K_Q(t,t')$ is independent of $X(t)$; such an assumption is commonly satisfied in linear optomechanics and optical-phase-modulation

systems. The expected $J(\hat{\sigma})$ becomes

$$\begin{aligned}\mathbb{E}_{Z|\theta}\left[J_{\mu\nu}(\hat{\sigma})\right] = \frac{4}{\hbar^2}\int_0^T dt \int_0^T dt' K_Q(t,t') \\ \times \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' K_Z(\tau,\tau'|\theta) \\ \times \left[\partial_\mu g(t-\tau|\theta)\right]\left[\partial_\nu g(t'-\tau'|\theta)\right], \end{aligned} \tag{4.25}$$

where

$$K_Z(\tau,\tau'|\theta) \equiv \mathbb{E}_{Z|\theta}\left[Z(\tau)Z(\tau')\right] \tag{4.26}$$

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

$$S_Q(\omega) \equiv \int_{-\infty}^{\infty} d\tau K_Q(t,t+\tau)\exp(i\omega\tau). \tag{4.27}$$

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. Defining a transfer function as

$$G(\omega|\theta) \equiv \int_{-\infty}^{\infty} dt g(t|\theta)\exp(i\omega t), \tag{4.28}$$

restricting G to be nonzero for all frequencies of interest, noting that the PSD of $Z(t)$ is $S_X/|G|^2$, and making the SPLOT assumption, Eq. (4.25) can be rewritten as

$$\mathbb{E}_{Z|\theta}\left[J_{\mu\nu}(\hat{\sigma})\right] = T\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{4S_Q S_X}{\hbar^2}\left(\partial_\mu \ln G\right)\left(\partial_\nu \ln G^*\right). \tag{4.29}$$

The Fisher information $j(p_Z)$ can be obtained by applying Eq. (4.12) to the Bhattacharyya distance between two stationary Gaussian processes [44].

The result is

$$j_{\mu\nu}(p_Z) = T\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{1}{2}\left(\partial_\mu \ln\frac{S_X}{|G|^2}\right)\left(\partial_\nu \ln\frac{S_X}{|G|^2}\right). \tag{4.30}$$

Combining Eqs. (4.29) and (4.30) according to Eq. (4.7), we obtain

$$\begin{aligned} u_\mu \mathcal{J}_{\mu\nu} u_\nu &= T\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\left[\frac{4S_Q S_X}{\hbar^2}|\lambda|^2 + \frac{1}{2}\left(\Lambda - \lambda - \lambda^*\right)^2\right], \\ \lambda &\equiv u_\mu\partial_\mu \ln G, \quad \Lambda \equiv u_\mu \partial_\mu \ln S_X. \end{aligned} \tag{4.31}$$

Since Eq. (4.31) is quadratic with respect to λ, the λ and thus G that minimizes Eq. (4.31) for each u can be found analytically. Straightforward algebra then leads to a variational upper bound on the quantum Fisher information given by

$$J \le \tilde{\mathcal{J}}, \qquad \tilde{\mathcal{J}}_{\mu\nu} \equiv T\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2 + \hbar^2/(S_Q S_X)}. \tag{4.32}$$

This is the first main result of this chapter. Note that the quantum state $|\psi\rangle$ need not be Gaussian for the result to hold.

For mechanical force measurements, the straightforward choice of the Hamiltonian leads to S_X being the force PSD and S_Q being the mechanical position PSD. For linear cavity optomechanics, the interaction-picture purification technique explained in Sec. 4.1.3 leads to an alternative Hamiltonian such that S_X is the PSD of the forced displacement and S_Q is proportional to the cavity photon-number PSD. For continuous optical phase modulation [10, 122–124], S_X is the phase PSD and $S_Q/\hbar^2$ is the photon-flux PSD. In all cases, the frequency-domain integral given by Eq. (4.32), together with the matrix inequalities

$$\Sigma \ge j^{-1} \ge J^{-1} \ge \tilde{\mathcal{J}}^{-1} \tag{4.33}$$

that follow from Eqs. (4.2), (4.6), 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. This means that a longer observation time can improve the parameter estimation even if the SNR is low, as is well known in statistics [125] but missed by some of the previous quantum studies [102, 103].

4.2 Continuous optical phase modulation

4.2.1 Error bounds

To illustrate our theory, consider the optics experiment depicted in Fig. 4.1(c) or (d). 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

$$\hat{H} = \hbar \hat{I}(t) X(t), \tag{4.34}$$

where $\hat{I}(t)$ is the photon-flux operator, $S_X(\omega|\theta)$ is the source PSD, and $S_I(\omega) = S_Q(\omega)/\hbar^2$ is the photon-flux PSD. This model also applies to quantum optomechanics if the dynamics can be linearized around a strong optical mean field and a suitable interaction picture is used, as discussed in Sec. 4.1.3. The quantum limit given by Eq. (4.32) becomes

$$\tilde{\mathcal{J}}_{\mu\nu} = T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2 + 1/(S_I S_X)}. \tag{4.35}$$

Equation (4.35) together with Eq. (4.33) represent an uncertainty relation between the phase spectrum-parameter estimation error and the photon-flux PSD.

We can compare our bound with the Fisher information for homodyne detection, a standard experimental phase measurement method [10, 122–124], as illustrated in Fig. 4.2(a). If the mean field is strong, and the modulation is weak or tight phase locking is achieved, the output process can be linearized as

$$Y(t) \approx X(t) + \eta(t), \tag{4.36}$$

where $\eta(t)$ is the phase-quadrature noise. The information $j(p_Y^{(\mathrm{hom})})$ can be computed analytically if η is Gaussian and stationary with power spectral density $S_\eta(\omega)$ such that Y is also Gaussian and stationary [44]; the result with the SPLOT assumption is

$$j_{\mu\nu}\left(p_Y^{(\mathrm{hom})}\right) = T\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2(1+S_\eta/S_X)^2}. \tag{4.37}$$

The classical Cramér-Rao bound $\Sigma \geq j^{-1}(p_Y^{(\mathrm{hom})})$ is asymptotically attainable for long T using maximum-likelihood estimation [125].

With the quadrature uncertainty relation

$$S_\eta(\omega)S_I(\omega) \geq \frac{1}{4} \tag{4.38}$$

for the optical beam [19], the optimal homodyne information is

$$\begin{aligned} j\left(p_Y^{(\mathrm{hom})}\right) &\leq \tilde{\tilde{j}}, \\ \tilde{\tilde{j}}_{\mu\nu} &\equiv T\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{(\partial_\mu \ln S_X)(\partial_\nu \ln S_X)}{2+1/(S_I S_X)+1/(8S_I^2 S_X^2)}. \end{aligned} \tag{4.39}$$

We can compare this homodyne limit with the quantum limit in Eq. (4.35);

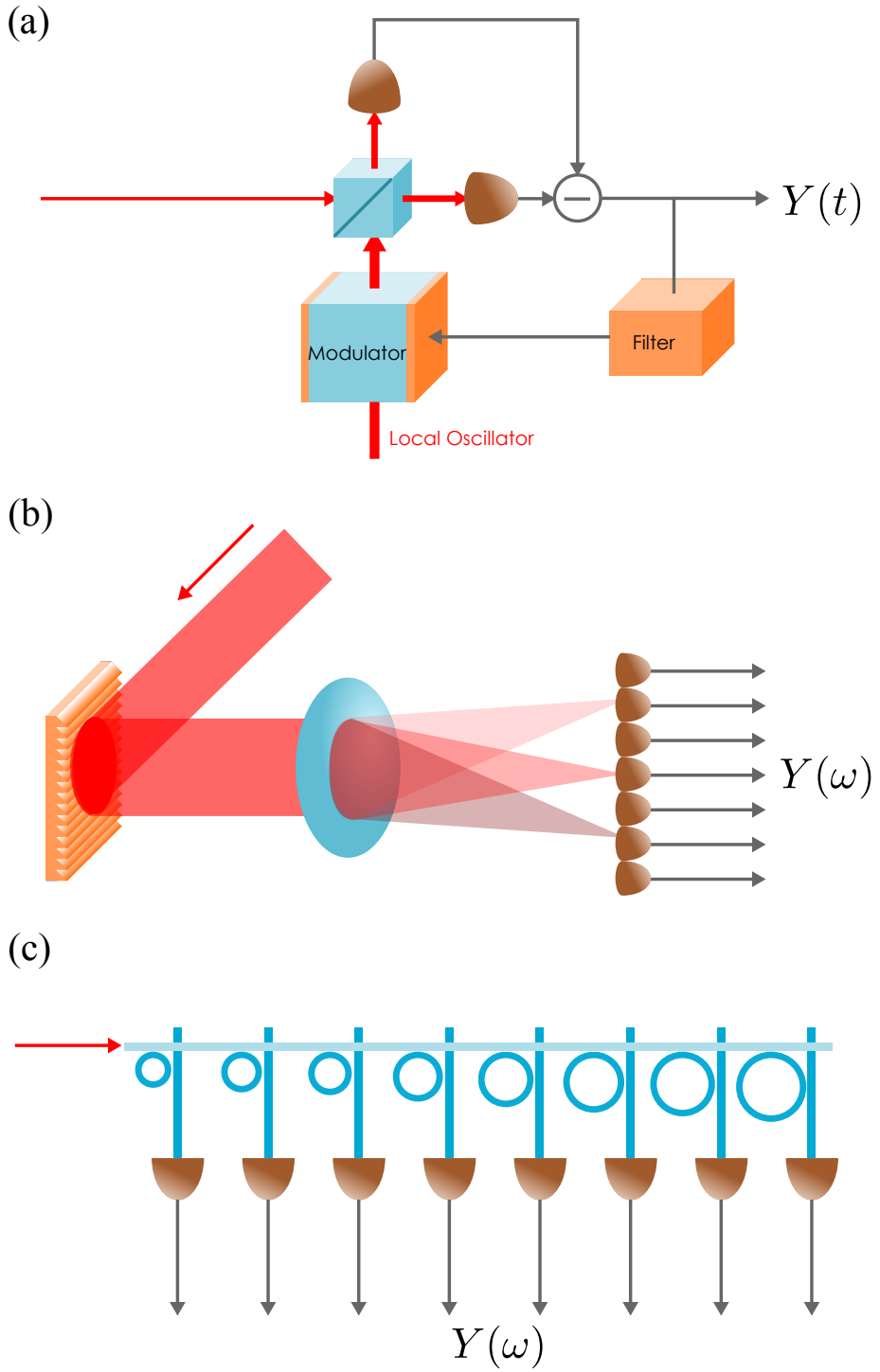


Figure 4.2: (Color online). (a) Adaptive homodyne detection. (b) Spectral photon counting with a diffraction grating and a lens. (c) Spectral photon counting with an optical-resonator array.

the expressions are similar, apart from a extra factor of $1/(8S_I^2S_X^2)$ that makes the homodyne limit strictly worse than our quantum limit, especially if S_IS_X is small.

4.2.2 Spectral photon counting

Although Eq. (4.33) sets rigorous lower bounds on the estimation error Σ, there is no guarantee that the the error for any measurement can attain the final bound $\tilde{J}^{-1}$. Inspired by our previous work on astronomical quantum optics [108, 126], here we analyze an alternative measurement that we call spectral photon counting. Physically, it is simply a conventional optical spectrometer with photon counting for each spectral mode [127, 128]. 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 [127] as depicted in Fig. 4.2(b), or an array of optical ring resonators with different resonant frequencies coupled to a cross grid of waveguides [129] as depicted in Fig. 4.2(c). The second step is a measurement of the photon numbers in the spectral modes, and the final step is a maximum-likelihood estimation of θ from the spectral photon counting results. For the phase spectrum-parameter estimation problem with weak modulation and a coherent-state input, this method turns out to have an information $j(p_Y^{(\mathrm{spc})})$ coinciding with $\tilde{\mathcal{J}}$ for all parameters.

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

$$\hat{E}^{(+)}(t) = \hat{A}(t)\exp(-i\Omega t), \tag{4.40}$$

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

commutation relation

$$[\hat{A}(t), \hat{A}^\dagger(t')] = \delta(t - t'), \tag{4.41}$$

and Ω is the optical carrier frequency. With a strong mean field

$$\alpha \equiv \langle\psi|\hat{A}(t)|\psi\rangle \tag{4.42}$$

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 [127] to express each frequency mode in terms of the mode annihilation operator as

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

with sideband frequencies

$$\omega_m = \frac{2\pi m}{T}, \qquad m \in \{\dots, -2, -1, 0, 1, 2, \dots\}, \tag{4.45}$$

and

$$[\hat{b}_m, \hat{b}_n^\dagger] = \delta_{mn}. \tag{4.46}$$

Assuming α to be time-constant,

$$\hat{b}_m \approx \hat{a}_m + i\alpha x_m, \tag{4.47}$$

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 [127]. 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 $S_X(\omega_m|\theta)$ [125], 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 [19] with mean number

$$\bar{N}_m = 2|\alpha|^2 S_X(\omega_m|\theta). \tag{4.48}$$

The Fisher information remains analytically tractable and is given by

$$j_{\mu\nu}(p_Y^{(\mathrm{spc})}) = \sum_{m>0} \frac{(\partial_\mu \ln \bar{N}_m)(\partial_\nu \ln \bar{N}_m)}{1 + 1/\bar{N}_m}. \tag{4.49}$$

If we use the SPLOT assumption to replace $\sum_{m>0}$ with $T\int_0^\infty d\omega/(2\pi)$ [7] and use the symmetry of the integrand to replace $T\int_0^\infty d\omega/(2\pi)$ with

$(T/2)\int_{-\infty}^{\infty} d\omega/(2\pi)$, the Fisher information becomes

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

where $\mathcal{N}$ is 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. (4.35). This is the second main result of this chapter. Comparing Eq. (4.50) with the homodyne limit given by Eq. (4.39), we can expect that spectral photon counting becomes significantly better than homodyne detection when $\mathcal{N}S_X$ is small.

4.2.3 Ornstein-Uhlenbeck spectrum analysis

For a more specific example, consider the experiments in Refs. [122, 123], which can be modeled as the continuous-optical-phase-modulation problem depicted in Fig. 4.1(d), with adaptive homodyne detection depicted in Fig. 4.2(a) and $X(t)$ given by an Ornstein-Uhlenbeck process. The PSD of $X(t)$ is

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

where $\theta_1 = \mathbb{E}_{X|\theta}[X^2(t)]$ is the area under S_X and θ_2 is the bandwidth. The experimental S_I can be assumed to be constant for all frequencies of interest, and the quantum limit given by Eq. (4.35) 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. (4.39) is

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

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.

Figure 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 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.

4.2.4 Experimental data analysis

To compare our theory with actual experimental performance, we analyze the data from the experiment reported in Ref. [122], which is in a high-SNR regime ($C \geq 23.5$) and 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. [123], as Eqs. (4.55) imply that squeezing offers insignificant improvement in this high-SNR regime.

The experiment reported in Ref. [122] used four different mean photon fluxes $\mathcal{N}_1 = 1.315 \times 10^6$ s^{-1}, $\mathcal{N}_2 = 3.616 \times 10^6$ s^{-1}, $\mathcal{N}_3 = 6.327 \times 10^6$ s^{-1}, $\mathcal{N}_4 = 1.418 \times 10^7$ s^{-1}. For each photon flux $\mathcal{N}_k$, M_k traces of $X(t)$ and M_k traces of $Y(t)$ were recorded ($M_1 = 21$, $M_2 = 23$, $M_3 = 24$, $M_4 = 27$). Each trace of $Y(t)$ was obtained using a different feedback gain for the filter in

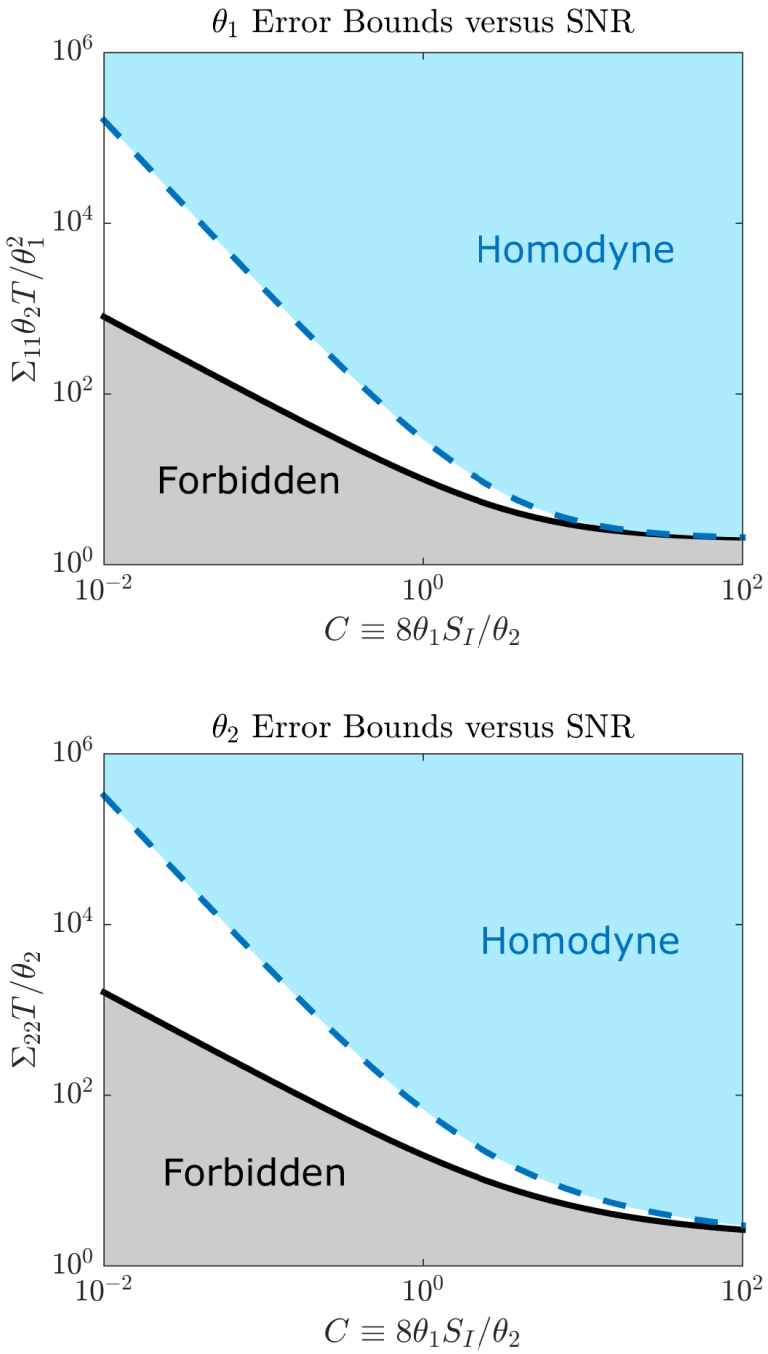


Figure 4.3: (Color online). Log-log plots of the quantum limit $\tilde{\mathcal{J}}^{-1}$ (inverse of Eqs. (4.52), black solid line) and homodyne limit $\tilde{j}^{-1}$ (inverse of Eqs. (4.54), blue dashed line) on the mean-square errors versus an SNR quantity $C \equiv 8\theta_1 S_I/\theta_2$. Top plot: limits on Σ_{11} (normalized in a unit of $\theta_1^2/(\theta_2 T)$), bottom plot: limits on Σ_{22} (normalized in a unit of θ_2/T). No measurement can achieve an error below the quantum limit (grey "forbidden" region), while the homodyne performance (blue "homodyne" region) cannot go below the homodyne limit. For $C \gg 1$, the limits approach constants, while for $C \ll 1$ the homodyne limit has a significantly worse error scaling.

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. [122], but it is also coincidentally appropriate in our present context, as θ_1 and θ_2 are supposed to be unknown here and the optimal filter is not supposed to be known. To make the data analysis tractable, we assume 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 reasonable. 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. Appendix B describes further calibrations to ensure that Eq. (4.58) is accurate.

For any observation time T, the maximum-likelihood estimation can be performed using an expectation-maximization algorithm [125, 130], but our numerical simulations suggest that it is safe here to use a simpler and faster method due to Whittle [131], which exploits the SPLOT assumption to simplify the likelihood function. Consider a real discrete-time series

$$\{Y(t_l); l = 0, 1, \ldots, L-1\}, \quad t_l = l\delta t, \tag{4.59}$$

and zero-mean Gaussian statistics conditioned on θ. Define the discrete

Fourier transform as

$$y_m = \frac{\delta t}{\sqrt{T}} \sum_{l=0}^{L-1} Y(t_l) \exp(i\omega_m t_l), \qquad \omega_m = \frac{2\pi m}{T}, \tag{4.60}$$

with integer m and $y_m = y^*_{L-m}$. It can be shown that, with the SPLOT assumption, the positive-frequency components $\{y_m; 0 < m < L/2\}$ are independent zero-mean complex Gaussian random variables with variances $S_Y(\omega_m|\theta)$ [125, 131]. This means that the log-likelihood function, up to a θ-independent additive constant $\mathcal{A}$, can be approximated as

$$\ln p_Y \approx \mathcal{A} - \sum_{0<m<L/2} \left[\ln S_Y(\omega_m|\theta) + \frac{|y_m|^2}{S_Y(\omega_m|\theta)} \right]. \tag{4.61}$$

Approximate maximum-likelihood estimation can then be performed by Fourier-transforming the time series into $\{y_m\}$ and finding the parameters that maximize Eq. (4.61). We use Matlab® and its `fft` and `fminunc` functions to implement this procedure on a desktop PC. 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 $\theta_1 = 0.1323$ and $\theta_2 = 5.909 \times 10^4$ rad/s. 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 es-

timation 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

$$\left\{ \check{\theta}_{\mu k}^{(m_k)}; \mu = 1, 2; k = 1, 2, 3, 4; m_k = 1, \ldots, M_k \right\}, \tag{4.62}$$

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

$$\varepsilon_{\mu k}^{(m_k)} \equiv \left(\check{\theta}_{\mu k}^{(m_k)} - \theta_\mu \right)^2. \tag{4.63}$$

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

$$\bar{\varepsilon}_{\mu k} \equiv \frac{1}{M_k} \sum_{m_k=1}^{M_k} \varepsilon_{\mu k}^{(m_k)} \tag{4.64}$$

to estimate the expected error

$$\Sigma_{\mu\mu} = \mathbb{E}_Y(\varepsilon_{\mu k}). \tag{4.65}$$

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,

$$V_{\mu k} \equiv \frac{1}{M_k - 1} \sum_{m_k=1}^{M_k} \left(\varepsilon_{\mu k}^{(m_k)} - \bar{\varepsilon}_{\mu k} \right)^2, \tag{4.66}$$

and divide it by the number of samples M_k. Our final results

$$\left\{ \bar{\varepsilon}_{\mu k} \pm \sqrt{\frac{V_{\mu k}}{M_k}}; \mu = 1, 2; k = 1, 2, 3, 4 \right\} \tag{4.67}$$

are plotted in normalized units in Fig. 4.4, together with the quantum limit given by the inverse of Eqs. (4.52) and the homodyne limit given by the inverse of Eqs. (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.

4.3 Conclusion

We have presented three key results in this chapter: a measurement-independent quantum limit to spectrum-parameter estimation, 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 gravitational-wave astronomy [26, 100–103, 109]. 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 explicitly 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 [132]. Tighter quantum limits that explicitly account for decoher-

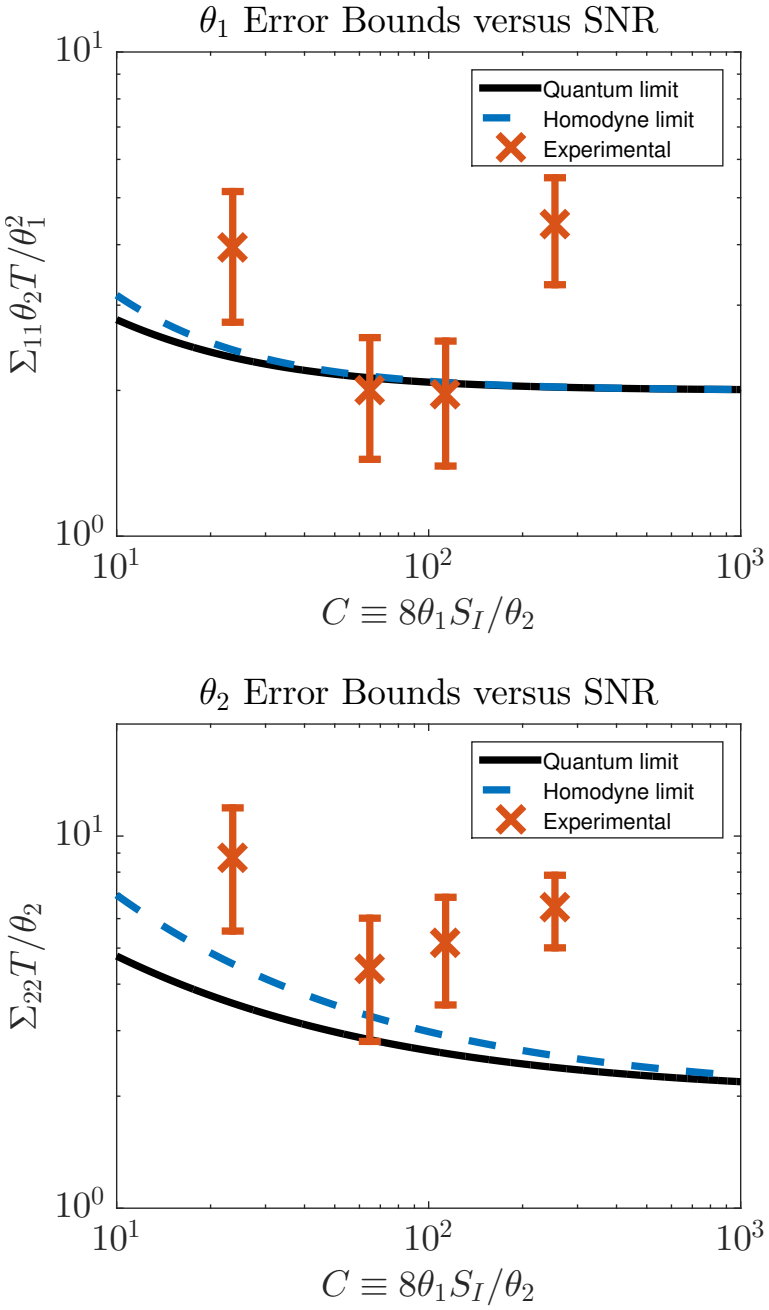


Figure 4.4: (Color online). Log-log plots of the quantum limit $\tilde{\mathcal{J}}^{-1}$ (inverse of Eqs. (4.52), black solid line), the homodyne limit $\tilde{j}^{-1}$ (inverse of Eqs. (4.54), blue dash line), and the experimental mean-square estimation errors Σ versus the SNR quantity $C \equiv 8\theta_1 S_I/\theta_2$. Top plot: Experimental $\Sigma_{11} = \{4.0 \pm 1.2, 2.0 \pm 0.6, 2.0 \pm 0.6, 4.4 \pm 1.1\}$ (in a unit of $\theta_1^2/(\theta_2 T)$) versus $C = \{23.5, 64.8, 113, 254\}$, compared with the homodyne limit and the quantum limit. Bottom plot: Experimental $\Sigma_{22} = \{8.7 \pm 3.2, 4.4 \pm 1.6, 5.2 \pm 1.7, 6.4 \pm 1.4\}$ (in a unit of θ_2/T) versus the same C values, compared with the homodyne limit and the quantum limit.

ence may be derived by applying the techniques in Refs. [111, 118, 119]. A Bayesian formulation that removes the unbiased-estimator assumption should be possible [32, 110, 111, 133, 134]. A more detailed study of our theory in the context of optomechanics can serve as an extension of Refs. [102, 103] 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 [120].

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 [11–13, 108, 126, 135, 136]. 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

Spectral analysis for quantum stochastic processes

[1]In the previous chapter, we have studied the case where a quantum system is used to measure the spectrum parameters of a classical stochastic process. However, in a lot of cases, the system-bath interaction requires a full quantum treatment, especially when the quantum system interacts with a bath of quantum fields. Inspired by quantum optical experiments [19–21], one of our main interests lies in studying the usage of cavity system as a measuring device. In a lot of these experiments, the cavity field is coupled to the continuous field outside the cavity, causing the cavity field to leak through the mirrors. To describe such an interaction between the cavity field and the continuous field, Gardiner and Collett developed a quantum input-output formalism for optical systems [137, 138]. Under this formalism, the Heisenberg equation of motion for a system operator is in the form of a quantum Langevin equation, where the continuous field serves as a quantum stochastic process driving the evolution of the system operator. Thus, measurements performed on the cavity field can be used to infer the properties of the quantum stochastic process.

As we have seen, the answer to the limit of estimation accuracy comes

[1]For notational simplicity, operators in this chapter *are not* denoted with a 'hat'.

in the form of quantum Cramér-Rao bound. The minimum mean-square error of any estimator, independent of any measurements is given by the inverse of quantum Fisher information. However, the direct evaluation of the quantum Fisher information is often too difficult, and we will focus on finding upper bounds for our problem. As seen in Chap. 4, we can apply a variational method to the extended convexity of quantum Fisher information [116] to find an upper bound if the stochastic process that couples to the system is classical. However, such method is not applicable to the case of a quantum stochastic process and generalization of the method in Chap. 4 is needed.

We will start in Sec. (5.1) to introduce the modified extended convexity of quantum Fisher information, which provides an upper bound on the quantum Fisher information of interest. As we will model the bath statistics with a special class of quantum Gaussian state, a quick reminder of Gaussian quantum information [139, 140] will be given in Sec. (5.2). These two sections form the core ideas of the method we use to derive our upper bounds to the quantum Fisher information. Finally, we state explicitly the model of quantum stochastic process we consider and present our results in Sec. (5.3).

5.1 The modified extended convexity of quantum Fisher information

Let us start with the Hilbert space $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B$, where $\mathcal{H}_S$ is the system Hilbert space while $\mathcal{H}_B$ is the bath Hilbert space. We assume that the bath degrees of freedom are not available to us. Measurements are performed on the system degrees of freedom to infer properties of the bath. The density

matrix of interest is given by the channel defined as

$$\rho(\theta) = \mathrm{tr}_B \left[U \rho_S \otimes \rho_B(\theta) U^\dagger \right], \tag{5.1}$$

where ρ_S (ρ_B) is the initial system (bath) density matrix. The system-bath interaction is specified by the unitary operator U on the system-bath Hilbert space $\mathcal{H}$. Although we are interested in the case where only the initial bath density matrix depends on an unknown parameter θ, the technique we present here would also be applicable when both ρ_B and U depend on the parameter.

To follow the similar procedure as in Chap. 4, we first find a modified convexity of quantum Fisher information for states given by Eq. (5.1), which would provide an upper bound on the quantum Fisher information. Inspired by the variational approach in Chap. 4 and [119, 141], we further assume that the initial bath density matrix is $\rho_B = \mathrm{tr}_{an}(U_{tot} \rho'_{tot} U_{tot}^\dagger)$ to tighten the upper bound, where an stands for the ancilla degree of freedom we introduced. A θ-dependent unitary operator U_{tot} on the enlarged bath Hilbert space $\mathcal{H}_{tot} = \mathcal{H}_B \otimes \mathcal{H}_{an}$ is included to serve as the variable for the variational approach. The upper bound specified by the modified extended convexity of quantum Fisher information then depends on (ρ'_{tot},U_{tot}) and can be improved by minimizing over a selected class of (ρ'_{tot},U_{tot}).

Given the assumption $\rho_B = \mathrm{tr}_{an}(U_{tot} \rho'_{tot} U_{tot}^\dagger)$, the density matrix in question is given equivalently by

$$\rho = \mathrm{tr}_{tot} \left[\tilde{U} \rho_S \otimes \rho'_{tot} \tilde{U}^\dagger \right], \tag{5.2}$$

$$\tilde{U} = U_{tot}^\dagger U U_{tot}. \tag{5.3}$$

Here, the unitary U acts on the system-bath Hilbert space $\mathcal{H}$, while the unitary U_{tot} acts on the enlarged bath Hilbert space $\mathcal{H}_{tot}$. For our purpose, let us consider $\rho'_{tot} = \sum_m p_m(\theta) |\psi_m\rangle\langle\psi_m|$, where $\{|\psi_m\rangle\}$ is an orthogonal

basis which does not depend on the unknown parameter. In that case, the SLD L_μ for the bath density matrix ρ'_{tot} is simply given by

$$L_\mu = \sum_m \partial_\mu \ln p_m(\theta) |\psi_m\rangle\langle\psi_m|, \tag{5.4}$$

where $\partial_\mu = \frac{\partial}{\partial\theta_\mu}$. Consider the purification of ρ'_{tot} given by

$$|\rho'_{tot}\rangle = \sum_m \sqrt{p_m(\theta)} |\psi_m, \phi_m\rangle, \tag{5.5}$$

where $\{\phi_m\}$ is an orthogonal basis on the extra Hilbert space $\mathcal{H}_{\tilde{tot}}$ used to purify ρ'_{tot}. Let $I_{\tilde{tot}}$ be the identity operator in $\mathcal{H}_{\tilde{tot}}$, $L_\mu \otimes I_{\tilde{tot}}$ solves the following Lyapunov equation:

$$\partial_\mu(|\rho'_{tot}\rangle\langle\rho'_{tot}|) = \frac{1}{2}(L_\mu \otimes I_{\tilde{tot}}|\rho'_{tot}\rangle\langle\rho'_{tot}| + |\rho'_{tot}\rangle\langle\rho'_{tot}|L_\mu \otimes I_{\tilde{tot}}), \tag{5.6}$$

$$\langle\rho'_{tot}|L_\mu \otimes I_{\tilde{tot}}|\rho'_{tot}\rangle = \sum_m \partial_\mu p_m = 0. \tag{5.7}$$

In other words, the operator $L_\mu \otimes I_{\tilde{tot}}$ is the SLD for the purification $|\rho'_{tot}\rangle$.

Given this purification of ρ'_{tot}, let us purify ρ by

$$|\rho\rangle = \tilde{U}|\rho_S\rangle|\rho'_{tot}\rangle, \tag{5.8}$$

where $|\rho_S\rangle$ is a purification of ρ_S in the space $\mathcal{H}_S \otimes \mathcal{H}_{\tilde{S}}$. For simplicity, we write $\tilde{U} = \tilde{U} \otimes \tilde{I}$ where $\tilde{I}$ is the identity operator in the combined ancilla Hilbert space $\tilde{\mathcal{H}} = \mathcal{H}_{\tilde{S}} \otimes \mathcal{H}_{\tilde{tot}}$. By the monotonicity of quantum Fisher information under completely positive map[2] [53, 54], the quantum Fisher information $\mathcal{J}$ of the pure state $|\rho\rangle$ then sets an upper bound on $J(\rho)$:

$$\mathcal{J} \geq J(\rho). \tag{5.9}$$

[2]It's easy to understand why this must be true: subsystem must always contains less information than the total system.

We now calculate the quantum Fisher information for the purification $|\rho\rangle$. Instead of using the SLD, the quantum Fisher information for pure states can also be calculated by using any nonsymmetric logarithmic derivatives $\mathcal{L}_\mu$ [142] satisfying

$$\begin{aligned} \partial_\mu|\rho\rangle\langle\rho| &= \frac{1}{2}(\mathcal{L}_\mu|\rho\rangle\langle\rho| + |\rho\rangle\langle\rho|\mathcal{L}_\mu^\dagger), \\ \mathrm{tr}(\mathcal{L}_\mu|\rho\rangle\langle\rho|) &= 0, \end{aligned} \tag{5.10}$$

where the quantum Fisher information of the pure state $|\rho\rangle\langle\rho|$ is given by the expression

$$\mathcal{J}_{\mu\nu}(|\rho\rangle\langle\rho|) = \frac{1}{2}\,\mathrm{tr}\left[|\rho\rangle\langle\rho|(\mathcal{L}_\mu^\dagger\mathcal{L}_\nu + \mathcal{L}_\nu^\dagger\mathcal{L}_\mu)\right]. \tag{5.11}$$

The following nonsymmetric logarithmic derivative satisfies the conditions in Eqs. (5.10):

$$\mathcal{L}_\mu = \tilde{U}[L_\mu^U + L_\mu]\tilde{U}^\dagger \tag{5.12}$$

$$L_\mu^U = 2[\tilde{U}^\dagger\partial_\mu\tilde{U} - \mathrm{tr}(\rho_S\rho'_{tot}\tilde{U}^\dagger\partial_\mu\tilde{U})]. \tag{5.13}$$

Thus, the quantum Fisher information of pure state $|\rho\rangle$ is given by:

$$\begin{aligned} \mathcal{J} &= \mathcal{J}^U + J(\rho'_{tot}), \\ \mathcal{J}_{\mu\nu}^U &= \frac{1}{2}\,\mathrm{tr}\left[\rho_S\rho'_{tot}(L_\mu^{U\dagger}L_\nu^U + L_\nu^{U\dagger}L_\mu^U)\right], \\ J_{\mu\nu}(\rho'_{tot}) &= \frac{1}{2}\,\mathrm{tr}\left[\rho'_{tot}(L_\mu L_\nu + L_\nu L_\mu)\right]. \end{aligned} \tag{5.14}$$

The final bound on the quantum Fisher information $J(\rho)$ is then given by

$$\mathcal{J} = \mathcal{J}^U + J(\rho'_{tot}) \geq J(\rho). \tag{5.15}$$

We call Eq. (5.15) the modified extended convexity of quantum Fisher information for quantum channel defined by Eq. (5.1), where $\rho_B = \mathrm{tr}_{an}(U_{tot}\rho'_{tot}U_{tot}^\dagger)$.

The upper bound specified by Eqs. (5.14) and (5.15) can be further optimized by choosing a family of (ρ'_{tot}, U_{tot}) satisfying $\rho_B = \mathrm{tr}_{an}(U_{tot}\rho'_{tot}U_{tot}^\dagger)$ and finding the particular doublet (ρ'_{tot}, U_{tot}) with the minimum $\mathcal{J}$.

5.2 Gaussian bosonic systems

Before we discuss the selection of the family (ρ'_{tot}, U_{tot}), let us review some concepts of Gaussian systems [139, 140]. Consider the Hilbert space $\mathcal{H}_B$ of N bosonic modes where the quadrature operators are given by q_m and p_m. Define the $2N$-vector of quadrature operators $Q^T = (q_1...q_N, p_1...p_N)$, the canonical commutation relation can be summarized as

$$[Q_m, Q_n] = i\Omega_{B,mn}, \tag{5.16}$$

$$\Omega_B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \otimes I_N, \tag{5.17}$$

where I_N is the $N \times N$ identity matrix. It is customary to use the Wigner quasiprobability distribution to describe the state of a bosonic system, which is the Fourier transform of the Wigner characteristic function defined as

$$\phi_{\rho_B}(z) = \mathrm{tr}\left(\rho_B e^{iQ^\top z}\right), \tag{5.18}$$

where z is a $2N$-vector.

A Gaussian state ρ_B is fully characterized by its mean and quantum covariance matrix, defined as

$$\bar{Q}_{B,m} = \mathrm{tr}(\rho_B Q_m), \tag{5.19}$$

$$\alpha_{B,mn} = \frac{1}{2}\,\mathrm{tr}(\rho_B\{Q_m, Q_n\}), \tag{5.20}$$

where $\{Q_m, Q_n\} = Q_mQ_n + Q_nQ_m$ is the anticommutator. On top of being a real symmetric matrix, the quantum covariance matrix must also satisfy

the uncertainty relation $\alpha_B + \frac{i}{2}\Omega_B \geq 0$. The Wigner characteristic function of a Gaussian state is a Gaussian parametrized by $\bar{Q}_B$ and α_B:

$$\phi_{\rho_B}(z) = \exp\left(i\bar{Q}_B^\top z - \frac{1}{2}z^\top \alpha_B z\right). \tag{5.21}$$

Consider a Hamiltonian quadratic in terms of the quadratures $H(A) = Q^\top A Q$ where A is a real symmetric matrix, we have the following unitary transformation

$$e^{iH(A)t} Q^\top e^{-iH(A)t} = Q^\top D_A, \tag{5.22}$$

$$D_A = \exp(-2A\Omega_B t). \tag{5.23}$$

Therefore, under the action of the unitary $U = e^{-iH(A)t}$, the Gaussian state ρ_B transforms into another Gaussian state $U\rho_B U^\dagger$, with the mean $D_A^\top \bar{Q}_B$ and quantum covariance matrix $D_A^\top \alpha_B D_A$. Futhermore, since unitary transformations must preserve the commutator, D_A must be a symplectic transformation which satisfies

$$D_A^\top \Omega_B D_A = \Omega_B. \tag{5.24}$$

If we consider only Gaussian states and Gaussian unitaries, we can focus on the symplectic space $\mathbb{Z}_B$, which is a $\mathbb{R}^{2N}$ vector space equipped with the symplectic form Ω_B. Gaussian states can then be represented on the symplectic space, and Gaussian unitaries are symplectic transformations D_A[3]. If we introduce new bosonic modes, and call the extra Hilbert space $\mathcal{H}_{an}$, the resulting Hilbert space would be $\mathcal{H}_{tot} = \mathcal{H}_B \bigotimes \mathcal{H}_{an}$, whereas the total symplectic space becomes $\mathbb{Z}_{tot} = \mathbb{Z}_B \bigoplus \mathbb{Z}_{an}$, and the total symplectic form is given by $\Omega_{tot} = \Omega_B \oplus \Omega_{an}$.

Consider the case where ρ_B is a zero mean Gaussian state, that is $\bar{Q}_B =$

[3]We have ignored Hamiltonians linear in terms of the quadratures, which correspond to displacements of the mean.

0. We can always find a Gaussian state ρ_{tot} in a larger Hilbert space such that $\rho_B = \mathrm{tr}_{an}(\rho_{tot})$. This is best understood by considering the Wigner characteristic function:

$$\phi_{\rho_B}(z) = \mathrm{tr}_{tot}\left(\rho_{tot} e^{iQ_{tot}^\top z_{tot}}\right)\Big|_{z_{tot}=(z^\top,0)^\top} = \phi_{\rho_{tot}}(z_{tot})|_{z_{tot}=(z^\top,0)^\top},$$

where $Q_{tot} = Q \oplus Q_{an}$ is the quadrature vector in the enlarged space. Thus, the covariance matrix of ρ_{tot} must be in the form

$$\alpha_{tot} = \begin{pmatrix} \alpha_B & \beta \\ \beta^\top & \alpha_{an} \end{pmatrix}, \tag{5.25}$$

where α_{tot} satisfies the uncertainty relation $\alpha_{tot} + \frac{i}{2}\Omega_{tot} \geq 0$. For simplicity, we shall assume that the ancilla Hilbert space has the same dimension as the Hilbert space $\mathcal{H}_B$, hence $\Omega_{an} = \Omega_B$.

In order to facilitate future discussions, let us consider $\rho_B = \mathrm{tr}_{an}(\rho_{tot})$ and

$$\rho_{tot} = \exp(-iH_{tot})(\rho_1' \otimes \rho_2') \exp(iH_{tot}), \tag{5.26}$$

$$H_{tot} = Q_{tot}^\top A_{tot} Q_{tot}, \tag{5.27}$$

$$A_{tot} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & g \\ 0 & 0 & g^\top & 0 \\ 0 & g & 0 & 0 \\ g^\top & 0 & 0 & 0 \end{pmatrix}, \tag{5.28}$$

where g is a real matrix. Assume that ρ_j' is a Gaussian state with covariance

matrix given by the block matrix

$$\alpha'_j = \begin{pmatrix} \alpha'_{j,R} & -\alpha'_{j,I} \\ \alpha'_{j,I} & \alpha'_{j,R} \end{pmatrix}, \tag{5.29}$$
$$\alpha'^{\top}_{j,R} = \alpha'_{j,R}, \quad \alpha'^{\top}_{j,I} = -\alpha'_{j,I},$$

the covariance matrix α_{tot} can be calculated from the equations:

$$\alpha_{tot} = D^{\top}_{tot}(\alpha'_1 \oplus \alpha'_2)D_{tot}, \tag{5.30}$$
$$D_{tot} = \exp(-2A_{tot}\Omega_{tot}). \tag{5.31}$$

The matrix D_{tot} and its inverse can be evaluated explicitly:

$$D_{tot} = \begin{pmatrix} C & 0 & S & 0 \\ 0 & C^{\top} & 0 & -S^{\top} \\ S & 0 & C & 0 \\ 0 & -S^{\top} & 0 & C^{\top} \end{pmatrix}, \tag{5.32}$$
$$D^{-1}_{tot} = \begin{pmatrix} C & 0 & -S & 0 \\ 0 & C^{\top} & 0 & S^{\top} \\ -S & 0 & C & 0 \\ 0 & S^{\top} & 0 & C^{\top} \end{pmatrix}, \tag{5.33}$$
$$S \equiv \sinh(g), C \equiv \cosh(g). \tag{5.34}$$

Therefore, we can relate the covariance matrices of ρ_B and $\rho_{an} = \mathrm{tr}_B(\rho_{tot})$ to the covariance matrix of ρ'_j:

$$\begin{aligned} \alpha_B &= \begin{pmatrix} C^{\top}\alpha'_{1,R}C + S^{\top}\alpha'_{2,R}S & -C^{\top}\alpha'_{1,I}C^{\top} + S^{\top}\alpha'_{2,I}S^{\top} \\ C\alpha'_{1,I}C - S\alpha'_{2,I}S & C\alpha'_{1,R}C^{\top} + S\alpha'_{2,R}S^{\top} \end{pmatrix}, \\ \alpha_{an} &= \begin{pmatrix} S^{\top}\alpha'_{1,R}S + C^{\top}\alpha'_{2,R}C & S^{\top}\alpha'_{1,I}S^{\top} - C^{\top}\alpha'_{2,I}C^{\top} \\ -S\alpha'_{1,I}S + C\alpha'_{2,I}C & S\alpha'_{1,R}S^{\top} + C\alpha'_{2,R}C^{\top} \end{pmatrix}. \end{aligned} \tag{5.35}$$

Note that since α_B is given by the problem, by our constructions we assumed that the bath state ρ_B is a zero mean Gaussian state with covariance matrix as given in Eqs. (5.35). In other words, α_B is decomposable in terms of matrix g (which specifies the unitary U_{tot}) and the covariance matrix α'_j. The bound in Eq. (5.15) can then be minimized by varying the matrices g and α'_j while keeping α_B fixed to obtain a tighter bound.

5.3 Upper bounds on the Quantum Fisher information for quantum stochastic processes

Let us assume the total Hamiltonian

$$H = H_S + H_B + H_{int}, \tag{5.36}$$

$$H_B = \hbar \int_{-\infty}^{\infty} d\omega \quad \omega b^\dagger(\omega) b(\omega), \tag{5.37}$$

$$H_{int} = \hbar \sqrt{\frac{\gamma}{2\pi}} \int_{-\infty}^{\infty} d\omega \big[c^\dagger b(\omega) + (\omega) c b^\dagger(\omega) \big], \tag{5.38}$$

where c is an operator in the system's Hilbert space $\mathcal{H}_S$, $b(\omega)$ is the annihilation operator in the bath's Hilbert space $\mathcal{H}_B$ which satisfies the commutation relation $[b(\omega), b^\dagger(\omega')] = \delta(\omega - \omega')$. The above Hamiltonian was used originally by Gardiner and Collett to describe an input-output theory for damped quantum system [137, 138], where the bath annihilation operator $b(\omega)$ serves as quantum noise driving the evolution of a quantum system. Going into interaction picture using $U_0(t, t_0) = e^{-i(H_S + H_B)(t - t_0)}$, we obtain the interaction picture propagator describing the evolution under the total Hamiltonian:

$$U(t, t_0) = \mathcal{T} \exp \big[-i\sqrt{\gamma} \int_{t_0}^{t} \big(c_0^\dagger(t') b(t') + c_0(t') b^\dagger(t') \big) dt' \big], \tag{5.39}$$

where

$$c_0(t) \equiv U_0^\dagger(t,t_0) c U_0(t,t_0) = e^{\frac{i}{\hbar}H_S t} c e^{-\frac{i}{\hbar}H_S t}, \tag{5.40}$$

$$b(t) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega b(\omega) e^{-i\omega(t-t_0)}, \tag{5.41}$$

$$[b(t), b^\dagger(t')] = \delta(t-t'). \tag{5.42}$$

In order to simplify the problem, we assume that $c^\dagger = c$ and write $d = \sqrt{2\gamma}c$, the propagator becomes

$$\begin{aligned} U(t,t_0) &= \mathcal{T} \exp\big[-i \int_{t_0}^{t} d_0(t') x(t') dt' \big], \\ x(t) &= \frac{1}{\sqrt{2}} \big[b(t) + b^\dagger(t) \big]. \end{aligned} \tag{5.43}$$

The quantum operator $x(t)$ can now be interpreted as a quantum stochastic process playing the part of the classical process $X(t)$ in the previous chapter.

Assuming an initial product state, the density matrix of interest is given by

$$\tilde{\rho}(\theta) = \mathrm{tr}_B \left[e^{-\frac{i}{\hbar}H(t_f - t_0)} \rho_S \otimes \rho_B(\theta) e^{\frac{i}{\hbar}H(t_f-t_0)} \right], \tag{5.44}$$

where only the initial bath density matrix ρ_B depends on the unknown parameter θ. We go into the interaction picture described above, and study the interaction picture density matrix, which is

$$\rho(\theta) = \mathrm{tr}_B \left[U(t_f,t_0) \rho_S \otimes \rho_B(\theta) U^\dagger(t_f,t_0) \right]. \tag{5.45}$$

Instead of the quantum white noise statistics assumed in the original formulation by Gardiner and Collett, we assume the bath state to be a continuous version of the Gaussian states in Eq. (5.35). We will dedicate the next subsection to describe the bath density matrix ρ_B.

5.3.1 A family of bath states ρ_{tot} such that $\rho_B = \mathrm{tr}_{an}(\rho_{tot})$

In order to apply the theory in Sec. (5.1), first, we need to describe a family of Gaussian bath states $\rho_{tot} = U_{tot}\rho'_{tot}U^\dagger_{tot}$ such that $\rho_B = \mathrm{tr}_{an}(\rho_{tot})$. As we are now working with bosonic fields instead of discrete bosonic modes, we need to extend the construction in Sec. (5.2). From now on, let us relabel $\mathcal{H}_B$ as $\mathcal{H}_{B_1}$ and $\mathcal{H}_{an}$ as $\mathcal{H}_{B_2}$. We assume that $\mathcal{H}_{B_2}$ is a Hilbert space with similar structures as $\mathcal{H}_{B_1}$, i.e. there are field operators satisfying $[b_2(t), b_2^\dagger(t')] = \delta(t-t')$. The field operator $b(t)$ will be relabeled as $b_1(t)$ while $\rho_1 = \rho_B$ is the bath density matrix of interest and $\rho_2 = \mathrm{tr}_{B_1}(\rho_{tot})$ is the reduced density matrix of ρ_{tot} on the space of $\mathcal{H}_{B_2}$. Let us assume that $\rho_{tot} = U_{tot}\rho'_1 \otimes \rho'_2 U^\dagger_{tot}$ where ρ'_j is a Gaussian state with density matrix given by

$$\rho'_j = \exp\big[-\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' G'_j(t-t') b_j^\dagger(t) b_j(t')\big]\mathcal{Z}'_j \tag{5.46}$$

where $\mathcal{Z}'_j$ is a normalization constant[4]. Under finite time approximation, $\rho'_1 \otimes \rho'_2$ is indeed diagonal in the discrete Fourier modes[5], therefore the theory of Sec. (5.1) can be applied. Let $Q_j(t)$ be a vector of the field operator $b_j(t)$

$$Q_j(t) = \begin{pmatrix} x_j(t) \\ p_j(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}\big(b_j(t) + b_j^\dagger(t)\big) \\ \frac{1}{\sqrt{2}i}\big(b_j(t) - b_j^\dagger(t)\big) \end{pmatrix} \tag{5.47}$$

and define the Fourier transform by:

$$G'_j(\omega) = \int_{-\infty}^{\infty} dt G'_j(t) e^{-i\omega t}, \tag{5.48}$$

[4]In terms of $b_j(\omega)$ the bath state is given by $\rho'_j = \mathcal{Z}'_j \exp[-\int_{-\infty}^{\infty} G'_j(\omega) b^\dagger(\omega) b(\omega) d\omega]$, where $G'_j(\omega) = \int_{-\infty}^{\infty} G'_j(t) e^{-i\omega t} dt$

[5]See Appendix C

the quantum covariance matrix α'_j of ρ'_j, defined as

$$\alpha'_{j,mn}(t,t') = \frac{1}{2}\,\mathrm{tr}\left(\rho_B\{Q_{j,m}(t), Q_{j,n}(t')\}\right), \tag{5.49}$$

is completely characterized by a complex function $K'_j(t)$ satisfying $K'_j(t) = K'^*_j(-t)$. The quantum covariance matrix is given by

$$\begin{aligned} \alpha'_j(t,t') &= \begin{pmatrix} \Re\{K'_j(t-t')\} & -\Im\{K'_j(t-t')\} \\ \Im\{K'_j(t-t')\} & \Re\{K'_j(t-t')\} \end{pmatrix}, \\ K'_j(t) &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \coth\frac{G'_j(\omega)}{2} e^{i\omega t}, \end{aligned} \tag{5.50}$$

where $\Re$ and $\Im$ denote the real part and imaginary part of a complex number. We note that the above covariance function is the continuous-time version of Eqs (5.29). See Appendix C for the proof that Eq. (5.46) leads to Eqs. (5.50).

The unitary U_{tot} is chosen to be the continuous-time analogue of $\exp(-iH_{tot})$ as in Sec. (5.2), where matrix multiplications become time integrals. For example,

$$g^2 \to \int_{-\infty}^{\infty} dt_1 g(t-t_1)g(t_1-t'). \tag{5.51}$$

The action of U_{tot} on the quadratures is given by

$$\begin{aligned} U^\dagger_{tot} Q^\top(t) U_{tot} &= \int_{-\infty}^{\infty} dt' Q^\top(t') D(t',t), \\ D(t,t') &= \begin{pmatrix} C(t,t') & 0 & S(t,t') & 0 \\ 0 & C(t',t) & 0 & -S(t',t) \\ S(t,t') & 0 & C(t,t') & 0 \\ 0 & -S(t',t) & 0 & C(t',t) \end{pmatrix}, \end{aligned} \tag{5.52}$$

where $Q(t) = Q_1(t) \oplus Q_2(t)$. The real functions $C(t,t')$ and $S(t,t')$ are defined in terms of a real function $g(t,t')$, analogous to Eq. (5.34). Assuming that $g(t,t') = g(t-t')$, they can be given in terms of the Fourier transform

of $g(t)$:

$$C(t-t') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \cosh g(\omega) e^{i\omega(t-t')}, \tag{5.53}$$

$$S(t-t') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \sinh g(\omega) e^{i\omega(t-t')}. \tag{5.54}$$

The functions $C(t-t')$ and $S(t-t')$ satisfy the following useful properties

$$\int_{-\infty}^{\infty} dt_1 [C(t-t_1)C(t_1-t') - S(t-t_1)S(t_1-t')] = \delta(t-t'), \tag{5.55}$$

$$\int_{-\infty}^{\infty} dt_1 [C(t-t_1)S(t_1-t') - S(t-t_1)C(t_1-t')] = 0, \tag{5.56}$$

which can be checked in the frequency domain. As shown in Sec. (5.2), the transformation $D(t-t')$ describes a symplectic transformation which transforms the Gaussian state $\rho_1' \otimes \rho_2'$ into another Gaussian state ρ_{tot}. The quantum covariance matrix for ρ_{tot} is given by

$$\alpha_{tot,\sigma\nu} = \sum_{k,m} \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 D_{k\sigma}(t_1-t)\alpha'_{km}(t_1-t_2)D_{m\nu}(t_2-t'), \tag{5.57}$$

where $\alpha' = \alpha_1' \oplus \alpha_2'$. This equation and the triplet $g(t)$, $K_1'(t)$ and $K_2'(t)$ then define the density matrix ρ_{tot} where $\rho_1 = \mathrm{tr}_{B_2}(\rho_{tot})$ as needed. The covariance matrix of ρ_1 is given in terms of $g(t)$, $K_1'(t)$ and $K_2'(t)$ analogous to Eqs. (5.35), hence we can vary the triplet $g(t)$, $K_1'(t)$ and $K_2'(t)$ while keeping ρ_1 fixed to tighten the quantum Fisher information bound Eq. (5.15) by using the technique in Sec. (5.1). We first calculate the upper bound Eq. (5.15).

5.3.2 Upper bound on the quantum Fisher information

For U_{tot} specified by Eqs. (5.52), $\tilde{U}$ in Eqs. (5.3) is given by

$$\begin{aligned} \tilde{U} &= \mathcal{T}\exp[-i\int_{t_0}^{t_f} d_0(t)x_1'(t)dt], \\ x_1'(t) = U_{tot}^\dagger x_1(t) U_{tot} &= \int_{-\infty}^{\infty} dt_1[C(t_1-t)x_1(t_1) + S(t_1-t)x_2(t_1)]. \end{aligned} \tag{5.58}$$

To calculate the partial derivatives of $\tilde{U}$, we first discretize the time axis into $N-1$ equal segments. $\tilde{U}$ becomes

$$\begin{aligned} \tilde{U} &= \prod_{j=0}^{N-1} \tilde{U}_j, \\ \tilde{U}_j &= \exp\big[-id_0(t_j)x_1'(t_j)\delta t\big]. \end{aligned} \tag{5.59}$$

To get the partial derivatives of $\tilde{U}$, it is then sufficient to calculate the partial derivatives of exponential $\tilde{U}_j$, given by

$$\partial_\mu \tilde{U}_j = \big[-i\delta t d(t_j)\partial_\mu x_1'(t_j)\big]\tilde{U}_j, \tag{5.60}$$

$\partial_\mu \tilde{U}$ is then given by taking the continuous time limit $\delta t \to 0$, the result of which is similar to the one in [110]:

$$\partial_\mu \tilde{U} = -i\tilde{U}\int_{t_0}^{t_f} h_\mu(t)dt, \tag{5.61}$$

$$h_\mu(t) = U_{tot}^\dagger d_H(t) U_{tot}\partial_\mu x_1'(t), \tag{5.62}$$

$$d_H(t) = U^\dagger(t,t_0)d_0(t)U(t,t_0). \tag{5.63}$$

Here, $d_H(t)$ is the Heisenberg picture of the operator d, before the transformation U_{tot}. Eq. (5.13) is then given by

$$L_\mu^U = -2i\int_{t_0}^{t_f}\big[h_\mu(t) - \mathrm{tr}\left(\rho_S\rho_1'\otimes\rho_2' h_\mu(t)\right)\big]. \tag{5.64}$$

To proceed, we assume that $d_H(t)$ is an operator in $\mathcal{H}_S$, $h_\mu(t)$ becomes a zero mean process. We note that this assumption is more restrictive than the assumptions made in the last chapter, however, it is still applicable to situations such as pure dephasing systems and continuous phase modulation. The unitary part of the bound on Fisher information in Eq. (5.15) can then be calculated as:

$$\begin{gathered}\mathcal{J}^U_{\mu\nu} = \\ 4\int_{t_0}^{t_f} dt \int_{t_0}^{t_f} dt' \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 R_d(t,t')\partial_\mu g(t_1-t) R_2(t_1,t_2)\partial_\nu g(t_2-t'), \\ R_d(t,t') = \frac{1}{2}\,\mathrm{tr}(\rho_S\{d_H(t), d_H(t')\}), \\ R_j(t,t') = \frac{1}{2}\,\mathrm{tr}(\rho_{tot}\{x_j(t), x_j(t')\}).\end{gathered} \tag{5.65}$$

We further assume that the experiment duration $t_f - t_0$ must be long compared to all timescales of the problem. To calculate the minimum bound, we assume that $d_H(t)$ is a stationary process and make a finite time approximation by first letting $t_0 = -T/2$ and $t_f = T/2$, then take the long time limit $T \to \infty$. In terms of the unnormalized Fourier components of function $f(t)$

$$f_m = \int_{-T/2}^{T/2} f(t) e^{-i\frac{2\pi}{T}mt}, \tag{5.66}$$

$\mathcal{J}^U_{\mu\nu}$ is given by

$$\mathcal{J}^U_{\mu\nu} = 4\sum_m R_{d,m} R_{2,m} \partial_\mu g_m \partial_\nu g^*_m, \tag{5.67}$$

On the other hand, the quantum Fisher information of state ρ'_j is given by[6]:

$$J_{\mu\nu}(\rho'_j) = \frac{\partial_\mu K'_{j,m} \partial_\nu K'_{j,m}}{K'^2_{j,m} - 1/4}. \tag{5.68}$$

[6]The calculation of the quantum Fisher information is given in Appendix C

Hence, the quantum Fisher information bound $\mathcal{J}$ in Eq. (5.15) becomes

$$J(\rho) \leq \mathcal{J},$$
$$\mathcal{J}_{\mu\nu} = \sum_m \Big[4R_{d,m}R_{2,m}\partial_\mu g_m \partial_\nu g_m^* + \sum_{j=1,2} \frac{\partial_\mu K'_{j,m}\partial_\nu K'_{j,m}}{K'^2_{j,m} - 1/4} \Big]. \tag{5.69}$$

To obtain a tighter upper bound on the quantum Fisher information by varying U_{tot} and ρ'_j while keeping ρ_1 fixed, we will first need to express Eq. (5.69) in terms of the covariance matrix of ρ_1. For simplicity, we will assume that $K'_{j,m} = K'_{j,-m}$, as a result the function $K'_j(t)$ does not have an imaginary part and equals a real function

$$R'_j(t,t') = \frac{1}{2}\,\mathrm{tr}(\rho'\{x_j(t), x_j(t')\}). \tag{5.70}$$

We first study the symmetric case where $g(t) = g(-t)$.

5.3.3 Tightening the upper bound for the symmetric case

When we have a symmetric $g(t)$, from Eqs. (5.35) and (5.57),

$$\begin{aligned} R_{1,m} &= C_m^2 R'_{1,m} + S_m^2 R'_{2,m}, \\ R_{2,m} &= C_m^2 R'_{2,m} + S_m^2 R'_{1,m}. \end{aligned} \tag{5.71}$$

To simplify the analysis, we first assume that $R'_{1,m} = R'_{2,m}$, as a result $R_{2,m} = R_{1,m}$. Eq. (5.69) can then be expressed in terms of $R_{1,m}$ and the free variable g_m instead of the primed variables. As Eq. (5.69) is a matrix inequality, a tighter bound $\mathcal{J}_{tight}$ on the quantum Fisher information is given by the matrix inequality $\mathcal{J}_{tight} < \mathcal{J}$. Hence, to obtain such a tighter upper bound it is necessary to minimize $\mathcal{F} = a_\mu a_\nu \mathcal{J}_{\mu\nu}$ for all vector a_μ, while varying g_m. We note that it is not necessary to find a single $g(t)$ that would minimize the bound for all values of θ. The minimization process

can be done by first assuming that the true value is $\theta = \theta_0$, and repeat the process for other values of θ. Hence, the problem of minimizing the above bound can then be reduced to optimization problem involving the variables g_m and $\partial_\mu g_m$. Defining the following new variables:

$$A_m = a_\mu \partial_\mu \ln R_{1,m}, \tag{5.72}$$

$$B_m = a_\mu \partial_\mu \ln \chi_m, \tag{5.73}$$

$$R_m = R_{d,m} R_{1,m}, \tag{5.74}$$

$$\chi_m = C_m^2 + S_m^2 = \cosh^2 g_m + \sinh^2 g_m, \tag{5.75}$$

$\mathcal{F}$ is given by

$$\mathcal{F} = \sum_m \frac{8R_{1,m}^2 (A_m - B_m)^2}{4R_{1,m}^2 - \chi_m^2} + \frac{R_m \chi_m^2 B_m^2}{\chi_m^2 - 1}. \tag{5.76}$$

Since $\mathcal{F}$ contains a sum of functions in (B_m, χ_m), they can be minimized independently for different m. The problem of minimizing $\mathcal{F}$ becomes an optimization problem in terms of the variables (χ_m, B_m). The critical points are the roots of the simultaneous equations $\frac{\partial \mathcal{F}}{\partial \chi_m} = 0$ and $\frac{\partial \mathcal{F}}{\partial B_m} = 0$, where the minimum point is given by

$$\chi_m = \sqrt[4]{\frac{8R_{1,m}}{R_{d,m}}}, \tag{5.77}$$

$$B_m = \frac{8R_{1,m}^2(\chi_m^2 - 1)A_m}{8R_{1,m}^2(\chi_m^2 - 1) + R_m \chi_m^2 (4R_{1,m}^2 - \chi_m^2)}. \tag{5.78}$$

In order for χ_m to describe a valid unitary transformation, we have to make sure $R'_{1,m} \geq \frac{1}{2}$ so that the uncertainty principle is respected. Together with the properties of hyperbolic functions which give us the requirement $\chi_m \geq 1$, we have the following contraint on $R_{d,m}$:

$$\frac{1}{2R_{1,m}^3} \leq R_{d,m} \leq 8R_{1,m}. \tag{5.79}$$

We note that in the limit of large $R_{1,m}$ there is effectively no constraint on $R_{d,m}$ as the L.H.S. goes to zero while the R.H.S. scales with $R_{1,m}$. Thus this bound should be useful in the classical bath regime where $R_{1,m}$ is large. Assuming that the constraint is fulfilled, the best bound is given by

$$\mathcal{J}^{(1)}_{\mu\nu} = \sum_m \frac{\partial_\mu \ln R_{1,m} \partial_\nu \ln R_{1,m}}{\frac{1}{2} + \frac{1}{R_{1,m}R_{d,m}} - \frac{1}{R_{1,m}\sqrt{2R_{1,m}R_{d,m}}}} \tag{5.80}$$

$$= T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\partial_\mu \ln R_1(\omega) \partial_\nu \ln R_1(\omega)}{\frac{1}{2} + \frac{1}{R_1(\omega)R_d(\omega)} - \frac{1}{R_1(\omega)\sqrt{2R_1(\omega)R_d(\omega)}}}, \tag{5.81}$$

where the second line is obtained by making the long observation time limit $\sum_m \to T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}$. See Appendix D for proof that this bound is minimum.

Instead of having $R'_{2,m} = R'_{1,m}$, it is possible to obtain another bound by assuming that ρ'_2 does not depend on θ and $R'_{2,m} = \frac{1}{2}$. Again, we express $\mathcal{J}$ in terms of $R_{1,m}$ and g_m. From Eq. (5.69), $\mathcal{F}$ becomes

$$\mathcal{F} = \sum_m \frac{R_{d,m}B_m^2}{2}\left(2R_{1,m} + 1 + \frac{\chi_m}{\chi_m - 1}\right) + \frac{(2R_{1,m}+1)(A_m - B_m)^2}{2R_{1,m} + 1 - 2\chi_m}, \tag{5.82}$$

where we have redefined the following variables:

$$A_m = a_\mu \partial_\mu \ln\left(R_{1,m} + \frac{1}{2}\right), \tag{5.83}$$

$$B_m = 2a_\mu \partial_\mu \ln\cosh g_m, \tag{5.84}$$

$$\chi_m = C_m^2 = \cosh^2 g_m. \tag{5.85}$$

Repeating the procedures as in the case where $R'_{2,m} = R'_{1,m}$, we obtain the minimum point at

$$\chi_m = \frac{1}{2(R_{1,m}+1)}\left(2R_{1,m} + 1 + \sqrt{\frac{2R_{1,m}+1}{R_{d,m}}}\right), \tag{5.86}$$

$$B_m = \frac{H_m A_m}{H_m + G_m}, \tag{5.87}$$

where

$$H_m = 2(\chi_m - 1)(2R_{1,m} + 1), \tag{5.88}$$

$$G_m = R_{d,m}(2R_{1,m} + 1 - 2\chi_m)[2(R_{1,m} + 1)\chi_m - (2R_{1,m} + 1)]. \tag{5.89}$$

Imposing the uncertainty principle $R'_{1,m} + \frac{1}{2} = \frac{R_{1,m}+\frac{1}{2}}{\chi_m} \geq 1$ and $\chi_m \geq 1$, the constraints on the value of χ_m is given by

$$\frac{1}{R^2_{1,m}(2R_{1,m} + 1)} \leq R_{d,m} \leq 2R_{1,m} + 1. \tag{5.90}$$

The constraint behaves similarly as in the case of $R'_{2,m} = R'_{1,m}$ when $R_{1m} \gg 1$. Assuming that χ_m obey the above inequality, the minimized bound is given by

$$\mathcal{J}^{(2)}_{\mu\nu} = \sum_m \frac{\partial_\mu \ln(R_{1,m} + \frac{1}{2})\partial_\nu \ln(R_{1,m} + \frac{1}{2})}{1 + \frac{1}{R_{1,m}R_{d,m}} - \frac{2}{R_{1,m}\sqrt{R_{d,m}(2R_{1,m}+1)}}} \left(1 + \frac{1}{R_{1,m}}\right) \tag{5.91}$$

$$= T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\partial_\mu \ln(R_1(\omega) + \frac{1}{2})\partial_\nu \ln(R_1(\omega) + \frac{1}{2})}{1 + \frac{1}{R_1(\omega)R_d(\omega)} - \frac{2}{R_1(\omega)\sqrt{R_d(\omega)(2R_1(\omega)+1)}}} \left(1 + \frac{1}{R_1(\omega)}\right), \tag{5.92}$$

where we have also taken the long observation time limit in the second line. The proof for minimality is also in Appendix D.

5.3.4 Tightening the upper bound for the antisymmetric case

Lastly, we find another bound by assuming that $g(t)$ is antisymmetric where $g(-t) = -g(t)$. In this case, we have

$$R_{1,m} = C^2_m R'_{1,m} - S^2_m R'_{2,m}, \tag{5.93}$$

$$R_{2,m} = C^2_m R'_{2,m} - S^2_m R'_{1,m}. \tag{5.94}$$

First, we note that if we assume $R'_1 = R'_2$, we have $R_1 = R_2 = R'_1$. This is uninteresting as $J(\rho_1) = J(\rho'_j)$ and $\mathcal{J}^U \geq 0$. The minimum bound is then given by the case where $\partial_\mu g(t-t') = 0$ and $\mathcal{J}^U = 0$. The best bound on the quantum Fisher information is then simply the quantum Fisher information of the initial bath density matrix, $J(\rho_1)$. We proceed to the case where $R'_{2,m} = 1/2$. Redefining the new variables

$$A_m = a_\mu \partial_\mu \ln\left(R_{1,m} - \frac{1}{2}\right), \tag{5.95}$$

$$B_m = 2a_\mu \partial_\mu \ln C_m, \tag{5.96}$$

$$\chi_m = C_m^2 = \cosh^2 g_m, \tag{5.97}$$

$\mathcal{F}$ is given by

$$\mathcal{F} = \sum_m R_{d,m} B_m^2 \left(R_{1,m} - \frac{1}{2} - \frac{\chi_m}{2(\chi_m - 1)}\right) + \frac{4(R_{1,m} - \frac{1}{2})(A_m - B_m)^2}{4(R_{1,m} - \frac{1}{2} + \frac{\chi_m}{2})^2 - \chi_m^2}. \tag{5.98}$$

The critical points are similarly given by the solutions of the simultaneous equations $\frac{\partial \mathcal{F}}{\partial B_m} = 0 = \frac{\partial \mathcal{F}}{\partial \chi_m}$, where the minimum point is given by

$$\chi_m = \frac{1}{2(R_{1,m} - 1)} \left(2R_{1,m} - 1 - \sqrt{\frac{2R_{1,m} - 1}{R_{d,m}}}\right), \tag{5.99}$$

$$B_m = \frac{H_m A_m}{H_m + G_m}, \tag{5.100}$$

where

$$H_m = 2(\chi_m - 1)(2R_{1,m} - 1), \tag{5.101}$$

$$G_m = R_{d,m}[2R_{1,m} - 1 + 2\chi_m][(2R_{1,m} - 1)(\chi_m - 1) - \chi_m]. \tag{5.102}$$

We give the proof that this critical point is a minimum point when $R_{1,m} > 1$ in the Appendix. χ_m is a valid transformation when $0 \leq \chi_m \leq 1$, or

equivalently when

$$\frac{1}{2R_{1,m}-1} \leq R_{d,m} \leq 2R_{1,m}-1. \tag{5.103}$$

We note that the uncertainty relation on $R'_{1,m}$ does not impose new constraint to the problem. Similar to the symmetric case, the bound should be most useful when $R_{1,m} \gg 1$. Assuming that the constraints on $R_{d,m}$ is met, the minimum bound is then given by

$$\mathcal{J}^{(3)}_{\mu\nu} = \sum_m \frac{[\partial_\mu \ln(R_{1,m}-1/2)][\partial_\nu \ln(R_{1,m}-1/2)]}{1+\frac{1}{R_{1,m}R_{d,m}}-\frac{2}{R_{1,m}\sqrt{R_{d,m}(2R_{1,m}-1)}}}\left(1-\frac{1}{R_{1,m}}\right), \tag{5.104}$$

$$= T\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{[\partial_\mu \ln(R_1(\omega)-1/2)][\partial_\nu \ln(R_1(\omega)-1/2)]}{1+\frac{1}{R_1(\omega)R_d(\omega)}-\frac{2}{R_1(\omega)\sqrt{R_d(\omega)(2R_1(\omega)-1)}}}\left(1-\frac{1}{R_1(\omega)}\right). \tag{5.105}$$

5.4 Discussion and Conclusion

We proposed a generalization of our results in Chap. 4 to the case where the stochastic process $X(t)$ is replaced by a quantum stochastic process $x(t)$. We assumed that the statistics of $x(t)$ is given by a quantum Gaussian state, where its power spectral density depends on some unknown parameters. Because of the assumption we made on $d_H(t)$, the applicability of our bounds is more limited than the one in Chap. 4. Nevertheless, it should be applicable for pure dephasing models as well as continuous phase modulation problems, where the noise is of quantum origin. A possible future direction is to study whether the assumption on $d_H(t)$ can be lifted to include more complex systems. It would also be interesting to make comparisons between our theory and experiments, as well as study experiment methods and conditions to reach the quantum bounds we derived.

On the other hand, we followed a modified variational approach which is similar to the one in Chap. 4 to derive our upper bounds. The main

ingredient of our recipe is the modified extended convexity for the channel defined in Eq. (5.1) and a variational procedure based on equivalent families of Eq. (5.1). By considering three different families, we found three different upper bounds to the quantum Fisher information. A price to pay for going from a classical bath to a quantum bath is that the quantum bath must satisfy the Heisenberg uncertainty principle. This requirement is manifested in the conditions in which our derived bounds are valid and it is interesting to see whether upper bounds that don't suffer from these limitations exists.

We have assumed the simplified evolution given by Eq. (5.43) throughout this chapter. However, the more general evolution in Eq. (5.39) is more suitable to describe the case of quantum dissipation [19, 138]. Hence, more efforts should be invested into studying quantum Fisher information bounds for these models. The main difficulty lies in taking the derivative of the time-ordered exponential $\tilde{U}$, which now contains time-integrals that do not commute. A discrete time approximation like the one we used in Eqs. (5.59) can be applied. However, care must be taken as the time integral now involves field operator $b(t)$ which has a delta function commutator. After making the appropriate discrete time approximation, the time-ordered exponential should be given in terms of multiplications of exponential operators as in Eq. (5.59). The method of taking derivatives of an exponential operator found in [143] can then be used to evaluate the derivative of $\tilde{U}$.

Chapter 6

Thesis conclusion

In this thesis, we studied detection and estimation theory for continuously measured quantum systems. We will give an overall summary in this chapter, and conclude the thesis with a discussion of future outlook.

6.1 Summary

To start our discussions, we gave a review of theoretical backgrounds in Chap. 2. We introduced the readers to stochastic calculus, in particular the Itô calculus, which is widely used to analyze stochastic processes. We gave an account of various concepts of detection and estimation theory, for example, the likelihood ratio tests and the Fisher information. We highlighted the goal of detection and estimation theory, which is to extract information by improving the data analysis process and optimizing experimental techniques.

We presented the first problem we studied in Chap. 3, where we explored the qubit readout problem of determining the initial state of a continuously measured qubit. In contrast to earlier studies, we considered two different noise models, namely the Gaussian and Poissonian noise models. The qubit readout problem is framed as a hypothesis testing problem, and we proposed to use the optimal likelihood ratio test to process the

measurement data. The likelihood ratio test involves solving a stochastic differential equation, and we have found analytic solutions for the cases where there is no excitation or no relaxation in the qubit dynamics. We note that numerical solutions are available for the other cases. Another important feature of our work is the consistent use of Itô calculus, which makes our protocol easier to implement and less likely to cause confusion over the interpretation of stochastic integrals in our protocol.

In Chap. 4, we studied the problem of estimating the spectrum parameters of a classical stochastic processes. We assumed that the stochastic process is coupled to a quantum system, and measurement is made on the quantum system to infer the value of the spectrum parameter. We found a fundamental limit to the accuracy of spectrum-parameter estimation based on the quantum Cramêr-Rao bound. To compare our theory with experiments, we analyzed an experiment of continuous optical phase estimation and demonstrated that the experimental performance using homodyne detection is close to our quantum limit. In the case of weak modulation and a coherent state input, we proposed the spectral photon counting method to outperform homodyne detection and attain quantum-optimal performance.

Lastly, we provided a generalization to the problem of spectrum-parameter estimation by considering quantum stochastic processes. We considered a simplified model of interaction between a quantum system and a quantum bath, and found quantum limits similar to the ones in Chap. 4. Because of the assumption that the stochastic process is quantum, we find that our derived bounds is valid only in certain regimes.

6.2 Future outlook

In Chap. 3, we have assumed that the continuous measurement made on the qubit is quantum nondemolition and replaced the quantum model with

an equivalent classical one. For more general quantum dynamics and measurements, it would be interesting to see how one applies the formalism in [78]. Another open problem is the evaluation of the performance of our protocol beyond the case of deterministic signal detection. A possible approach is to look for upper bounds on the error probability, for example in terms of the Chernoff coefficient.

Although we have provided an extension to the theory in Chap. 4, our theory for quantum bath in Chap. 5 suffers from several limitations. For instance, the derived limits are only valid under certain conditions. Our theory also only works for specific system-bath interactions which include pure dephasing systems and continuous phase modulation experiments. Further work needs to be done to lift these limitations, as well as to extend our results to the more interesting case of dissipation caused by a quantum bath. Since our derived limits are based on an upper bound on the quantum Fisher information, it is unclear if the limits are achievable by any experimental method. It is then important to study the conditions under which these quantum limits are achievable.

Throughout this thesis, we consider only the case of nonrandom parameter estimation. We have already encountered random parameter estimation in Chap. 3, and other possible scenarios include stochastic waveform estimation. For these problems, the mean-square error of an estimator is bounded by Bayesian bounds [110, 133, 134, 144, 145]. Interestingly, a lot of the limitations of nonrandom parameter estimation are lifted in the Bayesian formulation. For example, to apply the Bayesian Cramér-Rao bound, we do not need to require the estimator to be unbiased. It would be interesting to seek an extension of our results to the case of random parameter estimation.

Bibliography

[1] M. Schlosshauer, *Decoherence: and the Quantum-To-Classical Transition* (Springer-Verlag, Berlin, 2007).

[2] W. H. Zurek, Rev. Mod. Phys. **75**, 715 (2003).

[3] W. H. Zurek, Nat Phys **5**, 181 (2009).

[4] J. Taylor, *An Introduction to Measure and Probability* (Springer-Verlag, New York, 1998).

[5] M. Grigoriu, *Stochastic Calculus: Applications in Science and Engineering* (Birkhuser Basel, 2002).

[6] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Scuola Normale Superiore, Pisa, 2011).

[7] C. W. Helstrom, *Quantum detection and estimation theory* (Academic Press, New York, 1976).

[8] A. Barchielli and M. Gregoratti, *Quantum Trajectories and Measurements in Continuous Time: The Diffusive Case* (Springer-Verlag, Berlin, 2011).

[9] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[10] H. Wiseman and G. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010).

[11] M. Tsang, R. Nair, and X.-M. Lu, Phys. Rev. X **6**, 031033 (2016).

[12] M. Tsang, R. Nair, and X.-M. Lu (2016) pp. 1002903–1002903–7.

[13] R. Nair and M. Tsang, Opt. Express **24**, 3684 (2016).

[14] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

[15] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[16] J. J. . Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A **54**, R4649 (1996).

[17] A. N. Boto, P. Kok, D. S. Abrams, S. L. Braunstein, C. P. Williams, and J. P. Dowling, Phys. Rev. Lett. **85**, 2733 (2000).

[18] M. Tsang, Phys. Rev. Lett. **102**, 253601 (2009).

[19] D. F. Walls and G. J. Milburn, *Quantum optics* (Springer-Verlag, Berlin, 2008).

[20] P. Meystre, Annalen der Physik **525**, 215 (2013).

[21] M. Aspelmeyer, T. J. Kippenberg, and F. Marquardt, Rev. Mod. Phys. **86**, 1391 (2014).

[22] Y. Chen, Journal of Physics B: Atomic, Molecular and Optical Physics **46**, 104001 (2013).

[23] T. Westphal, D. Friedrich, H. Kaufer, K. Yamamoto, S. Goßler, H. Müller-Ebhardt, S. L. Danilishin, F. Y. Khalili, K. Danzmann, and R. Schnabel, Phys. Rev. A **85**, 063806 (2012).

[24] G. Anetsberger, E. Gavartin, O. Arcizet, Q. P. Unterreithmeier, E. M. Weig, M. L. Gorodetsky, J. P. Kotthaus, and T. J. Kippenberg, Phys. Rev. A **82**, 061804 (2010).

[25] S. Schreppler, N. Spethmann, N. Brahms, T. Botter, M. Barrios, and D. M. Stamper-Kurn, Science **344**, 1486 (2014).

[26] B. P. Abbott et. al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016).

[27] R. Miller, T. E. Northup, K. M. Birnbaum, A. Boca, A. D. Boozer, and H. J. Kimble, Journal of Physics B: Atomic, Molecular and Optical Physics **38**, S551 (2005).

[28] R. Blatt and P. Zoller, European Journal of Physics **9**, 250 (1988).

[29] M. Acton, K.-A. Brickman, P. C. Haljan, P. J. Lee, L. Deslauriers, and C. Monroe, Quantum Info. Comput. **6**, 465 (2006).

[30] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Rev. Mod. Phys. **82**, 1155 (2010).

[31] K. Jacobs and D. A. Steck, Contemporary Physics **47**, 279 (2006).

[32] H. L. Van Trees, K. L. Bell, and Z. Tian, *Detection estimation and modulation theory. Part I*, 2nd ed. (Wiley, Hoboken, N.J., 2013).

[33] B. Levy, *Principles of Signal Detection and Parameter Estimation* (Springer US, New York, 2008).

[34] S. Kay, *Fundamentals of Statistical Signal Processing: Detection theory* (Prentice-Hall PTR, 1998).

[35] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* (John Wiley & Sons, Hoboken, New Jersey, 2006).

[36] M. Hayashi, *Quantum Information: An Introduction* (Springer-Verlag, Berlin, 2010).

[37] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer-Verlag, Berlin, 2009).

[38] X. Mao, *Stochastic Differential Equations and Applications* (Woodhead, Oxford, 2007).

[39] N. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland Personal Library (Elsevier, 2007).

[40] D. Snyder and M. Miller, *Random Point Processes in Time and Space* (Springer-Verlag, 1991).

[41] T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley & Sons, Hoboken, New Jersey, 2012).

[42] T. Kailath and H. V. Poor, IEEE Transactions on Information Theory **44**, 2230 (1998).

[43] T. Kailath, IEEE Transactions on Information Theory **15**, 350 (1969).

[44] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise*, Detection, Estimation, and Modulation Theory (John Wiley & Sons, New York, 2001).

[45] T. Kailath, IEEE Transactions on Communication Technology **15**, 52 (1967).

[46] M. Tsang, Quantum Measurements and Quantum Metrology **1**, 84 (2013).

[47] D. Qiu, Phys. Rev. A **77**, 012328 (2008).

[48] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Phys. Rev. Lett. **98**, 160501 (2007).

[49] M. Nussbaum and A. Szkoa, Ann. Statist. **37**, 1040 (2009).

[50] H. Yuen and M. Lax, IEEE Transactions on Information Theory **19**, 740 (1973).

[51] V. Belavkin, Quant-ph/0412030v1.

[52] M. Hayashi, *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (World Scientific, 2005).

[53] D. Petz, Journal of Physics A: Mathematical and General **35**, 929 (2002).

[54] P. Gibilisco, F. Hiai, and D. Petz, IEEE Transactions on Information Theory **55**, 439 (2009).

[55] A. Fujiwara, Journal of Physics A: Mathematical and General **39**, 12489 (2006).

[56] M. Hübner, Physics Letters A **163**, 239 (1992).

[57] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **95**, 060501 (2005).

[58] A. Lupascu, S. Saito, T. Picot, P. C. de Groot, C. J. P. M. Harmans, and J. E. Mooij, Nat Phys **3**, 119 (2007).

[59] J. E. Johnson, C. Macklin, D. H. Slichter, R. Vijay, E. B. Weingarten, J. Clarke, and I. Siddiqi, Phys. Rev. Lett. **109**, 050506 (2012).

[60] D. B. Hume, T. Rosenband, and D. J. Wineland, Phys. Rev. Lett. **99**, 120502 (2007).

[61] A. H. Myerson, D. J. Szwer, S. C. Webster, D. T. C. Allcock, M. J. Curtis, G. Imreh, J. A. Sherman, D. N. Stacey, A. M. Steane, and D. M. Lucas, Phys. Rev. Lett. **100**, 200502 (2008).

[62] P. Neumann, J. Beck, M. Steiner, F. Rempp, H. Fedder, P. R. Hemmer, J. Wrachtrup, and F. Jelezko, Science **329**, 542 (2010).

[63] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. A. Alkemade, and R. Hanson, Nature **477**, 574 (2011).

[64] J. M. Elzerman, R. Hanson, L. H. Willems van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, Nature **430**, 431 (2004).

[65] A. N. Vamivakas, C.-Y. Lu, C. Matthiesen, Y. Zhao, S. Falt, A. Badolato, and M. Atature, Nature **467**, 297 (2010).

[66] A. Morello, J. J. Pla, F. A. Zwanenburg, K. W. Chan, K. Y. Tan, H. Huebl, M. Mottonen, C. D. Nugroho, C. Yang, J. A. van Donkelaar, A. D. C. Alves, D. N. Jamieson, C. C. Escott, L. C. L. Hollenberg, R. G. Clark, and A. S. Dzurak, Nature **467**, 687 (2010).

[67] J. J. Pla, K. Y. Tan, J. P. Dehollain, W. H. Lim, J. J. L. Morton, F. A. Zwanenburg, D. N. Jamieson, A. S. Dzurak, and A. Morello, Nature **496**, 334 (2013), letter.

[68] L. Rondin, J.-P. Tetienne, T. Hingant, J.-F. Roch, P. Maletinsky, and V. Jacques, Reports on Progress in Physics **77**, 056503 (2014).

[69] P. O. Schmidt, T. Rosenband, C. Langer, W. M. Itano, J. C. Bergquist, and D. J. Wineland, Science **309**, 749 (2005).

[70] C. W. Chou, D. B. Hume, T. Rosenband, and D. J. Wineland, Science **329**, 1630 (2010).

[71] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **76**, 012325 (2007).

[72] B. D'Anjou and W. A. Coish, Phys. Rev. A **89**, 012313 (2014).

[73] T. E. Duncan, Information and Control **13**, 62 (1968).

[74] D. Snyder, IEEE Transactions on Information Theory **18**, 91 (1972).

[75] V. Braginsky, F. Khalili, and K. Thorne, *Quantum Measurement* (Cambridge University Press, 1992).

[76] M. Tsang and C. M. Caves, Phys. Rev. X **2**, 031016 (2012).

[77] J. Gough and M. R. James, IEEE Transactions on Automatic Control **54**, 2530 (2009).

[78] M. Tsang, Phys. Rev. Lett. **108**, 170502 (2012).

[79] J. Stockton, M. Armen, and H. Mabuchi, J. Opt. Soc. Am. B **19**, 3019 (2002).

[80] J. Gambetta and H. M. Wiseman, Phys. Rev. A **64**, 042105 (2001).

[81] B. A. Chase and J. M. Geremia, Phys. Rev. A **79**, 022314 (2009).

[82] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk, Phys. Rev. Lett. **105**, 170501 (2010).

[83] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, Phys. Rev. A **89**, 042314 (2014).

[84] J. Combes, C. Ferrie, C. Cesare, M. Tiersch, G. J. Milburn, H. J. Briegel, and C. M. Caves, arXiv:1405.5656 [quant-ph] .

[85] J. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York, 1985).

[86] T. E. Duncan, Ph.D. thesis, Standford University (1967).

[87] R. E. Mortensen, Ph.D. thesis, University of California, Berkeley (1966).

[88] M. Zakai, Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete **11**, 230 (1969).

[89] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems* (Springer-Verlag, New York, 1985).

[90] W. M. Wonham, Journal of the Society for Industrial and Applied Mathematics Series A Control **2**, 347 (1964).

[91] D. J. Higham, X. Mao, and A. M. Stuart, SIAM Journal on Numerical Analysis **40**, 1041 (2002).

[92] P. Kloeden, E. Platen, and H. Schurz, *Numerical Solution of SDE Through Computer Experiments* (Springer, Berlin, 2003).

[93] R. L. Cook, C. A. Riofrío, and I. H. Deutsch, Phys. Rev. A **90**, 032113 (2014).

[94] J. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation* (Wiley, New York, 1990).

[95] M. Tsang and R. Nair, Phys. Rev. A **86**, 042115 (2012).

[96] M. Tsang, in *IEEE International Symposium on Information Theory(ISIT), Honolulu, HI, USA* (IEEE, Piscataway, NJ, 2014) pp. 321–325.

[97] M. Tsang, Quantum Measurements and Quantum Metrology **1**, 84.

[98] M. Tsang, Phys. Rev. A **92**, 062119 (2015).

[99] B. D'Anjou, L. Kuret, L. Childress, and W. A. Coish, Phys. Rev. X **6**, 011017 (2016).

[100] The LIGO Scientific Collaboration and The Virgo Collaboration, Nature **460**, 990 (2009).

[101] B. P. Abbott et. al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 131102 (2016).

[102] S. Nimmrichter, K. Hornberger, and K. Hammerer, Phys. Rev. Lett. **113**, 020405 (2014).

[103] L. Diósi, Phys. Rev. Lett. **114**, 050403 (2015).

[104] T. M. Stace, Phys. Rev. A **82**, 011611 (2010).

[105] U. Marzolino and D. Braun, Phys. Rev. A **88**, 063609 (2013).

[106] M. Jarzyna and M. Zwierz, Phys. Rev. A **92**, 032112 (2015).

[107] L. A. Correa, M. Mehboudi, G. Adesso, and A. Sanpera, Phys. Rev. Lett. **114**, 220405 (2015).

[108] R. Nair and M. Tsang, The Astrophysical Journal **808**, 125 (2015).

[109] B. P. Abbott et. al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 131103 (2016).

[110] M. Tsang, H. M. Wiseman, and C. M. Caves, Phys. Rev. Lett. **106**, 090401 (2011).

[111] M. Tsang, New Journal of Physics **15**, 073005 (2013).

[112] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, New Journal of Physics **14**, 103013 (2012).

[113] M. Brunelli, S. Olivares, and M. G. A. Paris, Phys. Rev. A **84**, 032105 (2011).

[114] C. Benedetti, F. Buscemi, P. Bordone, and M. G. A. Paris, Phys. Rev. A **89**, 032114 (2014).

[115] S. Gammelmark and K. Mølmer, Phys. Rev. Lett. **112**, 170401 (2014).

[116] S. Alipour and A. T. Rezakhani, Phys. Rev. A **91**, 042104 (2015).

[117] O. E. Barndorff-Nielsen and R. D. Gill, Journal of Physics A: Mathematical and General **33**, 4481 (2000).

[118] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, Nature Communications **3**, 1063 EP (2012), article.

[119] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, Nat Phys **7**, 406 (2011).

[120] L. T. Hall, J. H. Cole, C. D. Hill, and L. C. L. Hollenberg, Phys. Rev. Lett. **103**, 220802 (2009).

[121] A. De Pasquale, D. Rossini, P. Facchi, and V. Giovannetti, Phys. Rev. A **88**, 052117 (2013).

[122] 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, Phys. Rev. Lett. **104**, 093601 (2010).

[123] 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, Science **337**, 1514 (2012).

[124] K. Iwasawa, K. Makino, H. Yonezawa, M. Tsang, A. Davidovic, E. Huntington, and A. Furusawa, Phys. Rev. Lett. **111**, 163602 (2013).

[125] R. Shumway and D. Stoffer, *Time Series Analysis and Its Applications: With R Examples*, Springer Texts in Statistics (Springer, New York, 2006).

[126] M. Tsang, Phys. Rev. Lett. **107**, 270402 (2011).

[127] J. H. Shapiro, Quantum and Semiclassical Optics: Journal of the European Optical Society Part B **10**, 567 (1998).

[128] D. Brady, *Optical Imaging and Spectroscopy* (Wiley, Hoboken, NJ, 2009).

[129] S. T. Chu, B. E. Little, W. Pan, T. Kaneko, S. Sato, and Y. Kokubun, IEEE Photonics Technology Letters **11**, 691 (1999).

[130] S. Z. Ang, G. I. Harris, W. P. Bowen, and M. Tsang, New Journal of Physics **15**, 103028 (2013).

[131] P. Whittle, Journal of the Royal Statistical Society. Series B (Methodological) **15**, 125 (1953).

[132] M. Tsang, Phys. Rev. A **88**, 021801 (2013).

[133] M. Tsang, Phys. Rev. Lett. **108**, 230401 (2012).

[134] D. W. Berry, M. Tsang, M. J. W. Hall, and H. M. Wiseman, Phys. Rev. X **5**, 031018 (2015).

[135] 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, Science **349**, 711 (2015).

[136] D. Gottesman, T. Jennewein, and S. Croke, Phys. Rev. Lett. **109**, 070503 (2012).

[137] C. W. Gardiner and M. J. Collett, Phys. Rev. A **31**, 3761 (1985).

[138] C. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics* (Springer-Verlag, Berlin, 2000).

[139] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Rev. Mod. Phys. **84**, 621 (2012).

[140] H. A. S., *Quantum Systems, Channels, Information, A Mathematical Introduction* (Berlin/Boston, 2012).

[141] B. M. Escher, L. Davidovich, N. Zagury, and R. L. de Matos Filho, Phys. Rev. Lett. **109**, 190404 (2012).

[142] A. Fujiwara and H. Nagaoka, Physics Letters A **201**, 119 (1995).

[143] R. M. Wilcox, Journal of Mathematical Physics **8**, 962 (1967).

[144] H. Van Trees and K. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (Wiley-IEEE Press, 2007).

[145] X.-M. Lu and M. Tsang, Quantum Science and Technology **1**, 015002 (2016).

Appendix A

Quantum formalism of continuous quantum nondemolition measurement for qubits

Let

$$\hat{f}_m(t) = \begin{pmatrix} f_m(0,0,t) & f_m(0,1,t) \\ f_m(1,0,t) & f_m(1,1,t) \end{pmatrix} \tag{A.1}$$

be the unnormalized density matrix for the qubit conditioned on the observation record Y^t and hypothesis $\mathcal{H}_m$. Consider the following linear stochastic quantum master equation [10]:

$$\begin{aligned} d\hat{f}_m = {} & dtL_m^- \left(\hat{\sigma}_- \hat{f}_m \hat{\sigma}_+ - \frac{1}{2}\hat{\sigma}_+\hat{\sigma}_-\hat{f}_m - \frac{1}{2}\hat{f}_m\hat{\sigma}_+\hat{\sigma}_- \right) \\ & + dtL_m^+ \left(\hat{\sigma}_+ \hat{f}_m \hat{\sigma}_- - \frac{1}{2}\hat{\sigma}_-\hat{\sigma}_+\hat{f}_m - \frac{1}{2}\hat{f}_m\hat{\sigma}_-\hat{\sigma}_+ \right) \\ & + dtL_m^x \left(\hat{x} \hat{f}_m \hat{x} - \frac{1}{2}\hat{x}^2\hat{f}_m - \frac{1}{2}\hat{f}_m\hat{x}^2 \right) \\ & + \frac{dy\sigma_m}{2} \left(\hat{x}\hat{f}_m + \hat{f}_m\hat{x} \right), \end{aligned} \tag{A.2}$$

where

$$\hat{\sigma}_- = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \hat{\sigma}_+ = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \hat{x} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{A.3}$$

and L_m^-, L_m^+, and $L_m^x \geq \sigma_m^2/4$ are the decay, excitation, and dephasing rates, respectively. The estimator in the quantum estimator-correlator formula [78] is

$$\sigma_m(t)\mathbb{E}\left(\hat{x}|Y^t, \mathcal{H}_m\right) = \frac{\sigma_m(t) f_m(1,1,t)}{f_m(0,0,t) + f_m(1,1,t)}. \tag{A.4}$$

The important point here is that the estimator involves only the diagonal components of $\hat{f}_m(t)$, which are decoupled from the off-diagonal components throughout the evolution:

$$df_m(0,0,t) = dt\left[-L_m^+(t) f_m(0,0,t) + L_m^-(t) f_m(1,1,t)\right], \tag{A.5}$$

$$\begin{aligned} df_m(1,1,t) = {} & dt\left[L_m^+(t) f_m(0,0,t) - L_m^-(t) f_m(1,1,t)\right] \\ & + dy(t)\sigma_m(t) f_m(1,1,t). \end{aligned} \tag{A.6}$$

This means that a classical stochastic model is sufficient. In particular, Eqs. (A.5) and (A.6) are identical to the classical DMZ equation given by Eq. (3.20). The argument in the case of Poissonian noise is similar.

Appendix B

Experimental data recalibration for Chap. 4

In the experiment described in Ref. [122], calibration procedures were used to convert applied and measured voltages to the various physical quantities defined throughout Ref. [122]. In the course of analysing that experimental data for the purposes 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. [122] – 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. [122]. 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 the purpose of this new estimation task, we refine the calibration of the data from Ref. [122] so that we can achieve an accurate estimate. To do this in a fair way we use two extra data sets ($k = 5, 6$), which were not shown in Ref. [122] 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. [122]. 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 refined calibration 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. [122] by making use of independent, but contemporaneously recorded data.

Appendix C

A special class of thermal states for Chap. 5

Let $b(t)$ be the field operators such that $[b(t), b^\dagger(t')] = \delta(t-t')$, we approximate the bath state (5.46) by a finite time approximation:

$$\begin{aligned}\rho_B^T &= \exp[-\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} G(t-t')b^\dagger(t)b(t')dtdt']\mathcal{Z}_B \\ &= \exp[-\sum_m G_m B_m^\dagger B_m]\mathcal{Z}_B, \end{aligned} \tag{C.1}$$

where $\rho_B = \lim_{T\to\infty}\rho_B^T$ and $\mathcal{Z}_B = \prod_m[1-\exp(-G_m)]$ is a normalization constant. B_m is the discrete Fourier modes while G_m is the unnormalized Fourier component of $G(t)$:

$$B_m = \frac{1}{\sqrt{T}}\int_{-T/2}^{T/2} b(t)e^{-i\frac{2\pi m}{T}t}dt, \tag{C.2}$$

$$G(t) = \sum_m \nu_{G,m}e^{i\frac{2\pi m}{T}t}, \tag{C.3}$$

$$G_m = T\nu_{G,m} = \int_{-T/2}^{T/2} G(t)e^{-i\frac{2\pi m}{T}t}dt. \tag{C.4}$$

Due to hermiticity of the density operators, the coefficients $\nu_{G,m}$ must be real. Also, the discrete Fourier modes B_m are independent modes as $[B_m, B_n^\dagger] = \delta_{mn}$. Thus ρ_B^T represents a product thermal state in the dis-

crete Fourier basis.

Let us define the quadratures of these Fourier modes by $x_m = \frac{1}{\sqrt{2}}(B_m + B_m^\dagger)$ and $p_m = \frac{1}{\sqrt{2}i}(B_m - B_m^\dagger)$, the quantum moments are given by:

$$\mathrm{tr}(\rho_B^T \frac{1}{2}\{x_m, x_n\}) = \mathrm{tr}(\rho_B^T \frac{1}{2}\{p_m, p_n\}) = K_j \delta_{mn}, \tag{C.5}$$

$$\mathrm{tr}(\rho_B^T \frac{1}{2}\{x_m, p_n\}) = 0, \tag{C.6}$$

$$\bar{N}_m = \mathrm{tr}(\rho_B^T N_m) = K_m - 1/2, \tag{C.7}$$

$$\mathrm{tr}[\rho_B^T (N_m - \bar{N}_m)(N_n - \bar{N}_n)] = (K_j^2 - 1/4)\delta_{mn}. \tag{C.8}$$

where $K_m = \frac{1}{2}\coth(\frac{G_m}{2})$.

Let $x(t) = \frac{1}{\sqrt{2}}[b(t) + b^\dagger(t)]$ and $p(t) = \frac{1}{\sqrt{2}i}[b(t) - b^\dagger(t)]$, and note that we can write $b(t) = \frac{1}{\sqrt{T}}\sum_m B_m e^{i\frac{2\pi m}{T}t}$. Denote

$$Q(t) = \begin{pmatrix} x(t) \\ p(t) \end{pmatrix}, \tag{C.9}$$

we can express $Q(t)$ as

$$Q(t) = \frac{1}{\sqrt{T}} \sum_m M(m,t) Q_m, \tag{C.10}$$

$$M(m,t) = \begin{pmatrix} \cos(\frac{2\pi m}{T}t) & -\sin(\frac{2\pi m}{T}t) \\ \sin(\frac{2\pi m}{T}t) & \cos(\frac{2\pi m}{T}t) \end{pmatrix}, \tag{C.11}$$

$$Q_m = \begin{pmatrix} x_m \\ p_m \end{pmatrix}. \tag{C.12}$$

Define $\alpha_{m_1 n_1}(t,t') = \mathrm{tr}[\rho_B^T \frac{1}{2}\{Q_{m_1}(t), Q_{n_1}(t')\}$,

$$\begin{aligned}\alpha_{m_1,n_1}(t,t') &= \frac{1}{T}\sum_{m,n,m_2,n_2} M_{m_1,m_2}(m,t)M_{n_1,n_2}(n,t')\,\mathrm{tr}[\rho_B^T \frac{1}{2}\{Q_{m,m_2}, Q_{n,n_2}\}] \\ &= \frac{1}{T}\sum_m [M(m,t)M^\top(m,t')]_{m_1,n_1} K_m \\ &= \frac{1}{T}\sum_m M_{m_1,n_1}(m,t-t')K_m, \end{aligned} \tag{C.13}$$

and thus

$$\alpha(t,t') = \begin{pmatrix} \Re\{K(t-t')\} & -\Im\{K(t-t')\} \\ \Im\{K(t-t')\} & \Re\{K(t-t')\} \end{pmatrix}, \tag{C.14}$$

$$K(t) = \sum_m \frac{K_m}{T} e^{i\frac{2\pi m}{T}t}. \tag{C.15}$$

By taking the long time limit we then have $K(t) = \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \coth(\frac{G(\omega)}{2}) e^{i\omega t}$, and proved the assertion that $K(t)$ characterize the density matrix ρ_B completely. Note that $\Re\{K(-t)\} = \Re\{K(t)\}$ and $\Im\{K(-t)\} = -\Im\{K(t)\}$.

Let's consider the case where the function $G(t,t')$ in Eq. (C.1) depends on the parameter θ, the quantum Fisher information matrix can be easily calculated, by first finding the s.l.d.:

$$\begin{aligned}\partial_\mu \rho_B^T &= \sum_m [-\partial_\mu G_m B_m^\dagger B_m + \partial_\mu G_m \bar{N}_m]\rho_B^T \\ &= -\sum_m \partial_\mu G_m [N_m - \bar{N}_m]\rho_B^T \\ &= -\sum_m \frac{\partial_\mu G_m}{2}[N_m - \bar{N}_m]\rho_B^T - \rho_B^T \sum_m \frac{\partial_\mu G_m}{2}[N_m - \bar{N}_m]. \end{aligned} \tag{C.16}$$

Therefore, we have $L_\mu = -\sum_m \partial_\mu G_m [N_m - \bar{N}_m]$. The quantum Fisher

information is given by

$$\begin{aligned} J_{\mu\nu}(\rho_B^T) &= \mathrm{tr}[\rho_B^T \frac{1}{2}\{L_\mu, L_\nu\}] \\ &= \sum_{m,n} \partial_\mu G_m \partial_\nu G_n [K_m^2 - 1/4]\delta_{mn}, \\ &= \sum_{m} \partial_\mu G_m \partial_\nu G_m [K_m^2 - 1/4] \\ &= \sum_{m} \frac{\partial_\mu K_m \partial_\nu K_m}{K_m^2 - 1/4}, \end{aligned} \tag{C.17}$$

since $\partial_\mu \bar{N}_m = -[K_m^2 - 1/4]\partial_\mu G_m$. Under the long time limit, we have

$$J_{\mu\nu}(\rho_B) = T \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\partial_\mu K(\omega) \partial_\nu K(\omega)}{K^2(\omega) - 1/4}. \tag{C.18}$$

Appendix D

Proofs for the upper bounds for Chap. 5

D.1 The symmetric case and $R_2' = R_1'$

First, we note that when $\chi_m = 1$, $R'_{1,m} = R_{1,m}$ and $\partial_\mu R'_{1,m} = \partial_\mu R_{1,m}$ from Eqs. (5.71). According to Eq. (5.69) we have the obvious bound $\mathcal{J} > j(\rho_1)$, which does not provide any insight. When $\chi_m = 2R_{1,m}$, the only finite case is when $\partial_\mu R'_{1,m}$. Eq. (5.69) becomes

$$\mathcal{J}_{\mu\nu} = \sum_m \frac{\partial_\mu \ln R_{1,m} \partial_\nu \ln R_{1,m}}{\frac{1}{R_{d,m}R_{1,m}} - \frac{1}{4R_{1,m}^3 R_{d,m}}}. \tag{D.1}$$

This bound is less tight than $\mathcal{J}^{(1)}$ in Eq. (5.80).

We proceed to find tighter bound for other cases by defining

$$F_m = \frac{H_m(A_m - B_m)^2 + G_m B_m^2}{(\chi_m^2 - 1)(4R_{1,m}^2 - \chi_m^2)}, \tag{D.2}$$

$$G_m = R_m \chi_m^2 (4R_{1,m}^2 - \chi_m^2), \tag{D.3}$$

$$H_m = 8R_{1,m}^2(\chi_m^2 - 1). \tag{D.4}$$

The equations determining the critical points are given by $\frac{\partial \mathcal{F}}{\partial B_m} = 0$ and

$\frac{\partial \mathcal{F}}{\partial \chi_m} = 0$:

$$\frac{G_m B_m - H_m(A_m - B_m)}{(\chi_m^2 - 1)(4R_{1,m}^2 - \chi_m^2)} = 0, \tag{D.5}$$

$$\frac{2\chi_m}{(\chi_m^2 - 1)(4R_{1,m}^2 - \chi_m^2)}\big[8R_{1,m}^2(A_m - B_m)^2 + 2C_m B_m^2(2R_{1,m}^2 - \chi_m^2) - F_m(4R_{1,m}^2 - 2\chi_m^2 + 1)\big] = 0. \tag{D.6}$$

Assuming $\chi_m^2 \neq 1$ and $\chi_m^2 \neq 4R_{1,m}^2$, we solve the equations by first expressing B_m as a function of χ_m from Eq. (D.5):

$$B_m = \frac{H_m A_m}{H_m + G_m}, \tag{D.7}$$

then plugging Eq. (D.7) into Eq. (D.6). We obtain

$$\chi_m^{(min)} = \sqrt[4]{\frac{8R_{1,m}}{R_{d,m}}}. \tag{D.8}$$

The condition for a local minimum is given by $\frac{\partial^2 \mathcal{F}}{\partial B_m^2} > 0$ and $\det = \frac{\partial^2 \mathcal{F}}{\partial B_m^2}\frac{\partial^2 \mathcal{F}}{\partial \chi_m^2} - (\frac{\partial^2 \mathcal{F}}{\partial B_m \partial \chi_m})^2 > 0$. Looking at

$$\frac{\partial^2 \mathcal{F}}{\partial B_m^2} = \frac{16R_{1,m}^2(\chi_m^2 - 1) + 2C_m\chi_m^2(4R_{1,m}^2 - \chi_m^2)}{(\chi_m^2 - 1)(4R_{1,m}^2 - \chi_m^2)}, \tag{D.9}$$

the first condition is always true since $\chi_m = \cosh^2 g_m + \sinh^2 g_m > 1$ and $R'_{1,m} = \frac{R_{1,m}}{\chi_m} > \frac{1}{2}$. The second condition on the determinant det is harder to prove, but at the critical point the determinant is equivalent to

$$\det = \frac{128\chi_m^4 R_{1,m}^2 A_m^2 C_m^2}{(\chi_m^2 - 1)(4R_{1,m}^2 - \chi_m^2)(H_m + G_m)} > 0. \tag{D.10}$$

Therefore the critical point given by Eqs. (D.7) and (D.8) is indeed a minimum, assuming that it describes a valid transformation.

D.2 The symmetric case and $R'_2 = \frac{1}{2}$

Again, when $\chi_m = 1$ we have $R'_{1,m} = R_{1,m}$ and $\partial_\mu R'_{1,m} = \partial_\mu R_{1,m}$ from Eqs. (5.71). This case gives the upper bound $\mathcal{J} > j(\rho_1)$ which is uninteresting. For $\chi_m = R_{1,m} + \frac{1}{2}$, the only finite upper bound is again when $\partial_\mu R'_{1,m} = 0$:

$$\mathcal{J}_{\mu\nu} = \sum_m R_{d,m} R_{1,m} (R_{1,m} + \frac{1}{2}) \frac{\partial_\mu \ln(R_{1,m} + \frac{1}{2}) \partial_\nu \ln(R_{1,m} + \frac{1}{2})}{R_{1,m} - \frac{1}{2}}. \quad \text{(D.11)}$$

This is again a less tighter upper bound than $\mathcal{J}^{(2)}$.

We proceed to the other cases. Define

$$G_m = R_{d,m}\big[2R_{1,m} + 1 - 2\chi_m\big]\big[2(R_{1,m}+1)\chi_m - (2R_{1,m}+1)\big], \quad \text{(D.12)}$$

$$H_m = 2(\chi_m - 1)(2R_{1,m} + 1). \quad \text{(D.13)}$$

The equations determining the critical points are given by

$$\frac{G_m B_m - H_m(A_m - B_m)}{(\chi_m - 1)(2R_{1,m} + 1 - 2\chi_m)} = 0 \quad \text{(D.14)}$$

$$-\frac{R_{d,m} B_m^2}{2(\chi_m - 1)^2} + \frac{2(2R_{1,m}+1)(A_m - B_m)^2}{(2R_{1,m} + 1 - 2\chi_m)^2} = 0. \quad \text{(D.15)}$$

Assuming that $\chi_m \neq 1$ and $\chi_m \neq R_{1,m} + \frac{1}{2}$, from Eq. (D.14) we can obtain B_m in terms of χ_m:

$$B_m = \frac{H_m A_m}{H_m + G_m}. \quad \text{(D.16)}$$

Plugging Eq. (D.16) into Eq. (D.15) and realizing that $\chi_m \geq 1$ for a valid transformation, we obtain

$$\chi_m^{(min)} = \frac{1}{2(R_{1,m}+1)} \left(2R_{1,m} + 1 + \sqrt{\frac{2R_{1,m}+1}{R_{d,m}}} \right). \quad \text{(D.17)}$$

We conclude that this critical point is a minimum point as

$$\frac{\partial^2 \mathcal{F}}{\partial B_m^2} = R_{d,m}\left(2R_{1,m}+1+\frac{\chi_m}{\chi_m-1}\right)+\frac{2(2R_{1,m}+1)}{2R_{1,m}+1-2\chi_m} > 0, \quad \text{(D.18)}$$

and the determinant $\det = \frac{\partial^2 \mathcal{F}}{\partial B_m^2}\frac{\partial^2 \mathcal{F}}{\partial \chi_m^2} - (\frac{\partial^2 \mathcal{F}}{\partial B_m \partial \chi_m})^2$ at the critical point is given by

$$\det = \frac{8R_{d,m}^2 A_m^2 (2R_{1,m}+1)^2(R_{1,m}+1)}{(\chi_m-1)(2R_{1,m}+1-2\chi_m)}\times \quad \text{(D.19)}$$

$$\frac{2R_{1,m}+1-2\chi_m+2\big[\chi_m-1\big]\big[2(R_{1,m}+1)\chi_m-(2R_{1,m}+1)\big]}{\Big(H_m+G_m\Big)^2}, \quad \text{(D.20)}$$

which is also positive.

D.3 The antisymmetric case

Similar to the other cases, when $\chi_m = 1$ the upper bound is given by $j(\rho_1)$. Since $0 \leq \chi_m \leq 1$, $2R_{1,m} - 1 + 2\chi_m \geq 2R_{1,m} - 1$. By the requirement $R_{1,m} > 1$, we have $2R_{1,m} - 1 + 2\chi_m \geq 1 \neq 0$. To proceed to the other cases, define

$$G_m = R_{d,m}\big[2R_{1,m}-1+2\chi_m\big]\big[2(R_{1,m}-1)\chi_m-(2R_{1,m}-1)\big], \quad \text{(D.21)}$$

$$H_m = 2(\chi_m-1)(2R_{1,m}-1). \quad \text{(D.22)}$$

The critical points are given by the solutions of:

$$\frac{G_m B_m - H_m(A_m-B_m)}{(\chi_m-1)(2R_{1,m}-1+2\chi_m)} = 0, \quad \text{(D.23)}$$

$$\frac{R_{d,m}B_m^2}{2(\chi_m-1)^2} - \frac{2(2R_{1,m}-1)(A_m-B_m)^2}{(2R_{1,m}-1+2\chi_m)^2} = 0, \quad \text{(D.24)}$$

Assuming that $\chi_m \neq 1$, B_m is then given by

$$B_m = \frac{H_m A_m}{H_m+G_m}. \quad \text{(D.25)}$$

Substituting this into Eq. (D.24), we have

$$\chi_m^{(min)} = \frac{1}{2(R_{1,m}-1)}\left(2R_{1,m}-1-\sqrt{\frac{2R_{1,m}-1}{R_{d,m}}}\right). \tag{D.26}$$

We first note that

$$(2R_{1,m}-1)(\chi_m^{(min)}-1)-\chi_m^{(min)} = -\frac{2R_{1,m}}{R_{d,m}} \leq 0, \tag{D.27}$$

since $2R_{1,m}-1+2\chi_m \geq 0$ for all valid χ_m, the second derivative

$$\frac{\partial^2 \mathcal{F}}{\partial B_m^2} = \frac{R_{d,m}[(2R_{1,m}-1)(\chi_m-1)-\chi_m]}{\chi_m-1} + \frac{2(2R_{1,m}-1)}{2R_{1,m}-1+2\chi_m} > 0 \tag{D.28}$$

at the critical point. The determinant $\det = \frac{\partial^2 \mathcal{F}}{\partial \chi_m^2}\frac{\partial^2 \mathcal{F}}{\partial B_m^2} - (\frac{\partial^2 \mathcal{F}}{\partial \chi_m \partial B_m})^2$ is given by

$$\det = -\frac{8R_{d,m}^2 A_m^2 (2R_{1,m}-1)^2(R_{1,m}-1)}{(\chi_m-1)(2R_{1,m}-1+2\chi_m)} \times \tag{D.29}$$

$$\frac{2R_{1,m}-1+2\chi_m+2(\chi_m-1)[(2R_{1,m}-1)(\chi_m-1)-\chi_m]}{(H_m+G_m)^2}, \tag{D.30}$$

and $\det/(R_{1,m}-1)$ is positive at the critical point $(\chi_m^{(min)}, B_m^{(min)})$. Thus, this critical point is a minimum point assuming that $R_{1,m} > 1$.