DOCTORAL DISSERTATION

博士論文

Quantum estimation theory for continuous data

（連続的データの量子推定理論）

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2019

令和元年12月博士（理学）申請

Department of Physics, Graduate School of Science,
The University of Tokyo

東京大学大学院理学系研究科物理学専攻

Naoto Kura

久良　尚任

Distinguishing the signal from the noise requires both scientific knowledge and self-knowledge: the serenity to accept the things we cannot predict, the courage to predict the things we can, and the wisdom to know the difference. — Nate Silver

Abstract

Quantum metrology, a field of study aimed at efficient measurements on quantum systems, has led to the distinctive separation between classical and quantum measurement resources: While the measurement error is bounded from below by the classical statistics, this bound can be surpassed by utilizing the quantum nature of the system and the measurement device. The classical limit is called the *standard quantum limit* (SQL), and is $O(N^{-1/2})$ where N generically represents the amount of resource. With quantum resource, the error can be lower than the SQL, but not below the *Heisenberg limit* of $O(N^{-1})$.

Continuous data, on the other hand, appears in many situations. In particular, time series and graphical images continuously vary over time and space, respectively, where novel, interesting phenomena may be found through measurements. However, quantum metrology is far from well established for the measurement on continuous data, which involves infinitely many parameters unlike the conventional cases. In this thesis, we explore the theory of quantum estimation on continuous data. We especially study when and how quantum metrology still benefits for the estimation on systems with unknown functions.

We begin by a preliminary study on the multiparameter estimation, which develops the author's master thesis. We show that the Heisenberg limit on multiparameter estimation can equivalently be achieved by two distinct schemes: namely, the sequential scheme and the parallelized scheme. In other words, we may measure a single system over many times, or we may measure an extensive number of systems in parallel; in either way, the total resource required is the same up to a constant factor independent of the complexity of the system.

We next develop a fundamental study on function estimation by quantum metrology, which constructs a framework for continuous data handling in the quantum world. We first set up the function estimation problem as the metrology on the position-dependent phase-shift gate. Then, we restrict the target function to a certain class of continuity, which is based on the Hölder continuity in the function analysis. We consider two measurement strategies: one localized in position and the other localized in wavenumber. We show that the quantum limits with these strategies are shown to be equivalent to each other, which suggests the quantum counterpart of Nyquist's sampling theorem. Furthermore, by using the results of the preliminary study mentioned above, we show that these quantum limits are actually optimal for any possible strategies.

We finally extend our fundamental theory to the more practical problem of edge detection. Edge detection is one of the most important techniques in continuous data processing, where edges can be located as the extrema of the wavelet trans-

formation. In order to detect an edge of a given continuous data, We propose the direct measurement of wavelet coefficients by modifying the measurement technique called *ghost imaging*. In this method, a wave packet is probed through the phase-shift gate and the momentum is measured; therefore, the measurement error is bounded by the fixed wavelet scale and the uncertainty relation. Notably, although the quantum limits for a single wavelet scale are the same as those of the multiparameter estimation, it is consistent with the function estimation when all wavelet scales are taken into consideration.

In this thesis, we connect multiparameter quantum metrology to our theory of the function estimation, and then to its application for the edge detection. These form a series of theoretical studies on how far quantum mechanics can be exploited for the better estimation of continuous data in the real world.

Publicaton List

[P1] N. Kura and M. Ueda. “Finite-error metrological bounds on multiparameter Hamiltonian estimation.” *Phys. Rev. A* **97**, 012101 (2018).

[P2] N. Kura and M. Ueda. *Theoretical limits and scaling of quantum edge detection*. Tech. rep. QIT2019-57, The Institute of Electronics, Information and Communication Engineers (2019).

[P3] N. Kura and M. Ueda. “Standard Quantum Limit and Heisenberg Limit in Function Estimation.” *Phys. Rev. Lett.* **124**, 010507 (2020).

[P4] Y. Murashita, N. Kura, and M. Ueda. “Transient fractality as a mechanism for emergent irreversibility in chaotic Hamiltonian dynamics.” *arXiv:1802.10483* (2018).

[P5] R. Hamazaki *et al.* “The Threefold Way in Non-Hermitian Random Matrices.” *arXiv:1904.13082* (2019).

[P6] Z. Gong *et al.* “Lieb-Robinson Bounds on Entanglement Gaps from Symmetry-Protected Topology.” *arXiv:1904.12464* (2019).

[P7] 久良 尚任. “量子計測と量子的な機械学習.” In **物理学者、機械学習を使う**. Ed. by 橋本 幸士. 朝倉書店. Chap. 8 (2019). ISBN: 978-4-254-13129-1.

The content of this thesis is mainly based on Refs. [P1–P3][1].

[1] The copyrights of Refs. [P1] and [P3] belongs to © 2018–2020 American Physical Society.

Contents

Acronyms

CR Cramér–Rao

EPR state Einstein–Podolsky–Rosen state

GHZ state Greenberger–Horne–Zeilinger state

LLS local linear smoothing

MES maximally entangled state

POVM positive-operator-valued measure

QFI quantum Fisher inforamtion

QST quantum state tomography

RMS root mean square

SQL standard quantum limit

Chapter 1

Introduction

1.1 Quantum metrology

Since the discovery of quantum mechanics, its counterintuitive nature has attracted scientists across the world, many of whom have sought for its impact on the practical application. Quantum information processing is considered a prospective application, since it allows various operations that are forbidden in classical statistics. There are three different, but mutually related, directions of quantum information processing: computation [1–3], cryptography [4–6], and metrology [7, 8]. In this thesis, we focus on quantum metrology, a field of study for measuring physical quantities with accuracy beyond the known classical limits.

The process of measurement is crucial to the understanding of the world we live in. It is not a trivial process in which we see the nature as it is, but a complex operation in which we obtain a mixture of desired signals and unwanted noises. The process of measurement can be described in the language of statistics, which tells us how and how far the measurement devices and algorithms become efficient. In particular, the central limit theorem has shown that the asymptotic error of ordinary parameter estimation approaches to the Gaussian noise with $\sim N^{-1/2}$ standard deviation for N trials of measurements. In quantum metrology, however, one can exploit quantum dynamics to attain $\sim N^{-1}$ errors for N measurements. The "classical" error scaling $\sim N^{-1/2}$ is called the *standard quantum limit* (SQL), also known as the shot-noise limit, and the "quantum" error scaling $\sim N^{-1}$ is referred to as the *Heisenberg limit*.

Quantum system	limited by SQL	limited by HL	Ref.
Qubits	separate $\|+\rangle$ states	generalized GHZ state	[7]
Cold atoms	Nondegenerate ensemble	Bose-Einstein condensate	[9]
Optical mode	single-photon state	N00N state	[10]
	coherent beam	squeezed beam	[11]

Table 1.1: Brief examples of quantum systems that provide different error bounds according to the form of probe states. The amount of resource can be measured by the average number of qubits, particles or photons.

The Heisenberg limit can be reached when coherent dynamics are to be estimated with quantum-correlated probe states. A minimal example of Heisenberg-limited measurement is estimation of a one-qubit gate $e^{i\theta\sigma_z}$, where the entangled state $\frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$ can be chosen as the probe state [7]. We briefly show in Table 1.1 various examples of quantum systems where Heisenberg limits can be reached. A similar argument leads to the benefit of quantum correlation for a general system with one unknown parameter. This covers common simple models including atomic clocks [12–14] and interferometry [11, 15–17], by which we can measure such physical properties as an optical path length and an excitation energy. We note that quantum correlations in the probe do not always appear as entanglement, as typified by squeezed beams in an optical system [11, 18].

However, physical systems in reality are often more complicated than a mere one-paraemter system. In particular, the system may contain unwanted noises and/or more than one parameter. For example, under certain noise models, the accuracy is known to decline from the Heisenberg limit to the SQL [19] or some intermediate limits [20–23]. Although the Heisenberg limit is still present in multiple-parameter estimation [24], there is another problem that the error bounds scales not only by the amount of resource but also by the size of the model. It is hence important to know how and how far quantum mechanics can be exploited in such complicated systems.

The error bounds on multiparameter estimation problems have been calculated in the cases of multiple-phase estimation [25] and unitary channel estimation [26], but there has not been a theory that quantitatively evaluates the size dependency of these quantum limits. In the master course, the author have established a general framework for multiparameter quantum metrology as the Hamiltonian estimation [27]. Multiparameter estimation problems are modeled as the estimation of a Hamiltonian characterized by a set of unknown parameters, which we regard as a parameter vector. The error bounds on such Hamiltonian models have been derived by using the information-geometrical analysis [28, 29]. For the case of spherical Hamiltonian models[1], that include both multiple-phase estimation and unitary channel estimation as subclasses, we have formulated the size dependence of the error bounds, which depends only on the dimensions of the Hilbert space and the parameter vector. As a result, the benefit of the Heisenberg limit is still present in the estimation of large-sized models, but a stronger correlation relative to the model size is required to attain the Heisenberg limit.

The methodology for attaining quantum limits has also been discussed. In particular, it is worthwhile to ask whether quantum metrology can be parallelized: If it is allowed to distribute the quantum resource over multiple systems, the total time of estimation may be reduced. An existing study [30] suggested that parallelization is costly in the estimation of larger Hamiltonians. The analysis in

[1]We formally define a Hamiltonian model H_θ with the parameter vector θ to be *spherical* if $\int (H_\theta)^2 \mu(d\theta) \propto I$, where μ is the Haar measure over a hypersphere. Intuitively, the sphericality indicates the uniformity of energy variance over all probe states.

the author's master thesis [31] remained inconclusive, since the chaotic effect of the long-time evolution was not successfully handled.

1.2 Continuous data analysis

Estimation of a parameter, may it be a scalar or a vector, relies on the principle of uniformity: The same setup of a system and the same measurement should give identical results, no matter when or where we measure it. This does not hold in the true sense; essentially any signal depends on the time and position of the source, which makes any physical variable a function of spacetime rather than a parameter. For example, a time series of measurement outcomes is a one-dimensional function and images formed by spatially distributed intensity are a two-dimensional function. Such functional variations are often targets of observation rather than just negligible noises. Synchronization of clocks is a typical example [12, 32], since the reference clock may not tick regularly and thus needs continuous monitoring. Alluring phenomena such as a gravitational wave [33] or an event horizon [34, 35] are also detected as an unusual surge in time series or a hole in an two-dimensional image.

Functions are also a common target of estimation in classical statistics. Despite the usefulness of parametric model fitting, not every signal source comes up with any model in the first place. In this case, we need to estimate a probability distribution over the real line. The property of the estimation of a cumulative distribution has been revealed by Donsker's theorem [36], where the limiting behavior is described by a Gaussian process. For the estimation of a probability density, there exists a standard algorithm called kernel density estimation [37, 38].

Studies in continuous data requires the mathematics of function analysis and stochastic processes. This is due to the infinite degrees of freedom of functions: When continuous data is seen an infinite number of independent variables, finite number of measurement outcomes cannot lead to any conclusion about the data. Hence, in order to estimate a function with finite error, smoothness of the function must be quantitatively assumed. The smoothness of a function can be defined by various measures, and the choice of measure is directly related to the error bound. Informally speaking, the smoother functions one tries to estimate, the more accurate estimation one can achieve. In this sense, function estimation requires advanced treatments beyond the multiparameter estimation, though the technique of multiparameter estimation can be applied.

In recent studies, the Bayesian approaches have been employed in quantum theory of signal detection [39–43]. Bayesian analysis imposes a certain stochastic process as the prior distribution, and the target function is considered to be drawn from that process. With a stationary process, the smoothness of sample functions are characterized by the power spectrum, and error bounds can be achieved by an optimally designed smoothing algorithm. The error bounds on Bayesian signal estimation has been analyzed in recent studies, where the SQL and the Heisenberg

limit are modified according to the ultraviolet behavior of the power spectrum.

However, the assumption of stochastic behaviour of a function is not granted in all cases. In particular, when the functional behavior reflects the variation of physical properties in the system, it is more natural to consider a deterministic system rather than a stochastic one. It is also unclear whether the Bayesian approach covers unlikely events in the stochastic process, such as a sudden rise and fall of the signal. Hence, a non-Bayesian approach will be of potential use for estimating a deterministic function encoded in physical systems.

On the other hand, detection of continuous data is not the ultimate goal of our analysis. We aim to find certain functional structures that mark interesting phenomena of physical systems. Various approaches can be taken to extract meaningful features out of a function, such as signal separation [44] and model decision [45]. Among the most basic algorithms is edge detection [46], which identifies the positions at which the function is rapidly changing. Many different algorithms have developed for the feature extraction problem, some of which make use of simulated annealing [47] or neural networks [48, 49]. Hence, we expect that such continuous data analysis can also be enhanced by quantum metrology, regarding its usefulness in continuous data detection.

A straightforward application of quantum metrology for continuous data detection is simple: If we measure the system with sufficient quantum resource to obtain the well approximated function, upon which the classical algorithm can be performed. However, since the process of feature extraction usually discards a large part of the continuous data, the measurement on the entire function is possibly redundant. In quantum metrology, it is more desirable to measure only pieces of information that are significant in further analysis regarding, among othre things, the high difficulty of the Heisenberg-limited measurement.

In respect of edge detection, the theory of multiscale analysis provides a mathematical framework to efficiently extract information out of continuous data. In fact, we can locate the edges by a method on the basis of the wavelet transformation of a function [50, 51], which contains information of a given length scale from the continuous data. In other words, the function can be divided into multiple components, each of which corresponds to the edges of its own lengthscale.

1.3 Structure of this thesis

In this thesis, we establish a series of theories on quantum estimation for continuous data, and show to what extent the quantum benefit can be exploited. We present the reviews in Chapters 2 and 3, the main contents in Chapters 4–6, and the conclusion in Chapter 7.

Chapter 2 is dedicated for a brief review of the conventional theories of quantum metrology in parameter estimation. In Sec. 2.1, we introduce classical and quantum information theories based on the Fisher information. In Sec. 2.2, we reproduce the derivation of error bounds on quantum metrology, where several ex-

amples of Heisenberg-limited quantum measurements are explained. Section 2.3 is on the multiparameter extension of quantum metrology and the theoretical bounds on error derived from the Fisher-information analysis. This section covers the author's master thesis, and further results will be developed in Chapter 4.

Chapter 3 introduces the basic concepts of function analysis and signal estimation, on which our analytical results in Chapters 5 and 6 rely. In Sec. 3.1, the regularity of function is defined by various types of norms including the Hölder norm and its square-integrability version, which comes into play in this thesis. In Sec. 3.2, we review the signal estimation in terms of Bayesian analysis, where we see that the regularity of the sample function is governed by the power spectrum.

Chapter 4 is on a new results on the multiparameter Hamiltonian estimation, in which we compare the efficiency of the sequential and parallel schemes concerning the multiparameter quantum metrology. In particular, we discuss how far we can circumvent chaotic behavior in the adaptive procedure by operator-algebraic calculation. We are led to the conclusion that the cost of parallelization does not depend on the size of the Hamiltonian model, contrary to the implication by a previous study [30].

Chapter 5 develops the framework of function estimation by quantum metrology, where we set up the system as a position-dependent phase-shift gate in a one-dimensional interval. Here, our analysis is based on the non-Bayesianism in which the functions are regularized by the vector norm rather than the stochastic process. In Sec. 5.1, we describe the fundamental setup of the problem. In Sec. 5.2, we present two estimation strategies: The probe states are localized in real space in one strategy, while they are localized in the wavenumber space in the other. Despite the clear difference between these strategies, both achieve the same error bounds up to a constant factor. We find that these error bounds are actually optimal; in Sec. 5.3, we derive the theoretical error bounds on function estimation with the help of the results from the multiparameter estimation in Chapter 4. We find two quantum limits corresponding to the SQL and the Heisenberg limit, which are weaker than the conventional quantum limits due to the behavior of functions.

In Chap. 6, we address the practical application of edge detection by quantum metrology. In Sec. 6.1, we propose the direct quantum measurement on the wavelet coefficients of the function, which can be performed by a modified version of ghost imaging [52–54]. In Sec. 6.2, we approximately derive the error bounds on the edge detection for a fixed lengthscale, which results in the SQL and the Heisenberg limit of parameter estimation. In Sec. 6.3, we also present the analytical derivation on the edge detection according to the regularity of the functions, where we obtain the quantum limits consistent with those of function estimation. The main analysis is on Gaussian wavefunction withdistinguishable particles Gaussian wavefunction, while we also provide preliminary results on non-Gaussian wavefunctions and indistinguishable photons in Sec. 6.4.

Some of the intricate mathematical topics are relegated to the Appendices. Appendix A contains operator-algebraic discussion for the evaluation of operator

norms used in the main chapters. Appendix B is dedicated for the notion of Hölder continuity, which resembles the C^q-class differentiability regarding functions but supposes non-integer q.

Chapter 2

Quantum Estimation and Metrology

2.1 Estimation theory and Fisher information

2.1.1 Classical estimation theory

We begin with a parameter-estimation problem in classical statistics. The goal here is to accurately estimate parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ from some observation X, whose probability density $p_{\boldsymbol{\theta}}(X)$ depends on the parameter $\boldsymbol{\theta}$. Here, it is more convenient to consider a single parameter vector $\boldsymbol{\theta}$ ranging over a convex region $\Theta \subset \mathbb{R}^m$ rather than to consider m independent parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$. In this way, the estimation problem is characterized by the family of probability densities $\{p_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ called a *stochastic model*. Moreover, a solution to this estimation problem can be represented by an *estimator* $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*(X)$, which is the mapping from the observation X to the estimate of the parameter vector $\boldsymbol{\theta}^* \in \Theta$.

Given the true parameter vector $\boldsymbol{\theta}$, the stochastic behavior of the estimator $\boldsymbol{\theta}^*$ can be described by its mean value $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and the covariance matrix $V_{\boldsymbol{\theta}}$:

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = \langle \boldsymbol{\theta}^* \rangle_{\boldsymbol{\theta}}, \qquad V_{\boldsymbol{\theta}} = \langle (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}})(\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}})^{\mathrm{T}} \rangle, \tag{2.1}$$

where $\langle A \rangle_{\boldsymbol{\theta}} = \int A(X) p_{\boldsymbol{\theta}}(X) dX$ denotes the expected value of A with respect to the distribution $p_{\boldsymbol{\theta}}$. The overall error δ can be defined by the *root mean square* (RMS) of the Euclidean norm $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|$:

$$\delta^2 = \langle \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \rangle_{\boldsymbol{\theta}}. \tag{2.2}$$

In fact, this statistics can be computed in terms of the mean $\mu_{\boldsymbol{\theta}}$ and the covariance $V_{\boldsymbol{\theta}}$ as

$$\begin{aligned} \delta^2 &= \langle \|(\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}})\|^2 \rangle_{\boldsymbol{\theta}} \\ &= \|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 + 2\langle (\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}) \cdot (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}}) \rangle_{\boldsymbol{\theta}} + \langle \|\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}}\|^2 \rangle_{\boldsymbol{\theta}} \\ &= \|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 + \operatorname{tr} V_{\boldsymbol{\theta}}. \end{aligned} \tag{2.3}$$

Therefore, the estimator is accurate when $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is close to $\boldsymbol{\theta}$ and the positive-semidefinite matrix $V_{\boldsymbol{\theta}}$ is close to zero.

If n independent and identically distributed (i.i.d.) observations $X_1, \ldots, X_n$ are available, an estimator is given by the average of the estimators for all observations

$1 \leq i \leq n$:

$$\boldsymbol{\theta}_n^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\theta}^*(X_i). \tag{2.4}$$

This estimator distributes around the mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and the covariance matrix $\frac{1}{n}V_{\boldsymbol{\theta}}$. In this case, the RMS error δ can be described as

$$\delta^2 = \|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 + \frac{1}{n}\operatorname{tr} V_{\boldsymbol{\theta}}. \tag{2.5}$$

We note that when n tends to infinity, the first term in Eq. (2.5) remains nonzero unless $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \boldsymbol{\theta}$ while the second term $\frac{1}{n}\operatorname{tr} V_{\boldsymbol{\theta}}$ vanishes. In this sense, the term $\|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\|$ is referred to as the *bias* and the term $\operatorname{tr} V_\theta$ is called the *variance*.

We say that the estimator $\boldsymbol{\theta}^*$ is *unbiased* if $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \boldsymbol{\theta}$ holds for all $\boldsymbol{\theta} \in \Theta$. An unbiased estimator is desirable in the sense that the RMS error δ in Eq. (2.5) tends to zero in the limit of $n \to \infty$.

In general, an observation X may occur from any parameter $\boldsymbol{\theta}$ with positive probability $p_{\boldsymbol{\theta}}(X) > 0$, implying the impossibility of zero-error estimation. In fact, when the estimator $\boldsymbol{\theta}^*$ is unbiased, the variance $\operatorname{tr} V_{\boldsymbol{\theta}}$ has a positive lower bound, no matter how we choose the estimator $\boldsymbol{\theta}^*$. This can be shown from a matrix inequality between the covariance matrix $V_{\boldsymbol{\theta}}$ and the inverse of the Fisher information matrix $J_{\boldsymbol{\theta}}$, the definition of which follows.

Definition 2.1 We assume that a stochastic model $\{p_{\boldsymbol{\theta}}\}$ is differentiable with respect to the parameter $\boldsymbol{\theta}$. More precisely, we assume the existence of the *logarithmic derivative* $\boldsymbol{L}p_{\boldsymbol{\theta}} = (L_1 p_{\boldsymbol{\theta}}, \ldots, L_m p_{\boldsymbol{\theta}})$ as a vector of measurable functions:

$$L_\alpha p_{\boldsymbol{\theta}}(X) = \frac{\partial}{\partial\theta_\alpha}\log p_{\boldsymbol{\theta}}(X) = \frac{1}{p_{\boldsymbol{\theta}}(X)}\frac{\partial p_{\boldsymbol{\theta}}(X)}{\partial\theta_\alpha}. \tag{2.6}$$

Then, the *Fisher information* matrix $J_{\boldsymbol{\theta}}$ of this model is defined as

$$\begin{aligned}[J_{\boldsymbol{\theta}}]_{\alpha\beta} &= \langle L_\alpha p_{\boldsymbol{\theta}} L_\beta p_{\boldsymbol{\theta}}\rangle_{\boldsymbol{\theta}} \\ &= \int L_\alpha p_{\boldsymbol{\theta}}(X) L_\beta p_{\boldsymbol{\theta}}(X) p_{\boldsymbol{\theta}}(X) dX \\ &= \int \frac{\partial p_{\boldsymbol{\theta}}(X)}{\partial\theta_\alpha}\frac{\partial p_{\boldsymbol{\theta}}(X)}{\partial\theta_\beta}\frac{dX}{p_{\boldsymbol{\theta}}(X)}.\end{aligned} \tag{2.7}$$

Theorem 2.2 (Cramér–Rao inequality [55]) *Suppose that the Fisher information matrix $J_{\boldsymbol{\theta}}$ in Def. 2.1 is positive semidefinite. For any unbiased estimator $\boldsymbol{\theta}^*$ with covariance matrix $V_{\boldsymbol{\theta}}$, we have the matrix inequality $V_{\boldsymbol{\theta}} \geq J_{\boldsymbol{\theta}}^{-1}$.*

Regarding Eq. (2.3), the RMS error δ of an unbiased estimator is bounded from below by the Fisher information as:

$$\delta \geq (\operatorname{tr} V_{\boldsymbol{\theta}})^{1/2} \geq (\operatorname{tr} J_{\boldsymbol{\theta}}^{-1})^{1/2}. \tag{2.8}$$

For the later convenience, we show a more generalized form of the Cramér–Rao (CR) inequality which can also be used for biased estimators. We define a matrix

$D_{\boldsymbol{\theta}}$ by $D_{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}$, which equals the Jacobian of the map $\boldsymbol{\theta} \mapsto \boldsymbol{\mu}_{\boldsymbol{\theta}}$. We note that when $\boldsymbol{\theta}^*$ is unbiased, the matrix $D_{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}$ is equal to the identity matrix I.

Proposition 2.3 *We consider an estimator $\boldsymbol{\theta}^*$ with mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and covariance matrix $V_{\boldsymbol{\theta}}$. Then, we have the matrix inequality $V_{\boldsymbol{\theta}} \geq D_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} D_{\boldsymbol{\theta}}^{\mathrm{T}}$, where the matrix $D_{\boldsymbol{\theta}}$ is defined as $D_{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}$.*

Proof. First, we show the following identity for an arbitrary estimator $\boldsymbol{\theta}^*$:

$$\langle \boldsymbol{L} p_{\boldsymbol{\theta}} (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}})^{\mathrm{T}} \rangle_{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}. \tag{2.9}$$

This identity can be derived as follows:

$$\begin{aligned}
\langle \boldsymbol{L} p_{\boldsymbol{\theta}} (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}})^{\mathrm{T}} \rangle_{\boldsymbol{\theta}} &= \int \frac{\nabla_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}}(X)}{p_{\boldsymbol{\theta}}(X)} [\boldsymbol{\theta}^*(X) - \boldsymbol{\mu}_{\boldsymbol{\theta}}]^{\mathrm{T}} p_{\boldsymbol{\theta}}(X) dX \\
&= \int \nabla_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}}(X) [\boldsymbol{\theta}^*(X) - \boldsymbol{\mu}_{\boldsymbol{\theta}}]^{\mathrm{T}} dX \\
&= \nabla_{\boldsymbol{\theta}} \int p_{\boldsymbol{\theta}}(X) [\boldsymbol{\theta}^*(X)]^{\mathrm{T}} dX - \left[\nabla_{\boldsymbol{\theta}} \int p_{\boldsymbol{\theta}}(X) dX \right] \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} \\
&= \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} - (\nabla_{\boldsymbol{\theta}} 1) \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}.
\end{aligned} \tag{2.10}$$

Now, let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ be arbitrary nonzero vectors. By the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}
(\boldsymbol{x}^{\mathrm{T}} V_{\boldsymbol{\theta}} \boldsymbol{x})(\boldsymbol{y}^{\mathrm{T}} J_{\boldsymbol{\theta}} \boldsymbol{y}) &= \langle \boldsymbol{x}^{\mathrm{T}} (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}}) (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}})^{\mathrm{T}} \boldsymbol{x} \rangle_{\boldsymbol{\theta}} \cdot \langle \boldsymbol{y}^{\mathrm{T}} (\boldsymbol{L} p_{\boldsymbol{\theta}}) (\boldsymbol{L} p_{\boldsymbol{\theta}})^{\mathrm{T}} \boldsymbol{y} \rangle_{\boldsymbol{\theta}} \\
&\geq \langle \boldsymbol{x}^{\mathrm{T}} (\boldsymbol{\theta}^* - \boldsymbol{\mu}_{\boldsymbol{\theta}}) (\boldsymbol{L} p_{\boldsymbol{\theta}})^{\mathrm{T}} \boldsymbol{y} \rangle_{\boldsymbol{\theta}}^2 = (\boldsymbol{x}^{\mathrm{T}} D_{\boldsymbol{\theta}} \boldsymbol{y})^2.
\end{aligned} \tag{2.11}$$

Substituting $\boldsymbol{y} = J_{\boldsymbol{\theta}}^{-1} D_{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{x}$ in Eq. (2.11), we finally obtain

$$\boldsymbol{x}^{\mathrm{T}} V_{\boldsymbol{\theta}} \boldsymbol{x} \geq \boldsymbol{x}^{\mathrm{T}} D_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} D_{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{x}. \tag{2.12}$$

The desired inequality follows since $\boldsymbol{x} \neq \boldsymbol{0}$ is arbitrary. □

The CR inequality $V_{\boldsymbol{\theta}} \geq J_{\boldsymbol{\theta}}^{-1}$ does not necessarily hold when the estimator is not unbiased. In fact, $V_{\boldsymbol{\theta}} = 0$ can be achieved by setting $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$. Such an estimator can be described as "extremely biased," since the estimator blindly assumes $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ regardless of the outcome.

Compared to the unbiased case, we may reduce the variance when the matrix $D_{\boldsymbol{\theta}}$ is smaller than the identity. However, keeping the bias $\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}$ within a small value requires the matrix $\nabla_{\boldsymbol{\theta}} D_{\boldsymbol{\theta}}$ be close to the identity, and vice versa. This leads to the tradeoff relation between the bias and the variance, which we discuss in Chapter 5.

2.1.2 Quantum estimation theory

The efficiency of quantum information processing may be enhanced by the use of appropriate quantum states instead of stochastic variables, owing to the presence of non-diagonal entries in quantum density matrices. Nonetheless, the problem of estimating quantum states can be treated in parallel to the classical estimation problem.

A *quantum state model* is a family of density matrices $\{\rho_{\boldsymbol{\theta}}\}$ parametrized by a parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^n$. Quantum states, however, are not equipped with a stochastic variable X; instead, we need to obtain a stochastic outcome by performing some quantum measurement of our choice.

In a general formulation, a quantum measurement is represented by a *positive-operator-valued measure* (POVM) $\mathcal{M}$ [56]. Let us denote by $\mathcal{X}$ the set of all possible measurement outcomes, and the POVM maps every event $E \subset \mathcal{X}$ to a nonnegative operator $\mathcal{M}(E)$. When a density operator ρ is to be measured, the measurement outcome falls in E with probability $\mathrm{tr}[\rho\mathcal{M}(E)]$. In particular, $\mathcal{M}(\mathcal{X}) = I$ is required since $\mathrm{tr}[\rho\mathcal{M}(\mathcal{X})] = 1$ must hold regardless of the state ρ.

Therefore, if we measure the quantum state $\rho_{\boldsymbol{\theta}}$ with a POVM $\mathcal{M}$, the probability density $p_{\boldsymbol{\theta}}^{\mathcal{M}}(X)$ of the measurement outcome X can be defined as

$$p_{\boldsymbol{\theta}}^{\mathcal{M}}(X)dX = \mathrm{tr}[\rho_{\boldsymbol{\theta}}\mathcal{M}(dX)], \tag{2.13}$$

where dX refers to an infinitesimal volume around X. If we denote by $J_{\boldsymbol{\theta}}^{\mathcal{M}}$ the Fisher information concerning the stochastic model $\{p_{\boldsymbol{\theta}}^{\mathcal{M}}\}$, any unbiased estimator $\boldsymbol{\theta}^*$ computed from this measurement outcome is bounded by $V_{\boldsymbol{\theta}} \geq (J_{\boldsymbol{\theta}}^{\mathcal{M}})^{-1}$.

Now, the question is whether the covariance matrix has a lower bound that is independent of the choice of the quantum measurement. The answer is yes, and there exists a matrix $J_{\boldsymbol{\theta}}$ depending only on the quantum state model $\{\rho_{\boldsymbol{\theta}}\}$, which satisfies $J_{\boldsymbol{\theta}}^{\mathcal{M}} \leq J_{\boldsymbol{\theta}}$ for all POVM $\mathcal{M}$. This matrix $J_{\boldsymbol{\theta}}$ is called a *quantum Fisher inforamtion* (QFI) matrix of the model $\{\rho_{\boldsymbol{\theta}}\}$.[1] The QFI has infinitely many variants due to the non-commutativity of operators [57], some of which are listed in Ref. [58]. In this thesis, we consider quantum dynamics without dissipative noises, in which case an analysis on pure states is sufficient. A pure state can be represented by a unit vector $|q\rangle$ in a Hilbert space $\mathcal{H}$ instead of the density matrix $\rho = |q\rangle\langle q|$.

Notably, the QFI matrix on a pure quantum state model $\{|q_{\boldsymbol{\theta}}\rangle\}$ can be uniquely determined [59, 60]:

$$[J_{\boldsymbol{\theta}}]_{\alpha\beta} = 4\,\mathrm{Re}\left\langle \frac{\partial q_{\boldsymbol{\theta}}}{\partial\theta_\alpha}\middle|(I - |q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}}|)\middle|\frac{\partial q_{\boldsymbol{\theta}}}{\partial\theta_\beta}\right\rangle. \tag{2.14}$$

Theorem 2.4 (Quantum Cramér–Rao inequality) *Let $\{|q_{\boldsymbol{\theta}}\rangle\}$ be a quantum state model consisting only of pure states in a Hilbert space $\mathcal{H}$, and $J_{\boldsymbol{\theta}}$ be the QFI*

[1] There exists additional properties required for QFI matrices: (i) monotonicity under the transformation by quantum channels and (ii) consistency with the classical Fisher information for diagonal quantum states.

matrix of this model defined by Eq. (2.14)*. Then, under any measurement* $\mathcal{M}$ *on this model, the Fisher information matrix* $J_{\boldsymbol{\theta}}^{\mathcal{M}}$ *of the measurement outcome satisfies* $J_{\boldsymbol{\theta}}^{\mathcal{M}} \leq J_{\boldsymbol{\theta}}$*.*

Proof. The goal is to show

$$\boldsymbol{x}^{\mathrm{T}} J_{\boldsymbol{\theta}}^{\mathcal{M}} \boldsymbol{x} \leq \boldsymbol{x}^{\mathrm{T}} J_{\boldsymbol{\theta}} \boldsymbol{x} \tag{2.15}$$

for an arbitrary nonzero vector $\boldsymbol{x} \in \mathbb{R}^m$. Let $D_{\boldsymbol{x}} = \sum_{\alpha=1}^{m} x_\alpha \frac{\partial}{\partial \theta_\alpha}$ denote the directional differentiation along $\boldsymbol{x}$. Then, Eq. (2.15) is equal to

$$\int \frac{[D_{\boldsymbol{x}} p_{\boldsymbol{\theta}}^{\mathcal{M}}(X)]^2}{p_{\boldsymbol{\theta}}^{\mathcal{M}}(X)} dX \leq 4\langle D_{\boldsymbol{x}} q_{\boldsymbol{\theta}} | (I - |q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}}|) | D_{\boldsymbol{x}} q_{\boldsymbol{\theta}} \rangle, \tag{2.16}$$

where $p_{\boldsymbol{\theta}}^{\mathcal{M}}(X)$ is the probability density of the measurement outcome.

We take the directional derivative $D_{\boldsymbol{x}}$ in Eq. (2.13) to obtain

$$\begin{aligned} D_{\boldsymbol{x}} p_{\boldsymbol{\theta}}^{\mathcal{M}}(X) dX &= D_{\boldsymbol{x}} (\langle q_{\boldsymbol{\theta}} | \mathcal{M}(dX) | q_{\boldsymbol{\theta}} \rangle) \\ &= 2\,\mathrm{Re} \langle q_{\boldsymbol{\theta}} | \mathcal{M}(dX) | D_{\boldsymbol{x}} q_{\boldsymbol{\theta}} \rangle. \end{aligned} \tag{2.17}$$

In fact, we can replace $|D_{\boldsymbol{x}} q_{\boldsymbol{\theta}}\rangle$ with $|v_{\boldsymbol{\theta}}\rangle = (I - |q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}}|)|D_{\boldsymbol{x}} q_{\boldsymbol{\theta}}\rangle$:

$$\begin{aligned} D_{\boldsymbol{x}} p_{\boldsymbol{\theta}}^{\mathcal{M}}(X) dX &= 2\,\mathrm{Re}\big[\langle q_{\boldsymbol{\theta}} | \mathcal{M}(dX) | D_{\boldsymbol{x}} q_{\boldsymbol{\theta}} \rangle - \langle q_{\boldsymbol{\theta}} | \mathcal{M}(dX) | q_{\boldsymbol{\theta}} \rangle \langle q_{\boldsymbol{\theta}} | D_{\boldsymbol{x}} q_{\boldsymbol{\theta}} \rangle\big] \\ &= 2\,\mathrm{Re} \langle q_{\boldsymbol{\theta}} | \mathcal{M}(dX) | v_{\boldsymbol{\theta}} \rangle, \end{aligned} \tag{2.18}$$

because $\langle q_{\boldsymbol{\theta}} | q_{\boldsymbol{\theta}} \rangle = 1$ and thus $\langle q_{\boldsymbol{\theta}} | D_{\boldsymbol{x}} q_{\boldsymbol{\theta}} \rangle$ is pure imaginary.

Furthermore, by the Cauchy-Schwartz inequality, we have

$$\begin{aligned} [D_{\boldsymbol{x}} p_{\boldsymbol{\theta}}^{\mathcal{M}}(X) dX]^2 &\leq 4 \big| \langle q_{\boldsymbol{\theta}} | \mathcal{M}(dX) | v_{\boldsymbol{\theta}} \rangle \big|^2 \\ &\leq 4 \langle v_{\boldsymbol{\theta}} | \mathcal{M}(dX) | v_{\boldsymbol{\theta}} \rangle \cdot \langle q_{\boldsymbol{\theta}} | \mathcal{M}(dX) | q_{\boldsymbol{\theta}} \rangle \\ &= 4 \langle v_{\boldsymbol{\theta}} | \mathcal{M}(dX) | v_{\boldsymbol{\theta}} \rangle p_{\boldsymbol{\theta}}^{\mathcal{M}}(X) dX. \end{aligned} \tag{2.19}$$

Since the POVM $\mathcal{M}(dX)$ yields I when integrated over all outcomes, we obtain

$$\int \frac{[D_{\boldsymbol{x}} p_{\boldsymbol{\theta}}^{\mathcal{M}}(X)]^2}{p_{\boldsymbol{\theta}}^{\mathcal{M}}(X)} dX \leq 4 \int \langle v_{\boldsymbol{\theta}} | \mathcal{M}(dX) | v_{\boldsymbol{\theta}} \rangle = 4 \langle v_{\boldsymbol{\theta}} | v_{\boldsymbol{\theta}} \rangle, \tag{2.20}$$

which is the desired inequality (2.16). □

By Theorem 2.4, both the CR inequality (Th. 2.2) or its generalization (Prop. 2.3) can be applied to a pure-state model, with the Fisher information matrix replaced by the QFI matrix in Eq. (2.14).

2.1.3 Additivity of Fisher information

Let us consider the estimation problem when independent observations $X_1, \ldots, X_n$ are available. Then, the stochastic model $\{p_{\boldsymbol{\theta}}\}$ for the combined observation $X_\bullet = (X_1, \ldots, X_n)$ will be

$$p_{\boldsymbol{\theta}}(X_\bullet) = p_{1,\boldsymbol{\theta}}(X_1) \cdots p_{n,\boldsymbol{\theta}}(X_n). \tag{2.21}$$

Then, the Fisher information of this model can be calculated as follows:

$$\begin{aligned}
[J_{\boldsymbol{\theta}}]_{\alpha\beta} &= \int \cdots \int \frac{\partial \log p_{\boldsymbol{\theta}}(X_\bullet)}{\partial \theta_\alpha} \frac{\partial \log p_{\boldsymbol{\theta}}(X_\bullet)}{\partial \theta_\beta} p_{\boldsymbol{\theta}}(X_\bullet) dX_1 \cdots dX_n \\
&= \int \cdots \int \left(\frac{\partial}{\partial \theta_\alpha} \sum_{i=1}^{n} \log p_{i,\boldsymbol{\theta}}(X_i) \right) \left(\frac{\partial}{\partial \theta_\alpha} \sum_{j=1}^{n} \log p_{j,\boldsymbol{\theta}}(X_j) \right) p_{\boldsymbol{\theta}}(X_\bullet) dX_1 \cdots dX_n \\
&= \sum_{i=1}^{n} \int \frac{\partial \log p_{i,\boldsymbol{\theta}}(X_i)}{\partial \theta_\alpha} \frac{\partial \log p_{i,\boldsymbol{\theta}}(X_i)}{\partial \theta_\beta} p_{i,\boldsymbol{\theta}}(X_i) dX_i \\
&\quad + \sum_{i \neq j} \iint \frac{\partial \log p_{i,\boldsymbol{\theta}}(X_i)}{\partial \theta_\alpha} \frac{\partial \log p_{j,\boldsymbol{\theta}}(X_j)}{\partial \theta_\beta} p_{i,\boldsymbol{\theta}}(X_i) p_{j,\boldsymbol{\theta}}(X_j) dX_i dX_j.
\end{aligned} \tag{2.22}$$

Here, the second-to-the-last line in (2.22) corresponds to the Fisher information matrix $[J_{i,\boldsymbol{\theta}}]_{\alpha\beta}$ with respect to the ith stochastic model $\{p_{i,\boldsymbol{\theta}}\}$, whereas the summand in the last line can be rewritten as

$$\frac{\partial}{\partial \theta_\alpha} \int p_{i,\boldsymbol{\theta}}(X_i) dX_i \cdot \frac{\partial}{\partial \theta_\beta} \int p_{j,\boldsymbol{\theta}}(X_j) dX_j = \frac{\partial 1}{\partial \theta_\alpha} \frac{\partial 1}{\partial \theta_\beta} = 0. \tag{2.23}$$

In this way, the additivity of the Fisher information matrix is shown:

$$J_{\boldsymbol{\theta}} = J_{1,\boldsymbol{\theta}} + \cdots + J_{n,\boldsymbol{\theta}}. \tag{2.24}$$

We can use the same formulation even when more than two observations are available. In particular, where the n observations are i.i.d., the Fisher information matrix is

$$J_{\boldsymbol{\theta}} = n J_{1,\boldsymbol{\theta}}. \tag{2.25}$$

Let $\boldsymbol{\theta}^*$ be an arbitrary unbiased estimator from n i.i.d. observations, which may not necessarily be written be as Eq. (2.4). Then, the bound on the RMS error in Eq. (2.8) becomes

$$\delta \geq (\operatorname{tr} J_{\boldsymbol{\theta}}^{-1})^{1/2} \geq n^{-1/2} (\operatorname{tr} J_{1,\boldsymbol{\theta}}^{-1})^{1/2}, \tag{2.26}$$

indicating the universal error scaling $\delta \geq O(n^{-1/2})$. In fact, the inequality (2.26) can be asymptotically achieved in the limit of large n by the maximum-likelihood estimator [55].

The same thing can be said for quantum estimation problem. When we have n separate pure states:

$$|q_{\boldsymbol{\theta}}\rangle = |q_{1,\boldsymbol{\theta}}\rangle \otimes \cdots \otimes |q_{n,\boldsymbol{\theta}}\rangle, \tag{2.27}$$

the QFI matrix $J_{\boldsymbol{\theta}}$ for $\{q_{\boldsymbol{\theta}}\}$ is the sum of QFI matrices $J_{i,\boldsymbol{\theta}}$ of $\{q_{i,\boldsymbol{\theta}}\}$ over $1 \leq i \leq n$. In fact, $[J_{\boldsymbol{\theta}}]_{\alpha\beta}$ is 4 times the inner product of the following two vectors:

$$\begin{aligned}
(I - |q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}}|)\left|\frac{\partial q_{\boldsymbol{\theta}}}{\partial \theta_\alpha}\right\rangle &= \sum_{i=1}^{n}\left[|q_{1,\boldsymbol{\theta}}\rangle \otimes \cdots \otimes (I - |q_{i,\boldsymbol{\theta}}\rangle\langle q_{i,\boldsymbol{\theta}}|)\left|\frac{\partial q_{i,\boldsymbol{\theta}}}{\partial \theta_\alpha}\right\rangle \otimes \cdots \otimes |q_{n,\boldsymbol{\theta}}\rangle\right], \\
(I - |q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}}|)\left|\frac{\partial q_{\boldsymbol{\theta}}}{\partial \theta_\beta}\right\rangle &= \sum_{j=1}^{n}\left[|q_{1,\boldsymbol{\theta}}\rangle \otimes \cdots \otimes (I - |q_{j,\boldsymbol{\theta}}\rangle\langle q_{j,\boldsymbol{\theta}}|)\left|\frac{\partial q_{j,\boldsymbol{\theta}}}{\partial \theta_\beta}\right\rangle \otimes \cdots \otimes |q_{n,\boldsymbol{\theta}}\rangle\right].
\end{aligned} \tag{2.28}$$

We see that the inner product can be written as the summation over $1 \leq i, j \leq n$. Each of the summand with $i = j$ is equal to $[J_{i,\boldsymbol{\theta}}]_{\alpha\beta}$, while the summands with $i \neq j$ vanish owing to the projection operator $I - |q_{i,\boldsymbol{\theta}}\rangle\langle q_{i,\boldsymbol{\theta}}|$. Hence, we obtain the same results as Eq. (2.24) for the quantum case:

$$J_{\boldsymbol{\theta}} = J_{1,\boldsymbol{\theta}} + \cdots + J_{n,\boldsymbol{\theta}}. \tag{2.29}$$

Consequently, Eqs. (2.25) and (2.26) can also be used for the QFI matrix for n identical quantum states. Note, however, that the asymptotic saturation of this lower bound is not guaranteed in the quantum case, unlike the maximum-likelihood estimator in the classical estimation problem.

2.1.4 Quantum state tomography

The *quantum state tomography* (QST) is another topic in the quantum estimation [56], where target is an arbitrary quantum state in a Hilbert space $\mathcal{H} = \mathbb{C}^d$. In this thesis, we only consider estimation of a pure state $|q\rangle \in \mathcal{H}$ rather than a possibly mixed state. The estimator of the QST is represented as a state $|q^*\rangle \in \mathcal{H}$, and the quality of the estimation can evaluated by the avarage *quantum fidelity* $\mathcal{F} = \langle|\langle q|q^*\rangle|^2\rangle$.[2] The fidelity takes value in $0 \leq \mathcal{F} \leq 1$ with $\mathcal{F} = 1$ indicates the perfect estimation. Therefore, we often consider *quantum infidelity*

$$\mathcal{I} = 1 - \mathcal{F} = 1 - \langle|\langle q|q^*\rangle|^2\rangle \tag{2.30}$$

as a measure of error in the estimation.

Let us analyze the QST in terms of the parameter estimation. We assume that the target quantum state is $|0\rangle$, and consider a local coordinate in $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{2(d-1)})$ for the quantum state in proximity to $|q_{\mathbf{0}}\rangle = |0\rangle$:

$$|q_{\boldsymbol{\theta}}\rangle = \sqrt{1 - \|\boldsymbol{\theta}\|^2}|0\rangle + \sum_{k=1}^{d-1}(\theta_{2k} + i\theta_{2k+1})|k\rangle, \tag{2.31}$$

where $\{|0\rangle, \ldots, |d-1\rangle\}$ forms a basis of $\mathcal{H} = \mathbb{C}^d$. Then, the QFI matrix at the target parameter $\boldsymbol{\theta} = \mathbf{0}$ is $[J_{\mathbf{0}}]_{\alpha\beta} = 4\delta_{\alpha\beta}$. By the quantum CR inequality (Th. 2.4), we obtain a lower bound on the RMS error $\delta \geq \sqrt{\frac{d-1}{2n}}$ for n copies of the quantum

[2] We note that the average $\langle\cdot\rangle$ is taken over the stochastic behavior of the estimator.

state. Furthermore, if we describe the estimator $|q^*\rangle$ in terms of the parametric representation $|q_{\boldsymbol{\theta}^*}\rangle$ in Eq. (2.31), we see that the quantum infidelity in Eq. (2.30) is equivalent to the mean-square error:

$$\mathcal{I} = 1 - \left\langle \left| \langle q_{\boldsymbol{\theta}^*} | q_{\mathbf{0}} \rangle \right|^2 \right\rangle = \langle \|\boldsymbol{\theta}^*\|^2 \rangle = \delta^2. \tag{2.32}$$

Therefore, we obtain the quantum CR bound on the infidelity: $\mathcal{I} \geq \frac{d-1}{2n}$

On the other hand, the minimal fidelity for estimating an arbitrary quantum state can be computed by using the geometrical symmetry of the pure state space, which yields [60]:

$$\mathcal{I} \geq \frac{d-1}{n+d}. \tag{2.33}$$

which is larger than the CR bound with an additional factor 2 in the limit of $n \to \infty$. Hence, this is especially the case in which the quantum CR inequality cannot be attained, while the asymptotic scaling is the same. We note that the minimal infidelity in Eq. (2.33) is lower than the CR bound for $n \leq d$ because, with such small n, the bias term dominates in the coordinate of $\boldsymbol{\theta}$.

2.2 Quantum Metrology

2.2.1 A minimal problem

In the previous section, we have seen that both classical and quantum estimation problems have an error bound scaling of $O(n^{-1/2})$ with respect to the amount of resource n. In other words, as far as separable states are concerned, there is no distinct advantage in the quantum estimation problem compared with the classical one.

However, quantum mechanics takes a unique role in quantum metrology, which aims at measurements on parameters in a quantum system rather than a specific quantum state. Since such parameters can be only seen through the dynamics of the system, one must first prepare a quantum state called the *probe state* for time evolution in that system. After the evolution, the profile of the probe state will depend on the parameters that govern the dynamics, which one can estimate by measuring the probe state.

We set $\hbar = 1$ for the rest part of this chapter. Then, a pure state is driven by a unitary operator $U = e^{-itH}$, where H is the Hamiltonian and t is the evolution time. A minimalist example is $H = \omega_0 \sigma_z$ with the Pauli matrix $\sigma_z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ in a two-level system $\mathcal{H} = \mathbb{C}^2$. Let $\theta = -2\omega_0 t$ be a dimensionless parameter, and the unitary operator U_θ can be parametrized as $U_\theta = \left(\begin{smallmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{smallmatrix}\right)$.

Now, we consider estimating the parameter θ, and prepare a probe state for the estimation. Noting that U_θ alters the eigenstates $|0\rangle$ and $|1\rangle$ only by the phase factor, we need to prepare the superposition of these eigenstates; in particular, we

consider the eigenstates $|\pm\rangle$ of the Pauli matrix σ_x:

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle), \qquad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.34}$$

We take the initial probe state to be $|+\rangle$. After the evolution, the state will be $|q_{1,\theta}\rangle = U_\theta|+\rangle = \frac{1}{\sqrt{2}}(e^{i\theta/2}|0\rangle + e^{-i\theta/2}|1\rangle)$, which can be measured with the basis $\{|+\rangle, |-\rangle\}$. In fact, the probabilities for observing $+$ and $-$ are

$$p_{1,\theta}(\pm) = |\langle \pm | q_\theta \rangle|^2 = \frac{1 \pm \cos\theta}{2}. \tag{2.35}$$

Therefore, the Fisher information $J_{1,\theta}$ on this stochastic model can be computed as

$$J_{1,\theta} = \frac{1}{p_{1,\theta}(+)}\left[\frac{dp_{1,\theta}(+)}{d\theta}\right]^2 + \frac{1}{p_{1,\theta}(-)}\left[\frac{dp_{1,\theta}(-)}{d\theta}\right]^2 \tag{2.36}$$

$$= \frac{1-\cos\theta}{2} + \frac{1+\cos\theta}{2} = 1, \tag{2.37}$$

indicating the CR bound $\delta \geq 1$.

Next, we consider the case in which n probe states evolve according to U_θ. This situation can be represented by a system consisting of n identical subsystems, where the combined unitary evolution is represented by the tensor product of n copies of $U_{\boldsymbol{\theta}}$:

$$U_\theta^{\otimes n} = \underbrace{U_\theta \otimes \cdots \otimes U_\theta}_{n \text{ times}}. \tag{2.38}$$

One idea for the probe states is just preparing n copies of the probe state $|+\rangle$, one for each subsystem, and separately measure the probe state after the evolution. This yields n i.i.d. stochastic variables subject to Eq. (2.35), where the CR bound can be computed from Eq. (2.25) as

$$\delta \geq n^{-1/2} J_{1,\theta}^{-1/2} = n^{-1/2}. \tag{2.39}$$

On the other hand, we do not necessarily need to prepare separate states in the n systems. In fact, we may set the combined probe state to what is called the *Greenberger–Horne–Zeilinger state* (GHZ state):

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}). \tag{2.40}$$

This quantum states evolves according to the combined unitary operator $U^{\otimes n}$ into

$$|q_\theta\rangle = U^{\otimes n}|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(e^{in\theta/2}|0\rangle^{\otimes n} + e^{-in\theta/2}|1\rangle^{\otimes n}), \tag{2.41}$$

where we note that each unitary operator has shifted the phase of $|0\rangle^{\otimes n}$ by $\theta/2$ and $|1\rangle^{\otimes n}$ by $-\theta/2$. The quantum state model in Eq. (2.41) can simply be estimated by

measuring the first (or any) subsystem with the basis $\{|+\rangle, |-\rangle\}$. The probabilities of observing $+$ or $-$ are:

$$p_\theta(\pm) = \langle q_\theta | [|\pm\rangle\langle\pm| \otimes I^{\otimes(n-1)}] | q_\theta \rangle = \frac{1 \pm \cos n\theta}{2}. \tag{2.42}$$

Now, the Fisher information J_θ for this stochastic model is

$$\begin{aligned} J_\theta &= \frac{1}{p_\theta(+)} \left[\frac{dp_\theta(+)}{d\theta}\right]^2 + \frac{1}{p_\theta(-)} \left[\frac{dp_\theta(-)}{d\theta}\right]^2 \\ &= \frac{n^2(1-\cos n\theta)}{2} + \frac{n^2(1+\cos n\theta)}{2} = n^2. \end{aligned} \tag{2.43}$$

This Fisher information results in the CR bound $\delta \geq n^{-1}$, which is $n^{1/2}$ times smaller than that in Eq. (2.39).

In this way, if we introduce an entangled state such as the GHZ state, one can reduce the estimation error by a significant factor compared with the estimation by separate probes.

2.2.2 General one-parameter metrology

We generalize the previous problem of quantum metrology, and consider the estimation of a parameter θ in the unitary dynamics:

$$U_\theta = e^{i\theta H}. \tag{2.44}$$

Here, H is a Hermitian operator with discrete eigenvalues $\lambda_1 \leq \cdots \leq \lambda_d$. We would like to derive the lower bound on this metrology problem, which must be independent of the choice of the probe state, following Ref. [7].

Let us denote by $|q\rangle$ the initial probe state, which evolves into

$$|q_\theta\rangle = U_\theta |q\rangle = e^{i\theta H} |q\rangle. \tag{2.45}$$

Regarding $\{|q_\theta\rangle\}$ as the quantum state model, the QFI J_θ can be directly computed as

$$\begin{aligned} J_\theta &= 4 \left\langle \frac{dq_\theta}{d\theta} \middle| (I - |q_\theta\rangle\langle q_\theta|) \middle| \frac{dq_\theta}{d\theta} \right\rangle \\ &= 4 \langle q_\theta | (-iH)(I - |q_\theta\rangle\langle q_\theta|)(iH) | q_\theta \rangle \\ &= 4 [\langle q_\theta | H^2 | q_\theta \rangle - \langle q_\theta | H | q_\theta \rangle^2], \end{aligned} \tag{2.46}$$

which is found to be four times the energy variance of $|q_\theta\rangle$ with respect to the Hamiltonian H. Since the unitary evolution $e^{i\theta H}$ preserves the energy distribution of $|q\rangle$, the QFI is maximal when the probe state $|q\rangle$ is the equal superposition of the eigenstates $|\lambda_1\rangle$ and $|\lambda_d\rangle$, in which case

$$J_\theta = (\lambda_d - \lambda_1)^2. \tag{2.47}$$

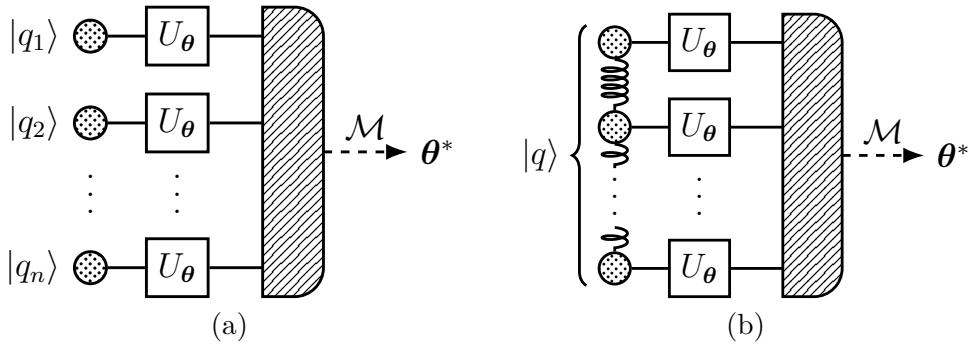


Figure 2.1: Schematic images for contrasting the classical and the quantum schemes. **(a)** Classical scheme, in which n separate probe states are evolved by the unitary dynamics $U_{\boldsymbol{\theta}}$. **(b)** Quantum scheme, in which n probes may be entangled together. In either scheme, the measurement process $\mathcal{M}$ is not restricted; it may be performed comprehensively in the combined system or individually in each of the subsystems. We also allow ancillary systems for the input probes, which are not shown.

That is, the QFI is upper bounded by the squared difference between the maximum and the minimum eigenvalues of H.

We can prepare probe states and perform measurement in the same manner as the two-level system in the last subsection, with $|0\rangle$ and $|1\rangle$ replaced by $|\lambda_1\rangle$ and $|\lambda_d\rangle$. Moreover, the Fisher information in Eq. (2.47) is preserved by the measurement with the basis $\{|+\rangle, |-\rangle\}$:

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|\lambda_1\rangle \pm |\lambda_d\rangle). \tag{2.48}$$

Hence, the CR bound on the RMS error can be written as

$$\delta \geq (\lambda_d - \lambda_1)^{-1}. \tag{2.49}$$

2.2.3 Standard Quantum Limit and Heisenberg Limit

Next, we consider the case in which we have n identical subsystems with unitary dynamics U_θ in Eq. (2.44). First, we consider what we call the *classical scheme* in this thesis, which is schematically depicted in Fig. 2.1 (a). We prepare the separate probe state $|q_1\rangle \otimes \cdots \otimes |q_n\rangle$ and perform measurement $\mathcal{M}$ on the state after the evolution:

$$|q_\theta\rangle = |q_{1,\theta}\rangle \otimes \cdots \otimes |q_{n,\theta}\rangle, \qquad |q_{i,\theta}\rangle = U_\theta |q_i\rangle. \tag{2.50}$$

Since each QFI $J_{i,\theta}$ for the quantum state model $\{|q_{i,\theta}\rangle\}$ is subject to the upper bound in Eq. (2.47), the overall error is lower-bounded by using the additivity of

QFI (Eq. (2.29)):

$$J_\theta = J_{1,\theta} + \cdots + J_{n,\theta} \leq n(\lambda_d - \lambda_1)^2,$$
$$\delta \geq n^{-1/2}(\lambda_d - \lambda_1)^{-1}. \tag{2.51}$$

The scaling $O(n^{-1/2})$ is consistent with the classical and quantum estimation problems, and the error bound is called the *standard quantum limit* (SQL).

On the other hand, the input probe may not necessarily separated; as depicted in Fig. 2.1 (b), we may consider an entangled state for the probe state. We call this scheme as the *quantum scheme*[3]. To derive the error bound in the quantum scheme, we rewrite the combined unitary operator as follows:

$$U_\theta^{\otimes n} = (e^{i\theta H})^{\otimes n} = e^{i\theta\{H\}_n}, \tag{2.52}$$

where the operator $\{H\}_n$ can be defined as

$$\{H\}_n = \sum_{i=1}^{n} I \otimes \cdots \otimes \underbrace{H}_{i\text{th}} \otimes \cdots \otimes I. \tag{2.53}$$

Since Eq. (2.52) again assumes the form of Eq. (2.44), we may apply the same analysis as we did for a single system. The eigenvalues of $\{H\}_n$ can be written as $\lambda_{j_1} + \cdots + \lambda_{j_n}$, with the corresponding eigenvectors $|\lambda_{j_1}\rangle \otimes \cdots \otimes |\lambda_{j_n}\rangle$ Therefore, the maximum and minimum eigenvalues of $\{H\}_n$ are $n\lambda_1$ and $n\lambda_d$, from which the CR bounds on the QFI and the RMS error follow:

$$J_\theta = n^2(\lambda_d - \lambda_1)^2, \tag{2.54}$$
$$\delta \geq n^{-1}(\lambda_d - \lambda_1)^{-1}. \tag{2.55}$$

We observe that the scaling $O(n^{-1})$ of the error bound in Eq. (2.55) decreases faster than the SQL, and is achievable provided that we can prepare the n-body entangled states $\frac{1}{\sqrt{2}}(|\lambda_1\rangle^{\otimes n} + |\lambda_d\rangle^{\otimes n})$. This bound is called the *Heisenberg limit*.

To summarize this section, quantum mechanics can be advantageous in quantum metrology, where an input probe state may be entangled between subsystems. In the classical scheme, there is no such entanglement and the error is bound by the standard quantum limit $O(n^{-1/2})$. In the quantum scheme, the Heisenberg limit $O(n^{-1})$ becomes the new error bound, which can be reached by utilizing the entangled state.

Regarding the analysis, two remarks are in order:

- The bounds given in this section is unaltered even if an ancillary systems are available. In fact, we may extend the unitary operator into the ancillary space: $\tilde{U}_\theta = e^{i\theta(H\otimes I_A)}$, where I_A denotes the identity operator in the ancillary space. Since H and $H \otimes I_A$ shares the same eigenvalues if with different multiplicity, we arrive at the same conclusion.

[3]In the next section, we will refer to this scheme as the *parallel scheme*, because we will introduce other types of quantum schemes.

- Use of mixed probe states instead of pure ones are useless in improving the bound. A mixed state can always be regarded as a pure state by a procedure called purification [56], which introduces an ancillary system and discards the ancillary system in the measurement process.

2.3 Multiparameter quantum metrology

The analysis presented in the last section shows the optimal error bound for a general class of quantum metrology, but with only one parameter to be estimated. However, many physical systems have more than one parameter in question, which makes the problems more complicated. In this section, we review some existing results on multiparameter quantum metrology, including the Hamiltonian estimation problem proposed in the author's master thesis.

2.3.1 Estimation of a unitary gate

Given a d-dimensional Hilbert space $\mathcal{H}_S = \mathbb{C}^d$, any non-dissipative dynamics is equivalent to a unitary gate, which can be described as a unitary operator with unit determinant[4]. Such operators constitute the special unitary group $\mathsf{SU}(d)$, whose dimension is known to be $d^2 - 1$. Therefore, quantum metrology on a completely unknown unitary gate $U_{\boldsymbol{\theta}} \in \mathsf{SU}(d)$ involves estimation of as many as $d^2 - 1$ parameters.

The estimation error of the parameter $\boldsymbol{\theta}$, of course, depends on the parametrization of the unitary matrix $U_{\boldsymbol{\theta}}$. For the purpose of normalization, we assume that $U_{\boldsymbol{\theta}}$ is locally parameterized so that

$$\mathrm{tr}[X_\alpha X_\beta] = \delta_{\alpha\beta}, \qquad X_\alpha = i\frac{\partial U_{\boldsymbol{\theta}}}{\partial \theta_\alpha} U_{\boldsymbol{\theta}}^\dagger. \tag{2.56}$$

This normalization has a geometrical meaning: The parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ can be regarded as the local coordinate of the compact Lie group $\mathsf{SU}(d)$, which is equipped with an $\mathsf{SU}(d)$-invariant Riemannian metric. Equation (2.56) is such an invariant metric, and is unique up to a constant factor. Therefore, Eq. (2.56) indicates that the parameter is consistent with the orthogonal coordinates with which $\mathsf{SU}(d)$ is naturally equipped. See Ref. [61] for detail, which covers geometrical aspects of Lie groups. We also note that $-iX_\alpha$ belongs to the Lie algebra $\mathsf{su}(d)$, meaning that X_α is Hermitian and traceless.

With the help of an ancillary system, in fact, estimation of a unitary gate can be efficiently mapped to the problem of quantum state estimation. We set the probe state to the *maximally entangled state* (MES) $|\mathrm{M}\rangle_{\mathcal{H}_S}$, which is defined as follows: let $\mathcal{H}_A = \mathbb{C}^d$ be the ancillary Hilbert space of the same dimension as $\mathcal{H}_S$,

[4] We ignore the ambiguity of the global phase by multiples of $2\pi/d$, since such discrete ambiguity does not affect the parameter estimation.

and the MES can be written as

$$|\mathrm{M}\rangle_{\mathcal{H}_S} = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |k\rangle_S \otimes |k\rangle_A, \tag{2.57}$$

where $\{|1\rangle_X, \ldots, |d\rangle_X\}$ is a basis of $\mathcal{H}_X$ for $X = S, A$. With the unitary gate $U_{\boldsymbol{\theta}}$ on $\mathcal{H}$, this state will evolve into

$$\begin{aligned} |q_{\boldsymbol{\theta}}\rangle = (U_{\boldsymbol{\theta}} \otimes I_A)|q\rangle_{SA} &= \frac{1}{\sqrt{d}} \sum_{k=1}^{d} U_{\boldsymbol{\theta}} |k\rangle_S \otimes |k\rangle_A \\ &= \frac{1}{\sqrt{d}} \sum_{k=1}^{d} \sum_{l=1}^{d} [U_{\boldsymbol{\theta}}]_{kl} |l\rangle_S \otimes |k\rangle_A. \end{aligned} \tag{2.58}$$

Thus, each coefficient of the basis $|l\rangle_S \otimes |k\rangle_A$ corresponds to an entry of the unitary matrix $U_{\boldsymbol{\theta}}$. In fact, this procedure gives a one-to-one correspondence between quantum channels and quantum states, which is called the Choi-Jamiołkowski isomorphism [62, 63]. The QFI matrix of the probe state in Eq. (2.58) can be calculated as

$$[J_{\boldsymbol{\theta}}]_{\alpha\beta} = \frac{4}{d} \operatorname{Re} \operatorname{tr}[X_\alpha X_\beta] - \operatorname{tr} X_\alpha \operatorname{tr} X_\beta = \frac{4}{d} \delta_{\alpha\beta}, \tag{2.59}$$

where we have used the assumption (2.56).

Now, we consider n copies of the identical unitary channel $U_{\boldsymbol{\theta}}^{\otimes n}$, which acts on the combined Hilbert space $\mathcal{H}_{\mathrm{com}} = \mathcal{H}_S^{\otimes n}$. If we are to estimate $U_{\boldsymbol{\theta}}^{\otimes n}$ as a unitary channel on $\mathcal{H}_{\mathrm{com}}$, we would prepare the MES $|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{com}}}$ in the d^n-dimensional Hilbert space $\mathcal{H}_{\mathrm{com}}$. Noting that

$$i \frac{\partial (U_{\boldsymbol{\theta}}^{\otimes n})}{\partial \theta_\alpha} (U_{\boldsymbol{\theta}}^{\otimes n})^\dagger = \{X_\alpha\}_n, \tag{2.60}$$

the Fisher information in this case can be calculated in the same way as Eq. (2.59):

$$[J_{\boldsymbol{\theta}}^{\mathrm{com}}] = \frac{4}{\dim \mathcal{H}_{\mathrm{com}}} \operatorname{Re} \operatorname{tr}[\{X_\alpha\}_n \{X_\beta\}_n] = \frac{4}{d^n} (n d^{n-1} \operatorname{tr}[X_\alpha X_\beta]) = \frac{4n}{d} \delta_{\alpha\beta}. \tag{2.61}$$

This QFI matrix is n times the QFI matrix (2.59) for the one-channel estimation. In fact, the MES in the combined system is nothing more than separate MES's in the subsystems: $|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{com}}} = |\mathrm{M}\rangle_{\mathcal{H}_S}^{\otimes n}$, and thus this estimation method is actually in the classical scheme. This is consistent with the fact that the QFI matrix in Eq. (2.61) implies the SQL $\delta = O(n^{-1/2})$.

Then, can we attain the error $\delta = O(n^{-1})$ in the quantum scheme? The answer is yes, and such a method can be found by using representation theory. We focus on the fact that the unitary operator $U_{\boldsymbol{\theta}}^{\otimes n}$ preserves the symmetry between the subsystems. This symmetry splits the Hilbert space $\mathcal{H}_{\mathrm{com}} = \mathcal{H}_S^{\otimes n}$ into several

invariant subspaces, the classification of which was given in 1900 by Young [64]. In particular, the largest irreducible subspace is the completely symmetric space $\mathcal{H}_{\rm sym}$, which consists of vectors that are invariant under all permutations of the subsystems. If we begin with the MES $|\mathrm{M}\rangle_{\mathcal{H}_{\rm sym}}$ of this subspace, the QFI matrix is given as

$$[J_{\boldsymbol{\theta}}^{\rm sym}]_{\alpha\beta} = \frac{4}{\dim \mathcal{H}_{\rm sym}} \mathrm{Re}\,\mathrm{tr}[P_{\rm sym}\{X_\alpha\}_n\{X_\beta\}_n], \tag{2.62}$$

where we denote by $P_{\rm sym}$ the projection operator onto $\mathcal{H}_{\rm sym}$. Both the dimension and the trace in Eq. (2.62) can be computed:

$$\dim \mathcal{H}_{\rm sym} = \frac{(n+d-1)!}{n!(d-1)!}, \tag{2.63}$$

$$\mathrm{tr}[P_{\rm sym}\{X_\alpha\}_n\{X_\beta\}_n] = \frac{(n+d)!}{(n-1)!(d+1)!}\,\mathrm{tr}[X_\alpha X_\beta] \tag{2.64}$$

and therefore

$$[J_{\boldsymbol{\theta}}^{\rm sym}]_{\alpha\beta} = \frac{4n(n+d)}{d(d+1)}\delta_{\alpha\beta}, \tag{2.65}$$

which is $\frac{n+d}{1+d}$ times larger than Eq. (2.61). For a sufficiently large number of subsystems: $n \geq d$, we obtain the scaling $O(n^2)$ in the QFI matrix, implying the Heisenberg limit of $O(n^{-1})$. We refer to Ref. [26] for further discussions regarding the optimality of the probe state $|\mathrm{M}\rangle_{\mathcal{H}_{\rm sym}}$ and the attainability of the Heisenberg limit.

2.3.2 Hamiltonian estimation

We have seen that owing to geometry and symmetry, quantum metrology on arbitrary unitary gate has met its optimal error bound. These error bounds cannot necessarily be applied to the estimation of unitary gates with a special property, such as a diagonalized matrix in a certain basis [65]. Thus, it will be convenient if there exists a model of quantum metrology with less specification on the target of estimation.

For this purpose, we consider the Hamiltonian $H_{\boldsymbol{\theta}}$ itself rather than the unitary gate $U_{\boldsymbol{\theta}}$. When we use n copies of the unitary gate $e^{-i\tau H_{\boldsymbol{\theta}}}$ with evolution time τ, the total resource of the estimation is evaluated as $T = n\tau$. It is also possible to use unitary gates with different evolution times, in which case the total resource T is the sum of all evolution times.

The general estimation scheme can be described as Fig. 2.2 (a), which we shall call the *quantum general scheme*. Between the applications of unitary gates $U_{\boldsymbol{\theta}} = e^{-i\tau H_{\boldsymbol{\theta}}}$, one is allowed to apply arbitrary control gates, which can be represented by unitary operators over the main and the ancillary systems. The quantum state after a series of operations will be

$$|q_{\boldsymbol{\theta}}\rangle = (U_{\boldsymbol{\theta}} \otimes I_A)V_{n-1}(U_{\boldsymbol{\theta}} \otimes I_A)V_{n-2}\cdots(U_{\boldsymbol{\theta}} \otimes I_A)V_1|q\rangle, \tag{2.66}$$

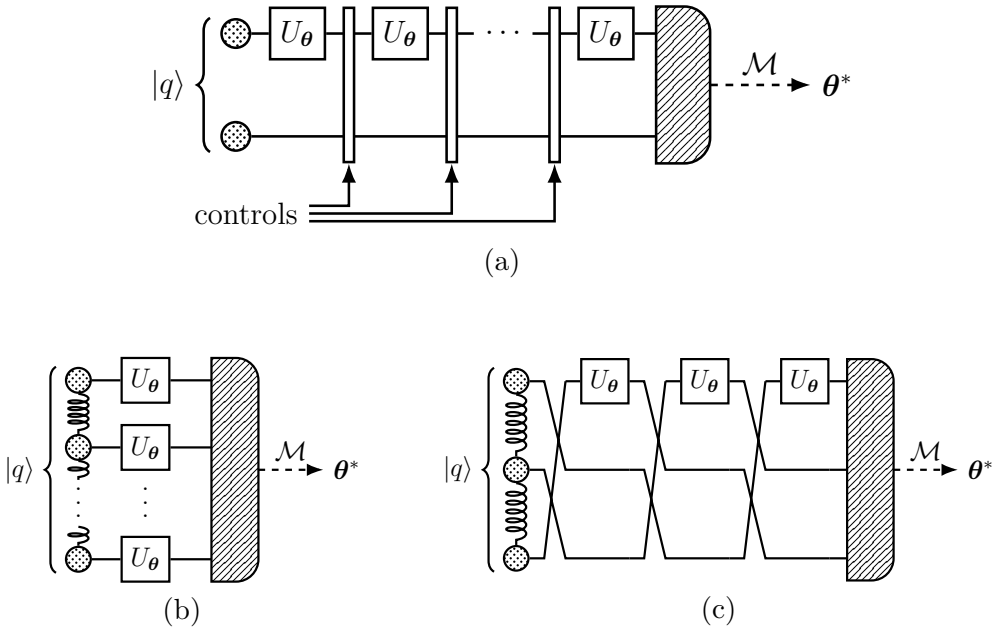


Figure 2.2: Schematic diagrams for contrasting the parallel and the general schemes. For simplicity, we denote by $U_{\boldsymbol{\theta}} = e^{-i\tau H_{\boldsymbol{\theta}}}$ the unitary evolution for the fixed time τ. **(a)** General scheme, where an arbitrary unitary operation is allowed at each of the "controls." **(b)** Parallel scheme, which is the same as Fig. 2.1 (b). Note that ancillary systems are omitted. **(c)** Emulation of the parallel scheme (b) by the corresponding general scheme (a). The bottom two lines are regarded as an ancillary system.

where V_j denotes the jth control unitary operator.

The quantum general scheme is more powerful than the *quantum parallel scheme* in Fig. 2.2 (b), which we have considered in the previous section. In fact, the combination of an ancillary system and unitary operations are sufficient to describe an arbitrary physical dynamics [66, 67], including stochastic processes, partial measurements and feedbacks. In particular, a parallel scheme can be emulated within a general scheme as shown in Fig. 2.2 (c). Therefore, theoretical lower bounds on the estimation error in the quantum general scheme also apply to the parallel quantum scheme, but not vice versa.

2.3.3 Lower error bounds in Hamiltonian estimation

Before we set up the parameter dependence of the Hamiltonian $H_{\boldsymbol{\theta}}$, we note that imposing $\operatorname{tr} H_{\boldsymbol{\theta}} = 0$ does not lose any generality, since adding cI ($c \in \mathbb{R}$) to the Hamiltonian makes no difference in the dynamics.

We assume a Hamiltonian $H_{\boldsymbol{\theta}}$ that linearly depends on Hermitian operators

$X_1, \ldots, X_m$:

$$H_{\boldsymbol{\theta}} = \sum_{j=1}^{m} \theta_j X_j. \tag{2.67}$$

Hamiltonians that can be written in form of Eq. (2.67) constitute an m-dimensional vector space with matrix basis $\{X_1, \ldots, X_m\}$. In order to fix the scale of the parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, we adopt the orthonormal matrix basis:

$$\operatorname{tr} X_j X_k = \delta_{jk}. \tag{2.68}$$

Such a matrix basis can always be taken by linearly transforming the parameter vector $\boldsymbol{\theta}$ with the help of the Gram–Schmidt algorithm.

Here are some practical choices of the matrix bases:

(1) The *full model*, in which an arbitrary Hamiltonian $H_{\boldsymbol{\theta}}$ on a d-dimensional Hilbert space can be estimated. The generalized Gell–Mann matrices can be used as the matrix basis:

$$\begin{cases} \sigma_x^{k,l} = \frac{1}{\sqrt{2}}\big(|k\rangle\langle l| + |l\rangle\langle k|\big) & (1 \le k < l \le d); \\ \sigma_y^{k,l} = \frac{i}{\sqrt{2}}\big(|k\rangle\langle l| - |l\rangle\langle k|\big) & (1 \le k < l \le d); \\ \sigma_z^{k} = \frac{1}{\sqrt{k(k+1)}}\big(k|k+1\rangle\langle k+1| - \sum_{l=1}^{k}|l\rangle\langle l|\big) & (1 \le k < d-1). \end{cases} \tag{2.69}$$

(2) The *multiple-phase model*, in which the Hamiltonian is diagonal with respect to a certain known eigenbasis. The matrix basis can be chosen as $\{\sigma_z^1, \ldots, \sigma_z^{d-1}\}$ from the generalized Gell–Mann matrices in Eq. (2.69). This model leads to the estimation of only diagonal Hamiltonians.

(3) Tensor products of Pauli matrices are often useful in qubit and spin systems. We may consider a matrix basis consisting of matrices of the form

$$\sigma_{(1)} \otimes \sigma_{(2)} \otimes \cdots \otimes \sigma_{(N)}, \tag{2.70}$$

where each of $\{\sigma_{(1)}, \sigma_{(2)}, \ldots, \sigma_{(N)}\}$ is $\sigma_x, \sigma_y, \sigma_z$ or I, but all of them cannot be I. We note that not all matrices are necessarily included in the basis. In fact, the model becomes equivalent to the full model if we consider all $4^N - 1$ matrices, and to the multiple-phase model if σ_x and σ_y are excluded from the choice.

With the assumption above, one can show an upper bound on the QFI matrix by an arbitrary procedure in the general scheme:

Proposition 2.5 *Let $|q_{\boldsymbol{\theta}}\rangle$ be the probe state after the evolution over time T in the quantum general scheme. Then, the QFI matrix $J_{\boldsymbol{\theta}}$ of the quantum state model $\{|q_{\boldsymbol{\theta}}\rangle\}$ satisfies*

$$\operatorname{tr} J_{\boldsymbol{\theta}} \le 4cT^2, \tag{2.71}$$

where $c = \|\mathbf{X}\|$ is defined as the operator norm of $\mathbf{X} = \sum_{j=1}^{m} X_j^2$.

Proof. Let $|q_{\boldsymbol{\theta}}(t)\rangle$ be the probe state after the evolution over time t, and $J_{\boldsymbol{\theta}}(t)$ be the corresponding QFI matrix. Since the unitary operator V_j in the control gate does not depend on the parameter $\boldsymbol{\theta}$, the control gate keeps the QFI matrix unaltered. Therefore, we only need to pursue the time derivative of the QFI matrix $J_{\boldsymbol{\theta}}(t)$ during the Hamiltonian evolution.

By using the Schrödinger equation, the time derivative of an diagonal entry $F_j(t) = [J_{\boldsymbol{\theta}}(t)]_{jj}$ can be evaluated as

$$\frac{dF_j(t)}{dt} = 8\,\mathrm{Im}\left\langle \frac{\partial q_{\boldsymbol{\theta}}(t)}{\partial \theta_j}\middle| [1 - |q_{\boldsymbol{\theta}}(t)\rangle\langle q_{\boldsymbol{\theta}}(t)|] X_j \middle| q_{\boldsymbol{\theta}}(t)\right\rangle \tag{2.72}$$

$$\leq 8\left\| [1 - |q_{\boldsymbol{\theta}}(t)\rangle\langle q_{\boldsymbol{\theta}}(t)|] \middle| \frac{\partial q_{\boldsymbol{\theta}}(t)}{\partial \theta_j}\right\rangle \right\| \cdot \| X_j |q_{\boldsymbol{\theta}}(t)\rangle \| \tag{2.73}$$

$$= 4\sqrt{F_j(t)}\sqrt{\langle q_{\boldsymbol{\theta}}(t)|X_j^2|q_{\boldsymbol{\theta}}(t)\rangle} \tag{2.74}$$

and therefore

$$\frac{d}{dt}(\mathrm{tr}\, J_{\boldsymbol{\theta}}(t)) = \sum_{j=1}^{n} \frac{dF_j(t)}{dt} \leq 4\sum_{j=1}^{n} \sqrt{F_j(t)}\sqrt{\langle q_{\boldsymbol{\theta}}(t)|X_j^2|q_{\boldsymbol{\theta}}(t)\rangle} \tag{2.75}$$

$$\leq 4\left[\sum_{j=i}^{n} F_i(t) \sum_{j=1}^{n} \langle q_{\boldsymbol{\theta}}(t)|X_j^2|q_{\boldsymbol{\theta}}(t)\rangle\right]^{1/2} \tag{2.76}$$

$$= 4[(\mathrm{tr}\, J_{\boldsymbol{\theta}}(t))\langle q_{\boldsymbol{\theta}}(t)|\mathbf{X}|q_{\boldsymbol{\theta}}(t)\rangle]^{1/2} \leq 4c\sqrt{\mathrm{tr}\, J_{\boldsymbol{\theta}}(t)}, \tag{2.77}$$

where we have used the Cauchy–Schwartz inequality in deriving Eq. (2.76). The differential inequality $\frac{d}{dt}(\mathrm{tr}\, J_{\boldsymbol{\theta}}(t)) \leq 4c\sqrt{\mathrm{tr}\, J_{\boldsymbol{\theta}}(t)}$ can be solved and has the maximal solution, which is $\mathrm{tr}\, J_{\boldsymbol{\theta}}(t) = 4ct^2$. We note that the QFI matrix at $t = 0$ is zero, since the initial probe state $|q\rangle$ does not depend on the parameter vector $\boldsymbol{\theta}$. By substituting $t = T$, we obtain the desired inequality (2.71). □

From the Prop. 2.5 and the quantum CR inequality (Th. 2.4), we can bound the RMS error δ of an unbiased estimator as

$$\delta \geq (\mathrm{tr}[J_{\boldsymbol{\theta}}^{-1}])^{1/2} \geq \frac{m}{(\mathrm{tr}\, J_{\boldsymbol{\theta}})^{1/2}} \geq \frac{m}{2c^{1/2}T}, \tag{2.78}$$

in which we have applied the Cauchy–Schwartz inequality again: $\mathrm{tr}[J^{-1}] \cdot \mathrm{tr}\, J \leq (\mathrm{tr}\, I)^2 = m^2$.

Aside from the number m of parameters, the lower bound in Eq. (2.78) contains the model-dependent constant c. Noting that $\mathrm{tr}\, \mathbf{X} = m$ follows from the orthonormal condition (2.68), the minimum possible value is found to be $c = m/d$, which is the case when $\mathbf{X}$ is in proportion to the identity matrix.

Proposition 2.6 *Let us call a Hamiltonian model in Eq.* (2.67) spherical *if and only if* $\sum_{j=1}^{m} X_j^2 = \frac{m}{d} I$ *holds. In this case, the quantum CR bound on the RMS error is given as*

$$\delta \geq \frac{(md)^{1/2}}{2T}. \tag{2.79}$$

The spherical model indicates that, when the parameter vector $\boldsymbol{\theta}$ is subject to a stochastic distribution with spherical symmetry, the variance of the energy does not depend on the probe state. The examples (1)–(3) on p. 33 are all spherical models.

2.3.4 Adaptive estimation method

An important problem arises when we consider a concrete estimation procedure: More than one candidate of parameter vector $\boldsymbol{\theta}$ may correspond to the same probe state after evolution. As the simplest case, we consider a one-parameter problem $H_\theta = \theta\sigma_z$. The unitary gate $U_\theta = e^{-i\tau H_\theta}$ becomes the identity if θ is a multiple of π/τ. Hence, one cannot distinguish between the candidates $\theta = 0, \pm\pi/\tau, \pm 2\pi/\tau, \ldots$, since the distribution of the measurement outcomes is exactly the same for any of the candidates.

This problem is inevitable when we consider the parameter in the unbounded space $\mathbb{R}^m$. To circumvent this problem, we must assume that the parameter vector $\boldsymbol{\theta}$ belongs to a certain bounded domain in the first place. More particularly, we introduce a parameter $E > 0$ such that the inequality

$$\|\boldsymbol{\theta}\| \leq E \tag{2.80}$$

is provided before the estimation. We note that under the normalization (2.68), the parameter $\boldsymbol{\theta}$ has the dimension of frequency or energy, whence E may be regarded as the energy scale. Under the assumption (2.80), the correspondence between the unitary gate $U_{\boldsymbol{\theta}} = e^{-i\tau H_{\boldsymbol{\theta}}}$ and the parameter $\boldsymbol{\theta}$ becomes one-to-one as long as $\tau \leq 1/E$. Conversely, if τ is larger than $O(1/E)$, such a one-to-one correspondence becomes impossible.

Let us fix the total time resource T. In the classical scheme, we prepare n separate probe states and let them evolve over $\tau = T/n$. As long as $\tau \leq 1/E$, the inequality in Prop. 2.5 can be satisfied up to a model-independent constant factor:

$$\operatorname{tr} J_{1,\boldsymbol{\theta}} = O(c\tau^2), \tag{2.81}$$

where $J_{1,\boldsymbol{\theta}}$ denotes the QFI matrix for one probe state. Consequently, the QFI for n probe states satisfy $\operatorname{tr} J_{\boldsymbol{\theta}} = O(cn\tau^2) = O(cT\tau)$, indicating that the evolution time τ per gate should be as large as possible. Since τ must be in the order of $O(1/E)$, we obtain $\operatorname{tr} J_{\boldsymbol{\theta}} = O(cT/E)$ and

$$\delta = O(m(E/cT)^{1/2}). \tag{2.82}$$

Here, the scaling $O(T^{-1/2})$ with respect to the total time resource T corresponds to the SQL as expected.

In the quantum general scheme, on the other hand, the theoretical bound on the RMS error $\delta = O(m/(c^{1/2}T))$ in Eq. (2.78), whose order is $O(T^{-1})$ and corresponds to the Heisenberg limit. Although no methods that satisfy $\delta = O(m/(c^{1/2}T))$ are currently known, we can attain an $O(T^{-1})$ error by what

is called the *adaptive estimation* [13, 68]. In the adaptive estimation, we compute a series of estimators $\boldsymbol{\theta}_0^*, \boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_K^*$ one by one such that the inequality

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}_k^*\| \leq 2^{-k}E \tag{2.83}$$

is satisfied with a sufficiently large probability. We begin with $\boldsymbol{\theta}_0^* = 0$, where the inequality (2.83) is satisfied because of Eq. (2.80). When an estimator $\boldsymbol{\theta}_k^*$ is obtained, we continue the estimation by adjusting the control gates depending on the value of $\boldsymbol{\theta}_k^*$. With this control, we can regard the system as Hamiltonian estimation with energy scale $2^{-k}E$, owing to which the effective per-gate evolution time can be increased: $\tau = O(2^k/E)$. Since larger τ implies more efficient estimation, we obtain a new estimator $\boldsymbol{\theta}_{k+1}^*$ with much better accuracy.

After K iterations of this procedure, the RMS error of the last estimator $\boldsymbol{\theta}_K^*$ is approximately $\delta = 2^{-K}E$, whence an arbitrarily small error can be attained. By adjusting the schedule on which the total evolution time T is consumed for each estimator, this error becomes $O(T^{-1})$ as desired. We refer to Sec. 4.1 for the detailed mathematics regarding this adaptive technique.

In the author's master thesis, the adaptive estimation in the quantum parallel scheme has also been presented. While it has been suggested that the estimation error in the parallel scheme is the same up to a constant factor as the non-parallel counterpart, a rigorous proof for this conjecture had remained open due to the chaotic behavior involved. In this thesis, this issue has been successfully handled in Sec. 4.2.

Chapter 3

Reviews on function estimation

In this chapter, we will review classical estimation theory for continuous data, which can be formulated as functions in mathematics. Among other forms of continuous data, estimation of time-dependent signal has been especially important in history, which is why many important technique has developed in signal detection theory. Here, we will focus on two problems: In the first section, we discuss how to restore the entire function from discretely sampled data, which can be solved by using *local linear smoothing* (LLS) [69–72]. In the second section, we introduce signal detection theory for continuously observed signals, where filters can be used to cancel stochastic noises.

3.1 Restoring functions

3.1.1 Significance of regularity

Compared with parameter estimation, the most significant characteristics of function estimation is the infinite degrees of freedom. In fact, given samples y_i at finite sample points $x_1 < x_2 < \cdots < x_n$, there are infinitely many functions f satisfying $y_i = f(x_i)$ for $i = 1, \ldots, n$. Furthermore, the value $f(x)$ can be arbitrarily large in positive or negative if, say, $x_1 < x < x_2$. Knowing that the function f is continuous, we can anticipate that $f(x)$ is close to the neighboring samples $y_1 = f(x_1)$ or $y_2 = f(x_2)$. However, since sole continuity does not provide any quantitative information, there is no way to estimate the value of $f(x)$ within a finite error without any further assumptions.

Therefore, the function estimation necessarily involves the quantitative assumption on the target function, which we refer to as the *regularity*. The simplest assumption is that the function be differentiable, and that its derivative $f'(X)$ be uniformly bounded:

$$|f'(x)| \leq M \tag{3.1}$$

for some positive constant M. With this assumption, one can estimate the value $f(x)$ as, e.g., $y_1 - M|x - x_1| \leq f(x) \leq y_1 + M|x - x_1|$. By taking sufficiently many samples, one can estimate the function value $f(x)$ at an arbitrary point x.

Equation (3.1) is not the only form of regularity. In particular, we may bound the derivative $f'(x)$ in terms of the *root mean square* (RMS) instead of uniformity:

$$\frac{1}{b-a}\int_a^b |f'(x)|^2 dx \leq M^2, \tag{3.2}$$

where we have denoted the domain of the function $f(x)$ by $a \leq x \leq b$. This is useful especially when we consider the RMS error by analogy with the parameter estimation problem. Let f^* be the estimator for the function f, and the RMS error δ is defined as

$$\delta^2 = \left\langle \frac{1}{b-a}\int_a^b |f^*(x) - f(x)|^2 dx \right\rangle, \tag{3.3}$$

where the expectation value $\langle\cdot\rangle$ is taken over the stochastic behavior of the estimator f^*.

We can also assume that the function f is differentiable more than once. In this case, different regularity can be defined by using the mth derivative $f^{(m)}(x)$:

$$|f^{(m)}(x)| \leq M \tag{3.4}$$

or

$$\frac{1}{b-a}\int_a^b \left|f^{(m)}(x)\right|^2 \leq M^2. \tag{3.5}$$

In some circumstances, such regularity can help us perform estimation with better accuracy, although the procedure of the estimation becomes more intricate.

3.1.2 Local linear smoothing

The goal of this section is to restore the entire function $f(x)$ ($x \in \mathbb{R}$) from the discrete samples $y_i = f(x_i)$ by performing the LLS. Here, we assume that the sample points $x_j = jh$ ($j \in \mathbb{Z}$) is taken with an equal spacing h.

An estimator of this function can be defined as

$$f^*(x) = \sum_{j\in\mathbb{Z}} y_j \phi\left(\frac{x - x_j}{h}\right), \tag{3.6}$$

by choosing an appropriate function $\phi(x)$. For the summation in Eq. (3.6) to be essentially finite, the function ϕ must be zero outside a finite interval. One choice of such function ϕ is

$$\phi(z) = \begin{cases} 1+z & (-1 \leq z \leq 0); \\ 1-z & (0 \leq z \leq 1); \\ 0 & (\text{otherwise}), \end{cases} \tag{3.7}$$

which results in a linear interpolation between each points:

$$f^*(\theta x_j + (1-\theta)x_{j+1}) = \theta y_j + (1-\theta)y_{j+1} \qquad (0 \leq \theta \leq 1). \tag{3.8}$$

This linear interpolation serves as a second-order approximation of the function $f(x)$; when the function f is equipped with the regularity $|f''(x)| \leq M$, the estimation error can be evaluated as $|f^*(x) - f(x)| \leq \frac{h^2 M}{2}$.

In general, the LLS can be performed on the basis of an mth-order approximation by choosing the function ϕ as follows:

Definition 3.1 If a function $\phi\colon \mathbb{R} \to \mathbb{R}$ satisfies

$$\sum_{j\in\mathbb{Z}} P(j)\phi(z-j) = P(z) \tag{3.9}$$

for an arbitrary polynomial P with degree at most m, the function ϕ is called *an LLS kernel* of degree m. Furthermore, if $\phi(x) = 0$ holds for $x > \frac{s}{2}$ for some positive integer s, the LLS kernel ϕ is said to have support s.

Let us see that such an LLS kernel can be constructed for $s \geq m+1$. We define the function ϕ as a piecewise polynomial function:

$$\phi(x) = \begin{cases} \phi_k(z-k) & (k - \frac{1}{2} \leq z < k + \frac{1}{2}); \\ 0 & (\text{otherwise}), \end{cases} \tag{3.10}$$

where $\phi_k(z)$ $(k = -\frac{s-1}{2}, -\frac{s-3}{2}, \ldots, \frac{s-1}{2})$ are polynomials with degree at most m.

Then, ϕ is an LLS kernel of degree m if and only if

$$\sum_{k=-\frac{s-1}{2}}^{\frac{s-1}{2}} k^l \phi_k(z) = z^l, \qquad \forall\, l = 0, 1, \ldots, m, \tag{3.11}$$

which can be interpreted as a system of linear equations. Since the coefficient matrix of this system is a Vandermonde matrix, this system has at least one solution as long as $s \geq m+1$. We show in Fig. 3.1 LLS kernels ϕ constructed in this way for $s = m+1$ and $m = 0, 1, 2$. In particular, the LLS kernel with $m = 1$ corresponds to the linear-interpolating function in Eq. (3.7). We note that the functions are not necessarily continuous; we can also choose smoother functions by taking a larger s [72].

We use an LLS kernel ϕ of degree $m-1$ with support s. To evaluate the estimation error $f^*(x) - f(x)$, we first remind the Taylor expansion $f(x') = P_x(x') + R_x(x')$ with the polynomial part P_x and the residual part R_x defined as

$$P_x(x') = \sum_{l=0}^{m-1} f^{(l)}(x) \frac{(x'-x)^l}{l!}, \tag{3.12}$$

$$R_x(x') = \int_x^{x'} \frac{(x'-t)^{m-1}}{(m-1)!} f^{(m)}(t)dt. \tag{3.13}$$

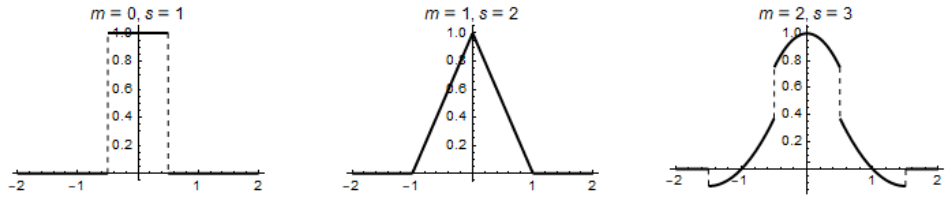


Figure 3.1: Plots of local linear smoothing kernels $\phi(x)$ $(-2 \leq x \leq 2)$ of degree m with support s, where $(m, s) = (0, 1), (1, 2)$, and $(2, 3)$.

Substitution of the Taylor expansion into Eq. (3.6) yields

$$\begin{aligned} f^*(x) &= \sum_{j\in\mathbb{Z}} \phi\left(\frac{x - x_j}{h}\right) f(x_j) = \sum_{j\in\mathbb{Z}} \phi\left(\frac{x}{h} - j\right)[P_x(x_j) + R_x(x_j)] \\ &= P_x(x) + \sum_{j\in\mathbb{Z}} \phi\left(\frac{x}{h} - j\right) R_x(x_j), \end{aligned} \tag{3.14}$$

and hence

$$f^*(x) - f(x) = f^*(x) - P_x(x) = \sum_{j\in\mathbb{Z}} \phi\left(\frac{x - x_j}{h}\right) R_x(x_j). \tag{3.15}$$

When the target function f is equipped with the regularity in Eq. (3.4), we obtain

$$|f^*(x) - f(x)| \leq \sum_{j\in\mathbb{Z}} \left|\phi\left(\frac{x - x_j}{h}\right)\right| \cdot \frac{|x_j - x|^m}{m!} M. \tag{3.16}$$

Here, $\phi(\frac{x-x_j}{h})$ is nonzero only if $|x - x_j| \leq \frac{sh}{2}$, which can be satisfied by at most s instances of j. By using these facts, the estimation error can be bounded as follows:

Theorem 3.2 *If φ is an LLS kernel of degree $m - 1$ and has support s, any function f with $|f^{(m)}(x)| \leq M$ can be estimated by the LLS within error*

$$|f^*(x) - f(x)| \leq Ch^m M, \tag{3.17}$$

where the constant C is set to be $C = \frac{s^{m+1}}{2^m m!} \max_z |\varphi(z)|$.

In other words, the error decreases down to the order $O(h^m)$ as the spacing h of the sample points gets smaller.

3.1.3 Hölder continuity

At this point, we assume the differentiability to the mth order as the regularity of the function, which results in the $O(h^m)$ scaling of the estimation error. However,

we may also define the regularities corresponding to a positive number q instead of an integer m, by which we can attain a finer scaling law. In particular, we may define the regularities with $0 < q < 1$, which provide the means to evaluate estimation errors for less smooth functions than the differentiable ones.

Definition 3.3 (Hölder continuity [73]) Let $0 < \sigma \leq 1$. A function $f\colon I \to \mathbb{R}$ defined over a real interval I is said to be *σ-Hölder continuous* if and only if there exists a number $M \geq 0$ such that

$$|f(x') - f(x)| \leq M|x' - x|^{\sigma} \tag{3.18}$$

for any $x, x' \in I$. The minimum of such a number M is said to be the *σ-Hölder norm* of the function f.

A typical example of a σ-Hölder continuous function is $f(x) = \alpha|x|^{\sigma} + \beta$, whose σ-Hölder norm equals $|\alpha|$. A function $f(x) = x$ defined over a finite interval $[a, b]$ is also σ-Hölder continuous function, whose Hölder norm is $(b-a)^{1-\sigma}$. However, the same function $f(x) = x$ defined over $\mathbb{R}$ is not σ-Hölder continuous unless $\sigma = 1$, because the said Hölder norm diverges for $b - a \to \infty$.

Any positive number q can be decomposed into the sum of a nonnegative integer $m \geq 0$ and a number $0 < \sigma \leq 1$. By combining the differentiability and the Hölder continuity, we can establish what can be regarded as the differentiability to the qth order.

Definition 3.4 A function $f\colon I \to \mathbb{R}$ is said to be of *class $C^{m,\sigma}$* if its mth derivative $f^{(m)}$ is σ-Hölder continuous. A regularity corresponding to this class can be defined as

$$|f^{(m)}(x') - f^{(m)}(x)| \leq M|x' - x|^{\sigma}. \tag{3.19}$$

The estimation error for the function with regularity (3.19) can be derived as follows: We consider an LLS kernel $\varphi(z)$ of degree m and support s, and deform Eqs. (3.12) and (3.13) in the Taylor expansion:

$$P_x(x') = \sum_{l=0}^{m} f^{(l)}(x)\frac{(x' - x)^l}{l!}, \tag{3.20}$$

$$R_x(x') = \int_x^{x'} \frac{(x' - t)^{m-1}}{(m-1)!}[f^{(m)}(t) - f^{(m)}(x)]dt. \tag{3.21}$$

Since the equation (3.15) also holds for this case, we can bound the estimation error in a similar manner to Th. 3.2.

Theorem 3.5 *If φ is an LLS kernel of degree m and support s, any function f with the regularity* (3.19) *can be estimated by the LLS within error*

$$|f^*(x) - f(x)| \leq Ch^q M, \tag{3.22}$$

where $q = m + \sigma$ and the constant C is set to be $C = \frac{s^{q+1}}{2^q q(m-1)!}\max_z|\varphi(z)|$.

We also note that the 1-Hölder continuity, also known as the *Lipschitz continuity*, is a generalized concept of the differentiability. The estimation error for the Lipschitz continuous functions are the same as that for the differentiable functions.

3.2 Canceling stochastic noises

3.2.1 Power-spectrum analysis

In this section, we consider a prior distribution of a continuous signal $s(t)$ according to Bayesianism. Such a family of real-parametrized stochastic variables $\{s(t)\}_{t\in\mathbb{R}}$ is called a *stochastic process*, and plays a central role in probabilistic theory.

In this review, we assume two basic properties of the signal as follows:

$$\langle s(t)\rangle = 0, \tag{3.23}$$

$$\langle s(t)s(t')\rangle = C_s(t-t'). \tag{3.24}$$

The former equation (3.23) implies that we assume the signal tends to be neither positive nor negative. The latter equation (3.24) indicates the *stationarity* of the stochastic process, by which we mean that the process remains invariant under the translation of time. Therefore, the signal can be characterized by the *correlation function* $C_s(t)$, which describes the second-order statistics of the signal $s(t)$. We may optionally assume that $s(t)$ is a *Gaussian process*, i.e., if any combination of snapshots $\big(s(t_1),\ldots,s(t_n)\big)$ is subject to a multivariate Gaussian distribution. In this case, the stochastic process is completely determined by the correlation function $C_s(t)$.

Let us consider the (unitary) Fourier transform of the signal $s(t)$ within a finite range:

$$s_T(\omega) = \frac{1}{\sqrt{T}}\int_{-T/2}^{T/2} s(t)e^{i\omega t}dt. \tag{3.25}$$

The *power spectrum* of the stochastic process $\{s(t)\}$ is defined by the $T\to\infty$ limit of the expected square amplitude of $s_T(\omega)$:

$$I_s(\omega) = \lim_{T\to\infty}\langle|s_T(\omega)|^2\rangle. \tag{3.26}$$

The Wiener–Khinchin theorem states that $I_s(\omega)$ is equal to the Fourier transform of the correlation function [74]:

$$I_s(\omega) = \int_{-\infty}^{\infty} C_s(t)e^{i\omega t}dt, \tag{3.27}$$

which can be confirmed by substituting Eq. (3.25) in Eq. (3.26). We also note that the power spectrum $I_s(\omega)$ is nonnegative and symmetric by definition.

The power spectrum describes the typical behavior of the signal $s(t)$ more clearly than the correlation function. That is, the number $I_s(\omega)$ denotes the weight of the fluctuation with frequency ω, which is integrated to give the total fluctuation: $\int_{-\infty}^{\infty} I(\omega)d\omega = \langle|s(t)|^2\rangle$. Since fluctuations at large frequencies lead to rapid variation of the sample function, the typical smoothness of the signal is determined by the speed at which the power spectrum decays in the limit $|\omega|\to\infty$. Hence, we assume a power-law decay $I_s(\omega)\sim|\omega|^{-p}$ $(p>1)$ of the power spectrum at

$|\omega| \to \infty$, with large p corresponding to a smoother signal. For the computational purpose, the power spectrum is often parametrized as [43]

$$I_s(\omega) = \kappa^2(\omega_0^2 + \omega^2)^{-p/2}, \tag{3.28}$$

where $\kappa > 0$ and $\omega_0 > 0$ characterizes the global amplitude and frequency of the signal, respectively.

In the next section, we also introduce the stochastic process of the noise $n(t)$ that affects the observation of the signal. We assume that the signal and noise are mutually independent, and thus free of the cross-correlation term: $\langle s(t)n(t')\rangle = 0$.

The correlation function $C_n(t)$ and the power spectrum $I_n(\omega)$ of the noise can be determined similarly. A common type of noise is the *white noise*, which is defined via its power spectrum

$$I_{n,\text{white}}(\omega) = \sigma^2. \tag{3.29}$$

The corresponding correlation function involves Dirac's delta function: $C_{n,\text{white}}(t) = \sigma^2\delta(t)$, indicating that the correlation $\langle n(t)n(t')\rangle$ is zero for $t \neq t'$ and infinite for $t = t'$. Due to this divergence, a white-noise process only appears as the theoretical limit of a physical stochastic process. An actual process is limited by the frequency cutoff Ω, which originates from the physical processes of the noise generation. The corresponding power spectrum satisfies $I_n(\omega) = \sigma^2$ only within the frequency $|\omega| \leq \Omega$ and must rapidly decay for $|\omega| > \Omega$. Since the time resolution of the physical measurement process is much larger than the timescale Ω^{-1} of the noise, the actual noise can be regarded as white, despite its unphysical nature.

3.2.2 Filtering

We suppose that a signal $s(t)$ is continuously measured, but suffers from stochastic noise $n(t)$. Therefore, the observed function $r(t)$ is

$$r(t) = s(t) + n(t). \tag{3.30}$$

The goal is to estimate the signal $s(t)$ from $r(t)$, given the power spectra $I_s(\omega)$ and $I_n(\omega)$ of the signal and the noise.

The linear estimator provides a simple way to perform this estimation, and is optimal for the Gaussian signal and noise [74].

Definition 3.6 A *linear estimator* $s^*(t)$ of this signal can be obtained as the convolution of the observation $r(t)$ and a fixed function $h(t)$:

$$s^*(t) = h(t) * r(t) = \int_{-\infty}^{\infty} h(t')r(t-t')dt'. \tag{3.31}$$

The function $h(t)$ is called the *filter*. If $h(t)$ is zero for negative t, the filter $h(t)$ is called *causal*; otherwise, $h(t)$ is called *non-causal*.

The value of the filter function $h(t')$ represents the contribution of the observation $r(t-t')$ to the estimator $s^*(t)$ at t' time later. Therefore, we can obtain the estimator $s^*(t)$ without having to wait for further observation after time t if and only if the filter is causal. In this review, we assume the use of non-causal filters, although the pursuit for causal filters is also important for the practical use.

Let us determine the optimal filter function $h(t)$. The error of the linear estimator $e(t)$ can be written as

$$\begin{aligned} e(t) = s^*(t) - s(t) &= h(t) * [s(t) + n(t)] - s(t) \\ &= [h(t) - \delta(t)] * s(t) + h(t) * n(t). \end{aligned} \tag{3.32}$$

The correlation function of $e(t)$ can be calculated as

$$\begin{aligned} C_e(t) &= \langle e(t) * e(-t) \rangle \\ &= [h(t) - \delta(t)] * \langle s(t) * s(-t) \rangle * [h(-t) - \delta(-t)] \\ &\quad + 2[h(t) - \delta(t)] * \langle s(t) * n(-t) \rangle * h(-t) \\ &\quad + h(t) * \langle n(t) * n(-t) \rangle * h(-t) \\ &= [h(t) - \delta(t)] * C_s(t) * [h(-t) - \delta(-t)] + h(t) * C_n(t) * h(-t). \end{aligned} \tag{3.33}$$

By the Fourier transform, the power spectrum of the error can be obtained:

$$\begin{aligned} I_e(\omega) &= [H(\omega) - 1] I_s(\omega) [\overline{H(\omega) - 1}] + H(\omega) * I_n(\omega) \overline{H(\omega)} \\ &= |H(\omega) - 1|^2 I_s(\omega) + |H(\omega)|^2 I_n(\omega), \end{aligned} \tag{3.34}$$

where $H(\omega)$ is defined as the Fourier transform of the filter $h(t)$. We notice from Eq. (3.34) that the optimal filter is a compromise between two extremal choices $H(\omega) = 0$ ($h(t) = 0$) and $H(\omega) = 1$ ($h(t) = \delta(t)$). In the former case, we completely discard the observed signal to avoid picking up any noise; in the latter case, we just trust the observation without considering the noise.

Since the RMS error δ is obtained as the integral $\delta^2 = \int_{-\infty}^{\infty} I_e(\omega) d\omega$, the optimal filter is obtained by minimizing Eq. (3.34) for each ω:

$$H(\omega) = \frac{I_s(\omega)}{I_s(\omega) + I_n(\omega)}, \tag{3.35}$$

in which case the error is given as

$$\delta^2 = \int_{-\infty}^{\infty} \frac{I_s(\omega) I_n(\omega)}{I_s(\omega) + I_n(\omega)} d\omega = 2 \int_0^{\infty} \frac{I_s(\omega) I_n(\omega)}{I_s(\omega) + I_n(\omega)} d\omega. \tag{3.36}$$

The linear filter is useful even when we do not precisely know the power spectra $I_s(\omega)$ and $I_n(\omega)$. In fact, an interesting choice of the filter is:

$$H(\omega) = \begin{cases} 1 & (|\omega| \leq \Omega); \\ 0 & (|\omega| > \Omega). \end{cases} \tag{3.37}$$

This corresponds to a sinc function $h(t) = t^{-1} \sin 2\Omega t$, whence the name *sinc filter*. This result in the estimation error

$$\delta^2 = 2 \int_0^\Omega I_n(\omega) d\omega + 2 \int_\Omega^\infty I_s(\omega) d\omega. \tag{3.38}$$

In most cases, the power spectrum of the signal $I_s(\omega)$ vanishes in the $\omega \to \infty$ limit faster than that of the noise $I_n(\omega)$. Hence, a suboptimal sinc filter can be obtained by choosing such a frequency Ω that $I_n(\Omega) = I_s(\Omega)$. In fact, we can show that the RMS is at most $\sqrt{2}$-time larger than the optimal filter:

$$\delta^2 = 2 \int_0^\infty \min\{I_s(\omega), I_n(\omega)\} d\omega \le 4 \int_0^\infty \frac{I_s(\omega) I_n(\omega)}{I_s(\omega) + I_n(\omega)} d\omega. \tag{3.39}$$

In particular, when we have $I_s(\omega) \approx \kappa^2 \omega^{-p}$ for $\omega \gg 1$ and $I_n(\omega) = \sigma^2$, the optimal error is in the order $O((\kappa \sigma^{1-p})^{2/p})$.

3.2.3 Connection to the function estimation

We recall the function class $C^{m,\sigma}$ (Def. 3.4), which introduces a wide variety of function regularities which are continuously parametrized by $q = m + \sigma > 0$. For the last part of this section, we see that a signal $s(t)$ with power spectrum $I_s(\omega) \sim |\omega|^{-p}$ can roughly be regarded as a function of class $C^{m,\sigma}$, where p and q are related to each other by $p = 2(m + \sigma) + 1 = 2q + 1$. This suggests an abstract relation between the estimation of deterministic functions and that of stochastic processes, which appears as the same scaling law of estimation error in quantum metrology in Chapter 5.

Let us assume that the power spectrum $I_s(\omega)$ of the stochastic process $\{s(t)\}$ satisfies

$$I_s(\omega) \le \min\{I_0, \kappa^2 |\omega|^{-p}\} \tag{3.40}$$

for $p > 1$ and $\kappa, I_0 > 0$. If $1 < p < 3$, the mean-square difference of the signal at two times t, t' with $\tau = |t - t'|$ satisfies

$$\begin{aligned} \left\langle |s(t') - s(t)|^2 \right\rangle &= 2C_s(0) - 2C_s(\tau) \\ &= \frac{2}{\pi} \int_0^\infty I_s(\omega)(1 - \cos \omega\tau) d\omega \\ &\le \frac{2}{\pi} \int_0^\infty \frac{\kappa^2 \omega^{-p}}{2} \min\{(\omega\tau)^2, 4\} d\omega \\ &= \frac{1}{2\pi} \kappa^2 \tau^{p-1} \int_0^\infty z^{-p} \min\{z^2, 4\} dz \le C \tau^{p-1}, \end{aligned} \tag{3.41}$$

where

$$C = \frac{2^{2-p} \kappa^2}{\pi (3 - p)(p - 1)}. \tag{3.42}$$

This indicates that the signal $s(t)$ satisfies σ-Hölder continuity in a stochastic manner:

$$|s(t') - s(t)| \lesssim \text{const.} \times |t' - t|^{(p-1)/2} \qquad (t' - t \to 0). \tag{3.43}$$

When $p = 3$, however, an analysis similar to Eq. (3.41) necessarily involves a logarithmic correction:

$$\langle |s(t') - s(t)|^2 \rangle \leq C' |t' - t|^2 \log |t' - t|^{-1}, \tag{3.44}$$

which is known to be inevitable [75].

For $p > 3$, it is known that the derivative $\dot{s}(t) = \frac{ds}{dt}(t)$ of the signal can be defined for almost all $t \in \mathbb{R}$ with probability one. When $\{\dot{s}(t)\}$ is seen as a stochastic process, its correlation function is

$$C_{\dot{s}}(t' - t) = \langle \dot{s}(t)\dot{s}(t') \rangle = \frac{\partial^2}{\partial t \partial t'} \langle s(t)s(t') \rangle = -C_s''(t' - t) \tag{3.45}$$

and its power spectrum becomes

$$I_{\dot{s}}(\omega) = \omega^2 I_s(\omega). \tag{3.46}$$

Therefore, when $3 < p \leq 5$, the signal s is almost surely differentiable and the derivative $\dot{s}$ is endowed with the mean-square $\frac{p-3}{2}$-Hölder continuity (3.41) except for $p = 5$, which leads to its corrected version (3.44).

The general theory for $p = 2(m + \sigma) + 1$ can be obtained by repeating this argument; the mth derivative $s^{(m)}$ is defined and satisfies the mean-square σ-Hölder continuity (with a logarithmic correction for $\sigma = 1$), which roughly corresponds to the $C^{m,\sigma}$ class.

We note that, although the stochastic Hölder continuity of the signal $s(t)$ holds, a typical signal $s(t)$ may not be Hölder continuous. However, if we compromise by an arbitrarily small $\epsilon > 0$ assume power spectrum $I_0(\omega) \sim |\omega|^{-p-\epsilon}$, the signal $s(t)$ belongs to the class $C^{m,\sigma}$ with probability one and satisfies the regularity in Eq. (3.19) for an appropriately chosen $M > 0$. For more advanced and complete discussions, please see Refs. [75, 76].

Chapter 4

On parallelization of multiparameter quantum metrology

In this chapter, we derive the estimation errors for two methods of Hamiltonian estimation. One method is in the *sequential scheme*, in which the Hamiltonian acts in only one system without quantum memory, and the other method is in the *parallel scheme*, in which the Hamiltonian acts in many subsystems in parallel without quantum control. They both exhibit the Heisenberg limit $O(T^{-1})$ with respect to the evolution time T (we note that the time is multiplied by the number of subsystems in the parallel scheme). In Ref. [30], both methods has already been shown to be optimal in terms of the Cramér–Rao (CR) bound, and the error bound on the sequential method was found to be $O(d^{1/2})$ times smaller than that on the parallel method.

However, the comparison with the same evolution time T is not appropriate, since whether or not the CR bound is attainable may depend on T. Therefore, we compute the optimal error with the attainability of the CR bound in consideration, and show that both methods require the same amount of resource up to a constant factor, regardless of the size of the Hamiltonian model. This means that the evolution time required for the Heisenberg-limited Hamiltonian estimation can be efficiently reduced by preparing many copies of a given physical system.

The first section is dedicated for the *hierarchical estimation*, an organized framework of adaptive estimation that can provide an upper bound on the estimation error. Then, we perform rigorous analysis on the estimation errors in both sequential and parallel schemes in the second section.

4.1 Hierarchical estimation

Let us recall how the adaptive estimation helps the Heisenberg-limited measurement.Following Sec. 2.3.2, We consider a spherical model in the Hamiltonian estimationm where the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$ is estimated from a Hamiltonian $H_{\boldsymbol{\theta}} = \sum_{j=1}^m \theta_j X_j$ over the d-dimensional Hilbert space $\mathcal{H}_S = \mathbb{C}^d$. The matrix basis $\{X_1, \ldots, X_m\}$ is taken to be traceless and orthonormal as in Eq. (2.68):

$$\operatorname{tr} X_j = 0, \qquad \operatorname{tr} X_j X_k = \delta_{jk}, \tag{4.1}$$

and satisfies $\sum_{j=1}^{m} X_j^2 = (m/d)I$. Furthermore, the Euclidean norm of the parameter $\boldsymbol{\theta}$ is bounded by the energy scale E:

$$\|\boldsymbol{\theta}\| \leq E. \tag{4.2}$$

The study on the *quantum Fisher inforamtion* (QFI) matrix has revealed that a longer time of quantum evolution leads to a smaller quantum CR bound. However, as the evolution time T become longer, the dependence of the unitary operator $U_{\boldsymbol{\theta}} = e^{-iTH_{\boldsymbol{\theta}}}$ on the parameter $\boldsymbol{\theta}$ becomes chaotic. In fact, given two parameters $\boldsymbol{\theta}, \boldsymbol{\theta}^*$, the difference in the unitary operators

$$\left\| e^{-iTH_{\boldsymbol{\theta}}} - e^{-iTH_{\boldsymbol{\theta}^*}} \right\| = \left\| I - e^{iTH_{\boldsymbol{\theta}}} e^{-iTH_{\boldsymbol{\theta}^*}} \right\| \tag{4.3}$$

becomes unpredictable for an extremely large T — it may be very small or as large as the unity depending on the small variation of $\boldsymbol{\theta}^*$, which makes the estimation infeasible [77, 78]. For the behavior of Eq. (4.3) to be predictable, the evolution time T must be at most of the order of E^{-1} for the energy scale E. However, with the help of the adaptive estimation, one can gradually decrease the effective energy scale E_{eff} based on the previous measurement, which leads to a longer evolution time.

As explained in Sec. 2.3.4, the basic idea is to compute a series of estimators $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_J^*$ one by one, where the estimation error becomes half on each iteration. As the number J of iterations increases, the estimation error decreases as $O(2^{-J})$ while the required amount T of resource increases as $O(2^J)$. The estimation error δ is hence in inverse proportion to the amount T of resource, resulting in the Heisenberg limit $\delta = O(T^{-1})$. The main problem here is to derive a rigorous bound on the *root mean square* (RMS) error, in which we need to take into consideration the failure in early iterations of the adaptive estimation.

Let us consider the estimation of a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$ in a ball-shaped domain $\|\boldsymbol{\theta}\| \leq E$. The *hierarchical estimation* employs a series of stochastic processes that compute the estimator $\boldsymbol{\theta}_j^*$ from the previous one $\boldsymbol{\theta}_{j-1}^*$ $(1 \leq j \leq J)$. Mathematically, this can be described by a Markov chain with the conditional probability distributions $p(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_{j-1}^*)$. This stochastic process must satisfy

$$\|\boldsymbol{\theta}_{j-1}^* - \boldsymbol{\theta}\| \leq E_{j-1} \implies \langle \|\boldsymbol{\theta}_j^* - \boldsymbol{\theta}\|^2 \rangle \leq \left(\frac{1}{4} E_j\right)^2, \tag{4.4}$$

where $E_j = 2^{-j}E$ is the magnitude of the error of $\boldsymbol{\theta}_j^*$. In other words, the estimation error of $\boldsymbol{\theta}_j^*$ is bounded as long as the input $\boldsymbol{\theta}_{j-1}^*$ is sufficiently close to the target parameter $\boldsymbol{\theta}$. In particular, the premise of Eq. (4.4) is automatically satisfied if we set $\boldsymbol{\theta}_0^* = \mathbf{0}$.

We further suppose that each attempt to compute the estimator $\boldsymbol{\theta}_j^*$ requires as much resource as $T_j = 2^{j-1}T$. The goal of the hierarchical estimation is to compute a new estimator $\tilde{\boldsymbol{\theta}}^*$, where the RMS error and the cost are comparable to E_J and T_J, respectively.

We begin by introducing the Chernoff bound [79], which bounds the probability of a tail event[1] from above with an exponentially decreasing quantity in the number of independent trials. In the particular case of a binomial distribution, the Chernoff bound appears as a result of the Chernoff–Hoeffding theorem.

For the rest of this chapter, we use $\mathbf{P}$ to denote the probability of an event.

Theorem 4.1 (Chernoff–Hoeffding theorem [80]) *We consider n independent trials each of which succeeds with probability p, and let k be the number of successful trials. In other words, k is a stochastic variable distributed as the binomial distribution $B(n,p)$. Then, for any $0 < r < p$, we have the following inequality:*

$$\mathbf{P}[k \leq nr] \leq e^{-nD(r,p)}, \tag{4.5}$$

where $D(r,p) = r\log\frac{r}{p} + (1-r)\log\frac{1-r}{1-p}$.

See Ref. [80] for the proof of this theorem. The law of large number ensures that k/n stochastically converges to p in the limit of $n \to \infty$, Theorem 4.1 reveals that only with an exponentially small probability in n may k/n take a value significantly smaller than p.

We use this theorem to bound the probability of the estimation error being significantly large.

Lemma 4.2 *Let $\boldsymbol{\theta}^* \in \mathbb{R}^m$ be an estimator of a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$, whose estimation error is bounded as*

$$\langle \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \rangle \leq a^2. \tag{4.6}$$

Then, for any $s > 1$, there exists another estimator $\tilde{\boldsymbol{\theta}}^*$ that can be obtained from $n = \lceil 7\log s \rceil$ independent instances of $\boldsymbol{\theta}^*$ such that

$$\mathbf{P}\big[\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\| \leq 4a\big] \geq 1 - s^{-1}. \tag{4.7}$$

Here, $\lceil\alpha\rceil$ indicates the minimal integer not less than α.

Proof. Let us say an estimator $\boldsymbol{\theta}^*$ is *successful* if and only if $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq 2a$ holds. Then, the probability of success is bounded by Chebyshev's inequality:

$$\mathbb{P}\big[\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq 2a\big] = \mathbb{P}\big[\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \leq 4a^2\big] \geq 1 - \frac{a^2}{4a^2} = \frac{3}{4}. \tag{4.8}$$

Now, we consider n independent instances $\boldsymbol{\theta}^{*(1)}, \ldots, \boldsymbol{\theta}^{*(n)}$ of the estimator $\boldsymbol{\theta}^*$. Let k be the number of successful estimators among the n independent trials. By Th. 4.1, more than $n/2$ are successful with probability

$$\begin{aligned}\mathbf{P}[k > n/2] &= 1 - \mathbf{P}[k \leq n/2] \\ &\geq 1 - e^{-nD(\frac{1}{2},\frac{3}{4})} = 1 - s^{-1},\end{aligned} \tag{4.9}$$

[1] "Tail event" is a term referring to an event that significantly differs from the expected behavior.

where we note $D(\frac{1}{2}, \frac{3}{4}) = \frac{1}{2}\log\frac{4}{3} \geq \frac{1}{7}$.

We define the new estimator $\tilde{\boldsymbol{\theta}}^*$ such that

$$\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{*(j)}\| \leq 2a \tag{4.10}$$

is satisfied for more than $n/2$ instances of j.

If more than $n/2$ estimators are successful, such $\tilde{\boldsymbol{\theta}}^*$ can be taken; moreover, the new estimator satisfies $\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\| \leq 4a$ since at least one successful estimator $\boldsymbol{\theta}^{*(j)}$ satisfies Eq. (4.10). Therefore, the statement of this lemma follows from Eq. (4.9) and the trigonometric inequality. □

In the hierarchical estimation, the sequence of estimators $\tilde{\boldsymbol{\theta}}_0^*, \tilde{\boldsymbol{\theta}}_1^*, \ldots, \tilde{\boldsymbol{\theta}}_J^*$ are calculated according to the following procedure:

(1) We set $\tilde{\boldsymbol{\theta}}_0^* = \mathbf{0}$.

(2) When $\tilde{\boldsymbol{\theta}}_{j-1}^*$ is calculated, we compute $n_j = \lceil (21\log 2)(J+1-j) \rceil$ independent instances of the estimator $\boldsymbol{\theta}_j^*$ by setting the previous estimator to $\tilde{\boldsymbol{\theta}}_{j-1}^*$.

(3) Following the assumption (4.4) and Lem. 4.2, we compute $\tilde{\theta}_j^*$ from the n_j instances of estimators such that

$$\|\tilde{\boldsymbol{\theta}}_{j-1}^* - \boldsymbol{\theta}\| \leq E_{j-1} \implies \mathbb{P}[\|\tilde{\boldsymbol{\theta}}_j^* - \boldsymbol{\theta}\| \leq E_j] \geq 1 - 8^{-(J+1-j)}. \tag{4.11}$$

(4) After the calculation of $\tilde{\boldsymbol{\theta}}_J^*$, the new estimator $\tilde{\boldsymbol{\theta}}^*$ is computed such that

$$\|\tilde{\boldsymbol{\theta}}^* - \tilde{\boldsymbol{\theta}}_j^*\| \leq E_j \qquad \forall\, j = 0, 1, \ldots, j' \tag{4.12}$$

is satisfied for as large j' as possible.

Now, we estimate the RMS error δ and the total cost of this estimator $\tilde{\boldsymbol{\theta}}^*$. The total cost can be calculated as

$$\begin{aligned} \sum_{j=1}^{J} n_j T_j &\leq \sum_{j=1}^{J} [(21\log 2)(J+1-j)+1] 2^{j-1} T \\ &\leq (42\log 2 + 1) 2^J T \leq 61 T_J. \end{aligned} \tag{4.13}$$

Let us say that the estimation is *successful at the kth iteration* if and only if

$$\|\tilde{\boldsymbol{\theta}}_j^* - \boldsymbol{\theta}\| \leq E_j \qquad \forall\, j = 0, 1, \ldots, k \tag{4.14}$$

is satisfied. We note that the estimation is always successful at the 0th iteration.

Let p_k be the probability of the estimation being *not* successful at the kth iteration. Then, by virtue of Eq. (4.11), the probability with which the kth iteration not being successful can be written as

$$p_k \leq \sum_{j=1}^{k} 8^{-(J+1-j)}. \tag{4.15}$$

Suppose that the estimation is successful at the kth iteration but not at the $(k+1)$th iteration, which occurs with probability $p_{k+1} - p_k$. Then, we find that the number j' in Eq. (4.12) is at least k, since Eq. (4.12) can be satisfied by setting $\tilde{\boldsymbol{\theta}}^* = \tilde{\boldsymbol{\theta}}$. Therefore, we obtain $\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\| \leq 2E_k$ by the trigonometric inequality. Similarly, if the estimation is successful at the Jth (i.e., last) iteration, we have $\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\| \leq 2E_J$. Therefore, the RMS error can be calculated as

$$\begin{aligned}\delta^2 &= \langle \|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\|^2 \rangle \\ &\leq \sum_{k=1}^{J} (p_k - p_{k-1})(2E_{k-1})^2 + (1 - p_J)(2E_J)^2. \\ &= \sum_{k=1}^{J} 4p_k(E_{k-1}^2 - E_k^2) + 4E_J^2. \end{aligned} \tag{4.16}$$

Noting that $E_k < E_{k-1}$ are all positive for $1 \leq k \leq J$, we can substitute Eq. (4.15) in Eq. (4.16):

$$\begin{aligned}\delta^2 &\leq \sum_{k=1}^{J} 4\Bigg(\sum_{j=1}^{k} 8^{-(J+1-j)}\Bigg)(E_{k-1}^2 - E_k^2) + 4E_J^2 \\ &= \sum_{k=1}^{J} 4 \cdot 8^{-(J+1-k)} E_{k-1}^2 + 4\Bigg(1 - \sum_{j=1}^{J} 8^{-(J+1-j)}\Bigg) E_J^2 \\ &\leq 4 \sum_{k=0}^{J} 8^{-(J-k)} E_k^2 = \sum_{k=0}^{J} 2^{2(1-k)-3(J-k)} E^2 \\ &\leq 2^{(3-2J)} E^2 = 8E_J^2. \end{aligned} \tag{4.17}$$

To summarize, the estimation error and the cost of the estimator $\tilde{\boldsymbol{\theta}}^*$ are the same up to a constant factor to those of the last estimator $\boldsymbol{\theta}_J^*$ in the process. Unlike $\boldsymbol{\theta}_J^*$, however, the estimator $\tilde{\boldsymbol{\theta}}^*$ only relies on the initial condition $\|\boldsymbol{\theta}\| \leq E$, which is the advantage of the hierarchical estimation.

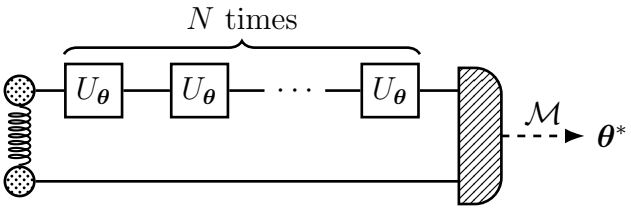
We note that although the Euclidean norm $\|\cdot\|$ is adopted to measure the estimation error, it can be replaced by any metric endowed with the trigonometric inequality. In that sense, our approach provides a general and rigorous framework for adaptive estimation.

4.2 Errors of Hamiltonian estimation methods

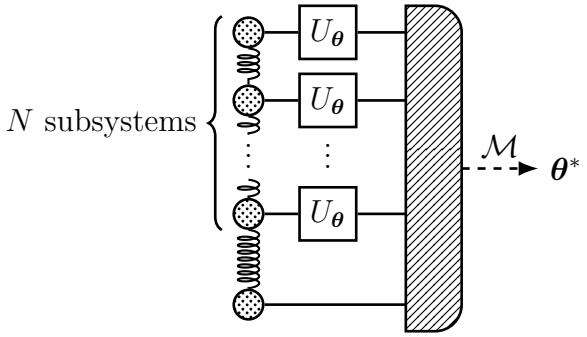
We consider estimation of a Hamiltonian $H_{\boldsymbol{\theta}} = \sum_{j=1}^{m} \theta_j X_j$ in a spherical Hamiltonian model, the definition of which we describe in Sec. 2.3.2. A lower bound on the estimation error is demonstrated in Prop. 2.6, is a lower bound applicable for any estimation method, which can be drawn in the *quantum general scheme*. However, whether or not this error bound can be attained is an open problem, owing to the huge degree of freedom in the general scheme. Hence, we focus on two specific

schemes of the Hamiltonian estimation, within which actual Heisenberg-limited methods have been found in the previous literature.

One of the schemes is the *quantum parallel scheme*, which we have already introduced in Sec. 2.2. The other is the *quantum sequential scheme*, in which the estimation methods consists of state preparations, local operations and quantum measurements. In other words, no quantum memory should be generated except for the preparation of entangled states between the system $\mathcal{H}_{\mathrm{S}}$ and the ancilla $\mathcal{H}_{\mathrm{A}}$.



(a) Sequential scheme



(b) Parallel scheme

Figure 4.1: Schematic images of optimal estimation methods in term of the quantum CR bound. We note that $U_{\boldsymbol{\theta}} = e^{-i\tau H_{\boldsymbol{\theta}}}$ is the unitary evolution over time τ, so that the total time T of the Hamiltonian evolution by $H_{\boldsymbol{\theta}}$ is $T = N\tau$. **(a)** Method in the sequential scheme, where the initial state is the MES of the system Hilbert space $\mathcal{H}_{\mathrm{S}}$. **(b)** Method in the sequential scheme, where the initial state is the MES of the completely symmetric space $\otimes^N_{\mathrm{sym}}\mathcal{H}_{\mathrm{S}}$.

Optimal estimation methods in term of the quantum CR bound δ_{CR} is known in both of the parallel and the sequential schemes, which are as simple as Fig. 4.1. The optimal quantum CR bounds are [30, 59]:

$$\delta_{\mathrm{CR}} = \frac{m^{1/2}d^{1/2}}{N\tau} \qquad \text{(sequential scheme);} \tag{4.18}$$

$$\delta_{\mathrm{CR}} = \frac{m^{1/2}d}{N\tau} \qquad \text{(parallel scheme).} \tag{4.19}$$

Since Eq. (4.18) is $O(d^{1/2})$ times lower than Eq. (4.19), the sequential scheme is seemingly advantageous over the parallel scheme for Hamiltonians in a large Hilbert space.

However, the quantum CR bounds described above cannot be attained for an arbitrarily large N. If we denote by E the energy scale of the Hamiltonian $H_{\boldsymbol{\theta}}$, the attainability of the quantum CR bound relies on the linear approximation $e^{-iN\tau E} \approx 1 - iN\tau E$. Hence, the number r of unitary gates allowed in each measurement is at most $r = O(1/(\tau E))$, meaning that the Heisenberg limit $\delta \propto T^{-1}$ is not attainable when $T \gg r\tau$.

In fact, this problem can be solved by introducing adaptive estimation in the two methods. As explained in Sec. 2.3.4, the benefit of adaptive estimation is intuitively understood as the reduction of the energy scale E, but no rigorous analysis has been performed for multiparameter Hamiltonian estimation. In this thesis, we compute the maximum number r of unitary gates that can be employed in the quantum measurement. Then, by using hierarchical estimation in the previous section, we derive the optimal error of the two methods of Hamiltonian estimation, which is equivalent up to a constant factor independent of the size of the Hamiltonian.

4.2.1 Classical scheme

First, we briefly review the basic idea of the Hamiltonian estimation in the classical scheme, which has already been studied in the author's master thesis. For an operator A, we denote by $\|A\|$ its operator norm and define the *Hilbert–Schmidt norm* by $\|A\|_{\mathrm{HS}} = (\operatorname{tr} A^\dagger A)^{1/2}$. We note that $\|H_{\boldsymbol{\theta}}\|_{\mathrm{HS}} = \|\boldsymbol{\theta}\|$ always holds due to the orthonormal condition (4.1).

Following the unitary gate estimation (Sec 2.3.1), the probe state is set to the MES $|\mathrm{M}\rangle_{\mathcal{H}_S}$ of the system Hilbert space $\mathcal{H}_S$. By using an ancillary Hilbert space $\mathcal{H}_A$, the MES can be written as

$$|\mathrm{M}\rangle_{\mathcal{H}_S} = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |k\rangle_S \otimes |k\rangle_A, \tag{4.20}$$

where $\{|k\rangle_S\}$ and $\{|k\rangle_A\}$ are orthonormal bases of $\mathcal{H}_S$ and $\mathcal{H}_A$. For any operator A acting on $\mathcal{H}_S$, a direct consequence of Eq. (4.20) is

$$\langle \mathrm{M}|A|\mathrm{M}\rangle_{\mathcal{H}_S} = \frac{1}{d} \operatorname{tr} A \tag{4.21}$$

and

$$\big\| A|\mathrm{M}\rangle_{\mathcal{H}_S} \big\|^2 = \frac{1}{d} \operatorname{tr} A^\dagger A = \frac{1}{d} \|A\|_{\mathrm{HS}}^2. \tag{4.22}$$

After the evolution with the Hamiltonian $H_{\boldsymbol{\theta}}$ over time τ, the probe state $|\mathrm{M}\rangle_{\mathcal{H}_S}$ evolves into

$$|q_{\boldsymbol{\theta}}\rangle = e^{-i\tau H_{\boldsymbol{\theta}}} |\mathrm{M}\rangle_{\mathcal{H}_S} \approx (I - i\tau H_{\boldsymbol{\theta}}) |\mathrm{M}\rangle_{\mathcal{H}_S}, \tag{4.23}$$

where the approximation is taken up to the first order in τ. In fact, the approximation error can be bounded from above as

$$\begin{aligned}\| |q_{\boldsymbol{\theta}}\rangle - (I - i\tau H_{\boldsymbol{\theta}})|\mathrm{M}\rangle_{\mathcal{H}_S} \| &= \| (e^{-i\tau H_{\boldsymbol{\theta}}} - I + i\tau H_{\boldsymbol{\theta}})|\mathrm{M}\rangle_{\mathcal{H}_S} \| \\ &= \frac{1}{\sqrt{d}} \| e^{-i\tau H_{\boldsymbol{\theta}}} - I + i\tau H_{\boldsymbol{\theta}} \|_{\mathrm{HS}} \\ &\leq \frac{\tau^2}{2\sqrt{d}} (\operatorname{tr} H_{\boldsymbol{\theta}}^4)^{1/2} \leq \frac{\tau^2 E^2}{2\sqrt{d}} \end{aligned} \tag{4.24}$$

and hence is negligible when $\alpha = \tau E$ is sufficiently small. For the rest of this section, the number α should be a sufficiently small number that does not depend on the other variables such as m and d.

Let $\mathcal{H}'_S \subset \mathcal{H}_S \otimes \mathcal{H}_A$ be the $(m+1)$-dimensional subspace spanned by

$$|\mathrm{M}\rangle_{\mathcal{H}_S}, X_1|\mathrm{M}\rangle_{\mathcal{H}_S}, \ldots, X_m|\mathrm{M}\rangle_{\mathcal{H}_S}. \tag{4.25}$$

Then, the right-hand side of Eq. (4.23) belongs to $\mathcal{H}'_S$ regardless of the parameter $\boldsymbol{\theta}$. Hence, we consider the projection measurement $\{P, 1-P\}$ with the projection operator P onto $\mathcal{H}'_S$, and discard the probe state corresponding to $1-P$. This postselection yields the quantum state $|\tilde{q}_{\boldsymbol{\theta}}\rangle \propto P|q_{\boldsymbol{\theta}}\rangle$, which is close to $(I - i\tau H_{\boldsymbol{\theta}})|\mathrm{M}\rangle_{\mathcal{H}_S}$ and belongs to the $(m+1)$-dimensional Hilbert space $\mathcal{H}'_S$. We note that the fraction of discarded probe states is negligible.

Therefore, we expect that the parameter $\boldsymbol{\theta}$ can be accurately estimated by the *quantum state tomography* (QST) on the postselected state $|\tilde{q}_{\boldsymbol{\theta}}\rangle$. In fact, the following proposition has been proved in the master thesis of the present author [27]:

Proposition 4.3 *When $\alpha = \tau E$ is a sufficiently small, there exists $c_1 > 0$ such that the quantum infidelity between two quantum states $|\tilde{q}_{\boldsymbol{\theta}^*}\rangle$ and $|\tilde{q}_{\boldsymbol{\theta}}\rangle$ can be bounded from below as*

$$1 - \left|\langle \tilde{q}_{\boldsymbol{\theta}^*} | \tilde{q}_{\boldsymbol{\theta}} \rangle\right|^2 \geq \frac{c_1 \tau^2}{d} \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2. \tag{4.26}$$

We note that this proposition relies on the fact that the approximation error in Eq. (4.24) is negligible. Although the proof involves a series of complicated analytical calculations, the bound given in Prop. (4.26) is equivalent to the quantum CR bound, indicating the optimality. Here is a more refined version of Prop. 4.3:

Proposition 4.4 *Suppose that three constants $0 < c_1 < 1$, $c_2 \geq 1$, and $c_3 \geq 1$ are given. Then, there exists a constant $\alpha > 0$ such that following statement holds: Let $0 < b \leq c_3$, and $\{H_{\boldsymbol{\theta}}\}$ be a Hamiltonian model in d-dimensional Hilbert space such that*

$$\tfrac{1}{d} \operatorname{tr} H_{\boldsymbol{\theta}}^2 = b^2 \|\boldsymbol{\theta}\|^2, \qquad \tfrac{1}{d} \operatorname{tr} H_{\boldsymbol{\theta}}^4 \leq c_2^2 b^2 \|\boldsymbol{\theta}\|^4. \tag{4.27}$$

Then, the postselected state $|\tilde{q}_{\boldsymbol{\theta}}\rangle$ satisfies

$$1 - \left|\langle \tilde{q}_{\boldsymbol{\theta}^*} | \tilde{q}_{\boldsymbol{\theta}} \rangle\right|^2 \geq c_1 b^2 \tau^2 \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \tag{4.28}$$

for arbitrary $\|\boldsymbol{\theta}^*\|, \|\boldsymbol{\theta}\| \leq \alpha/\tau$.

Proposition 4.3 is the special case where $b = d^{-1/2}$. Owing to this proposition, the RMS error can be bounded from above by Eq. (2.33)[2]. With n copies of such probe states, we obtain

$$\begin{aligned}\delta^2 &= \langle \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \rangle \\ &\leq \frac{d}{c_1 \tau^2} \left\langle 1 - \left| \langle \tilde{q}_{\boldsymbol{\theta}^*} | \tilde{q}_{\boldsymbol{\theta}} \rangle \right|^2 \right\rangle \\ &= \frac{d}{c_1 \tau^2} \frac{(m+1)-1}{(m+1)+n} \leq \frac{md}{c_1 \tau^2 n},\end{aligned} \tag{4.29}$$

where we note that $m+1$ is the dimension of the Hilbert space $\mathcal{H}'_S$.

Under the constraint by the assumption in Prop. 4.3 and the total resource $T = n\tau$, the error is minimized when we set $\tau = \alpha E^{-1}$. Consequently, the estimation error exhibits the *standard quantum limit* (SQL) $\delta = O(T^{-1/2})$:

$$\delta \leq \left(\frac{mdE}{c_1 \alpha T} \right)^{1/2}. \tag{4.30}$$

4.2.2 Quantum sequential scheme

In Fig. 4.2, we show how to introduce adaptive feedback in the optimal estimation method (Fig. 4.1 (a)) in the sequential scheme.

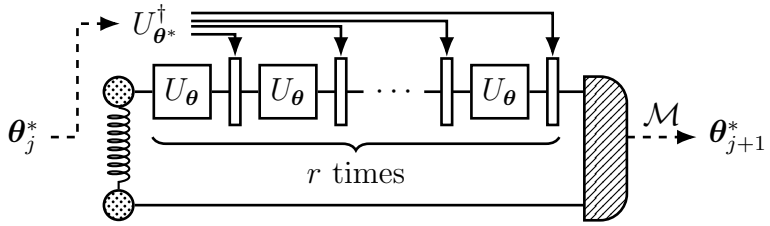


Figure 4.2: Schematic image for the adaptive estimation in the sequential quantum scheme. During the computation of the next estimator $\boldsymbol{\theta}^*_{j+1}$, each unitary evolution $U_{\boldsymbol{\theta}}$ is followed by a control gate $U^\dagger_{\boldsymbol{\theta}^*}$ based on the previous estimator $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*_j$.

Suppose that we have obtained an estimator $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*_j$ such that

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq E_j = 2^{-j} E \tag{4.31}$$

holds with a high probability. In the next iteration, we repeatedly apply the unitary evolution given by $U_{\boldsymbol{\theta}} = e^{-i\tau H_{\boldsymbol{\theta}}}$ for r times. Furthermore, each unitary evolution is followed by a control unitary gate $U^\dagger_{\boldsymbol{\theta}^*}$. Since the control operation is

[2]The equality in Eq. (2.33) can be satisfied by the optimal QST, which we adopt here.

local to the system $\mathcal{H}_{\mathrm{S}}$ and based only on the outcome of the previous measurements, this method is free of quantum memory and still in the quantum sequential scheme.

Now, we compute the maximum number r of subsystems such that the quantum CR bound in Prop. 4.4 can be attained. If we define the Hermitian operator $K_{\boldsymbol{\theta}}$ such that

$$U_{\boldsymbol{\theta}^*}^\dagger U_{\boldsymbol{\theta}} = \exp(i\tau H_{\boldsymbol{\theta}^*}) \exp(-i\tau H_{\boldsymbol{\theta}}) = \exp(-i\tau K_{\boldsymbol{\theta}}), \tag{4.32}$$

The entire evolution can be regarded as the Hamiltonian evolution by $K_{\boldsymbol{\theta}}$ over time $r\tau$:

$$(U_{\boldsymbol{\theta}^*}^\dagger U_{\boldsymbol{\theta}})^r = \exp(-ir\tau K_{\boldsymbol{\theta}}). \tag{4.33}$$

Such $K_{\boldsymbol{\theta}}$ is known as the Floquet Hamiltonian, and can be approximated as

$$K_{\boldsymbol{\theta}} = H_{\boldsymbol{\theta}} - H_{\boldsymbol{\theta}^*} + O(\tau) = H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*} + O(\tau) \tag{4.34}$$

for small τ. The residual term can be evaluated by the Baker–Campbell–Hausdorff formula: For small operators X, Y such that $\|X\|, \|Y\| \leq \frac{1}{6}$, there exists an operator Z satisfying $e^X e^Y = e^Z$ and

$$\big\|Z - (X+Y)\big\|_{\mathrm{HS}} \leq c_{\mathrm{B}} \big\|[X,Y]\big\|_{\mathrm{HS}}, \tag{4.35}$$

where the constant factor can be set to $c_{\mathrm{B}} = 2.24\ldots$. We may assume $\alpha \leq \frac{1}{6}$ without loss of generality, in which case we may substitute $X = \tau H_{\boldsymbol{\theta}}, Y = -\tau H_{\boldsymbol{\theta}^*}$ to obtain

$$\begin{aligned} \|K_{\boldsymbol{\theta}} - H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}\|_{\mathrm{HS}} &\leq c_{\mathrm{B}}\tau \big\|[H_{\boldsymbol{\theta}}, H_{\boldsymbol{\theta}^*}]\big\|_{\mathrm{HS}} = c_{\mathrm{B}}\tau \big\|[H_{\boldsymbol{\theta}}, H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}]\big\|_{\mathrm{HS}} \\ &\leq 2c_{\mathrm{B}}\tau E E_j = 2c_{\mathrm{B}}\alpha E_j, \end{aligned} \tag{4.36}$$

where we have used Eq. (4.31). Therefore, the Floquet Hamiltonian $K_{\boldsymbol{\theta}}$ can be approximated by $H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}$ when $\alpha = \tau E$ is sufficiently small. By Eqs. (4.24) and (4.36), the approximation error in the sequential scheme can be bounded as

$$\begin{aligned} &\big\||q_{\boldsymbol{\theta}}\rangle - (I - ir\tau H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*})|\mathrm{M}\rangle_{\mathcal{H}_S}\big\| = \frac{1}{\sqrt{d}} \big\|e^{-ir\tau K_{\boldsymbol{\theta}}} - I + ir\tau H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}\big\|_{\mathrm{HS}} \\ &\leq \frac{1}{\sqrt{d}} \Big(\big\|e^{-ir\tau K_{\boldsymbol{\theta}}} - I + ir\tau K_{\boldsymbol{\theta}^*}\big\|_{\mathrm{HS}} + r\tau \|K_{\boldsymbol{\theta}} - H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}\|_{\mathrm{HS}} \Big) \\ &\leq \frac{1}{2\sqrt{d}} \Big[r^2\tau^2 (\operatorname{tr} K_{\boldsymbol{\theta}}^4)^{1/2} + 2c_{\mathrm{B}} r\tau^2 \big(\operatorname{tr}(i[H_{\boldsymbol{\theta}}, H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}])^2\big)^{1/2} \Big] \qquad (4.37) \\ &\leq \frac{1}{2\sqrt{d}} (r^2\tau^2 E_j^2 + 4c_{\mathrm{B}} r\tau^2 E E_j) = \frac{\alpha^2}{2\sqrt{d}} (2^{-2j} r^2 + 4c_{\mathrm{B}} 2^{-j} r). \qquad (4.38) \end{aligned}$$

Therefore, Prop. 4.4 can be applied as long as the number of unitary gate r is at most comparable to 2^j.

We set $r = c_4 \cdot 2^j$ with a constant c_4, which depends on c_3 and c_{B}. Then, an upper bound on the estimation error is given by Eq. (4.29) with τ replaced by $r\tau$:

$$\delta^2 \leq \frac{md}{c_1 (r\tau)^2 n} = \frac{md}{c_1 c_4^2 2^{2j} \tau^2 n}. \tag{4.39}$$

By setting $n = \frac{64md}{c_1 c_4^2 \alpha^2}$, the estimation error can be further bounded by

$$\delta^2 \leq \frac{md}{c_1 c_4^2 2^{2j} \tau^2 n} \leq 2^{-2j-6} E^2 = \left(\frac{1}{4} E_{j+1}\right)^2, \tag{4.40}$$

implying that the adaptive estimation satisfies Eq. (4.4). The total resource required for this iteration is

$$T_{j+1} = nr\tau = 2^j \cdot \frac{64md}{c_1 c_4 \alpha E}. \tag{4.41}$$

By the hierarchical estimation, the estimation error $\delta = \sqrt{8} E_J = \sqrt{8} \cdot 2^{-J} E$ can be attained with resource at most $T = 61 T_J = 61 \cdot 2^J \frac{64md}{c_1 c_4 \alpha E} = \frac{3904\sqrt{8}}{c_1 c_4 \alpha} md\delta^{-1}$. Therefore, we obtain the Heisenberg limit

$$\delta = \frac{3904\sqrt{8}}{c_1 c_4 \alpha} mdT^{-1} = O(mdT^{-1}), \tag{4.42}$$

which is consistent with the result obtained in the master thesis of the present author.

4.2.3 Quantum parallel scheme

A rigorous error bound for the optimal method in the quantum parallel scheme is given in this thesis for the first time. The main obstacle is to derive the error bound on the approximation similar to Eq. (4.23), which is impossible without an intricate investigation into the extended Hilbert space. First, we consider the case without adaptive estimation (or equivalently, the first step of the adaptive estimation), and compute maximum number r of unitary gates such that the quantum CR bound can be attained. After that, we introduce adaptive estimation and derive the estimation error.

We recall that, in the parallel scheme, each subsystem undergoes unitary evolution with $U_{\boldsymbol{\theta}} = e^{-i\tau H_{\boldsymbol{\theta}}}$. The unitary evolution in the entire system is

$$U_\theta^{\otimes r} = (e^{-i\tau H_{\boldsymbol{\theta}}})^{\otimes r} = e^{-i\tau \{H_{\boldsymbol{\theta}}\}_r}, \tag{4.43}$$

$$\{A\}_r = \sum_{i=1}^{r} I \otimes \cdots \otimes \underbrace{A}_{i\text{th}} \otimes \cdots \otimes I. \tag{4.44}$$

Analogously to the unitary gate estimation (Sec 2.3.1), the Heisenberg limit can be attained when we set the probe state to be the MES $|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}}$ in the completely symmetric space $\mathcal{H}_{\mathrm{sym}} \subset \mathcal{H}_S^{\otimes r}$. In this case, postselection is performed onto the $(m+1)$-dimensional subspace

$$|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}}, \{X_1\}_r |\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}}, \ldots \{X_m\}_r |\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}}. \tag{4.45}$$

We have seen in Sec 2.3.1 that the QFI matrix is $\frac{r(r+d)}{1+d}$ times larger than that in the classical scheme. In accordance with this fact, we must set $b^2 = \frac{r(r+d)}{1+d}$ in Prop. 4.4 so that

$$\frac{1}{D}\operatorname{tr}[P_{\mathrm{sym}}\{H_{\boldsymbol{\theta}}\}_r^2] = b^2\|\boldsymbol{\theta}\|^2. \tag{4.46}$$

In this case, the error bound in Eq. (4.29) is replaced by

$$\delta^2 \leq \frac{m}{c_1\tau^2 b^2 n} = \frac{md(1+d)}{c_1\tau^2 r(r+d)n}. \tag{4.47}$$

Therefore, we need to choose a large number as r in order to reduce this error bound. The question is the maximal number r of subsystems that does not break the approximation

$$|q_{\boldsymbol{\theta}}\rangle := e^{-i\tau\{H_{\boldsymbol{\theta}}\}_r}|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}} \approx (I - i\tau\{H_{\boldsymbol{\theta}}\}_r)|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}}. \tag{4.48}$$

In fact, the approximation (4.48) can be bounded in a way similar to Eq. (4.24):

$$\big\||q_{\boldsymbol{\theta}}\rangle - (I - i\tau\{H_{\boldsymbol{\theta}}\}_r)|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}}\big\| \leq \frac{\tau^2}{2\sqrt{D}}\big(\operatorname{tr}[P_{\mathrm{sym}}\{H_{\boldsymbol{\theta}}\}_r^4]\big)^{1/2}, \tag{4.49}$$

where $D = \frac{(r+d-1)!}{r!(d-1)!}$ is the dimension of $\mathcal{H}_{\mathrm{sym}}$ and P_{sym} is the projection operator onto $\mathcal{H}_{\mathrm{sym}}$. The right-hand side of Eq. (4.49) can be bounded by the following proposition:

Proposition 4.5 *For an arbitrary traceless operator X, the following identity holds:*

$$\begin{aligned}\frac{1}{D}\operatorname{tr}[P_{\mathrm{sym}}\{X\}_r^4] = {}& \frac{r(r+d)(6r^2+6dr+d^2-d)}{d(d+1)(d+2)(d+3)}\operatorname{tr}X^4 \\ & + \frac{3r(r+d)(r-1)(d+r+1)}{d(d+1)(d+2)(d+3)}(\operatorname{tr}X^2)^2.\end{aligned} \tag{4.50}$$

In fact, we have found an algorithm to calculate

$$\operatorname{tr}[P_{\mathrm{sym}}\{X\}_r^m] \tag{4.51}$$

in terms of $\operatorname{tr}X, \ldots, \operatorname{tr}X^m$ for general m, where X is not necessarily traceless. The derivation of this algorithm uses the grand partition function of a free boson as the generating function of Eq. (4.51). To avoid a digression from the main argument, we refer to Appendix A for the detail.

We focus on the fact that the coefficients on the right-hand side of Eq. (4.50) is both $O((r/d)^4)$ for $r \geq d$. In fact, we have

$$\frac{1}{D}\operatorname{tr}[P_{\mathrm{sym}}\{H_{\boldsymbol{\theta}}\}_r^4] \leq \frac{9}{16}\left(1+\frac{2r}{d}\right)^4\|\boldsymbol{\theta}\|^4. \tag{4.52}$$

Therefore, the approximation error can be bounded as

$$\| |q_{\boldsymbol{\theta}}\rangle - (I - i\tau\{H_{\boldsymbol{\theta}}\}_r)|\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}} \| \leq \frac{3}{8}\left(1 + \frac{2r}{d}\right)^2 \tau^2 E^2, \tag{4.53}$$

whence we can choose $r = O(d)$ for the number of subsystems.

Now, we introduce the adaptive estimation in the parallel scheme, the schematics of which is shown in Fig. 4.3.

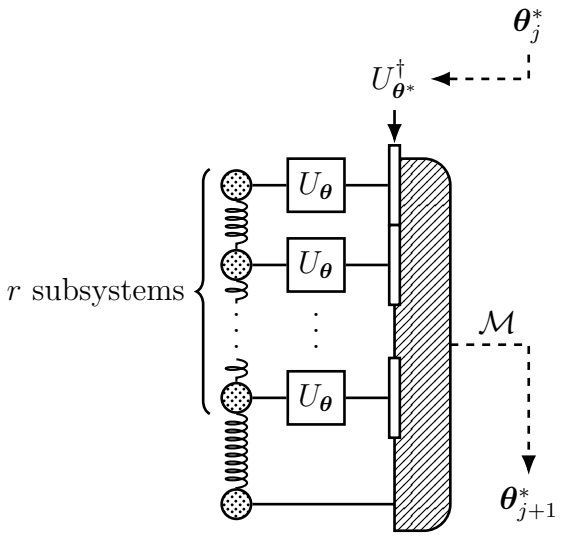


Figure 4.3: Schematics for the adaptive estimation in the parallel quantum scheme. The control gates $U_{\boldsymbol{\theta}^*}^\dagger$ are applied only once per each subsystem, which can be regarded as a part of the measurement process. The control gates on the $(j+1)$th sector depends on the previous estimator $\boldsymbol{\theta}^* = \boldsymbol{\theta}_j^*$.

Suppose that we have obtained an estimator $\boldsymbol{\theta}^* = \boldsymbol{\theta}_j^*$ such that

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq E_j = 2^{-j}E \tag{4.54}$$

holds with a high probability.

In the next iteration, the unitary evolution $(e^{-i\tau H_{\boldsymbol{\theta}}})^{\otimes r}$ is followed by a control unitary gate $(e^{i\tau H_{\boldsymbol{\theta}^*}})^{\otimes r}$ right before the measurement. As described in Fig. 4.3, the control unitary gate can be regarded as a part of the quantum measurement. Moreover, the evolution by the Hamiltonian $H_{\boldsymbol{\theta}}$ can be performed before any of the measurement. Therefore, this estimation method is still in the quantum parallel scheme without any feedback protocol, despite the measurement itself being adaptive.

In this case, the unitary evolution can also be described by the same Floquet Hamiltonian $K_{\boldsymbol{\theta}}$ as in (4.32):

$$|q_{\boldsymbol{\theta}}\rangle = e^{i\tau\{H_{\boldsymbol{\theta}^*}\}_r} e^{-i\tau\{H_{\boldsymbol{\theta}}\}_r} |\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}} = e^{-i\tau\{K_{\boldsymbol{\theta}}\}_r} |\mathrm{M}\rangle_{\mathcal{H}_{\mathrm{sym}}}. \tag{4.55}$$

The approximation error in the adaptive estimation can be evaluated in a way similar to Eq. (4.37):

$$\begin{aligned}
&\big\| |q_{\boldsymbol{\theta}}\rangle - (I - i\tau\{H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}\}_r)|\mathrm{M}\rangle_{\mathcal{H}_S} \big\| \\
&\leq \frac{1}{2\sqrt{D}}\Big[\tau^2(\operatorname{tr} P_{\mathrm{sym}}\{K_{\boldsymbol{\theta}}\}_r^4)^{1/2} + 2c_{\mathrm{B}} r\tau^2\big(\operatorname{tr} P_{\mathrm{sym}}\{i[H_{\boldsymbol{\theta}}, H_{\boldsymbol{\theta}-\boldsymbol{\theta}^*}]\}_r^2\big)^{1/2}\Big] \qquad (4.56) \\
&\leq \frac{3}{8}\left(1+\frac{2r}{d}\right)^2\tau^2 E_j^2 + 2c_{\mathrm{B}}\left(1+\frac{2r}{d}\right)\tau^2 E E_j \\
&= \alpha^2\left[\frac{3}{8}\left(1+\frac{2r}{d}\right)^2 2^{-2j} r^2 + 2c_{\mathrm{B}}\left(1+\frac{2r}{d}\right)2^{-j} r\right], \qquad (4.57)
\end{aligned}$$

whence we may choose $r = c_5 \cdot 2^j d$ for some constant $c_5 > 0$. The estimation error in Eq. (4.47) then becomes

$$\delta^2 \leq \frac{md(1+d)}{c_1\tau^2 r(r+d)n} = \frac{m}{c_1 c_5^2 \tau^2 2^{2j} n}, \qquad (4.58)$$

which becomes $\left(\frac{1}{4}E_{j+1}\right)^2$ if we set $n = \frac{64m}{c_1 c_5^2 \alpha^2}$. Finally, the total cost required in this iteration is

$$T_{j+1} = nr\tau = \frac{64m}{c_1 c_5^2 \alpha^2} \cdot c_5 2^j d \cdot \tau = 2^j \cdot \frac{64md}{c_1 c_5 \alpha E}, \qquad (4.59)$$

which is equivalent to (4.41) up to a constant factor. Therefore, with the hierarchical estimation, the estimation error in the parallel scheme reaches the same Heisenberg limit as the sequential scheme:

$$\delta = \frac{3904\sqrt{8}}{c_1 c_5 \alpha} mdT^{-1} = O(mdT^{-1}). \qquad (4.60)$$

The crucial point of the parallel scheme is that the number r of unitary gates in the quantum scheme involves a factor $O(d)$, which does not appear in the sequential scheme. This marks an essential distinction of our work from a preceding study by Yuan [30], in which a common number r of unitary gates are considered in comparing these two schemes. Yuan's analysis results in the advantage of sequential scheme by an $O(d^{-1/2})$ factor of estimation error.

We, on the other hand, consider a common energy scale E between both schemes, which appears to be a more natural situation. This assumption requires an analysis beyond the simple CR bound, which we have resolved by the adaptive estimation with a decreasing energy scale E_j. By calculating the optimal number r of unitary gates regarding the energy scale E_j, we have found that the parallel scheme can make use of more unitary gates than the sequential scheme. Finally, the advantages of both schemes in different aspects compensate each other, resulting in the same estimation error up to a constant factor.

Chapter 5

Quantum Metrology for Function Estimation

In this chapter, we provide a fundamental framework for quantum estimation of a continuously varying quantity, namely a function. As the first attempt to establish function estimation in quantum metrology, we work on a simple model of *phase function estimation*, which we introduce in Sec. 5.1. It is a minimal yet fundamental model in which the phase shift $\varphi(x)$ depends on the position x in a one-dimensional interval $[0, L]$. We define the regularity of the function $\varphi(x)$ by a mean-square variant of the Hölder continuity of order q, and the estimation error is also measured by the mean-square error (MSE).

The error of the function estimation is calculated in the three situations. The first two situations use specific estimation methods (Sec. 5.2), while the last situation does not specify the method (Sec. 5.3):

- Estimation error for a method that we call the *position-state* (PS) method, in which the probe state is localized in position. In this method, we consider all regularity ($q > 0$), although periodicity of the phase must be taken in consideration for $q > 1$; in the latter case, we introduce additional regularity (a uniform upper bound on $\varphi'(x)$) and use the mean-square periodic error (MSPE) as the error measure, .

- Estimation error for a method that we call the *wavenumber-state* (WS) method, in which the probe state is localized in wavenumber. In this method, we only consider regularity with $0 < q \leq 1$, because this method is not applicable for $q \geq 1$. We also use the mean-square periodic error (MSPE) as the error measure.

- Lower bound on the estimation error independent of the estimation method. In the quantum scheme, however, we assume some additional assumption on the initial probe state. Instead of arbitrarily entangled N particles, we assume that n_1 groups of n_2 entangled particles in a unique symmetric sector. We hypothesize that the Heisenberg limit is unaltered by clearing this assumption.

In any case of the above, all yields the same scaling laws: $\delta = O\big((LM^{1/q}N^{-1})^{q/(2q+1)}\big)$ without entanglement and $\delta = O\big((LM^{1/q}N^{-1})^{q/(q+1)}\big)$ with entanglement.

In Sec. 5.4, we consider two types of extension of our model. One is the estimation of functions over an infinite-length interval, where we find the relation to

the previous studies on Gaussian process estimation. The other is the estimation of multivariate functions, where a hypothesis on the error bounds is presented.

5.1 Phase function estimation

5.1.1 Physical system and resource

We adopt a phase-shift gate as our target of estimation, which respects the minimal model of quantum metrology for parameter estimation. Here we consider a situation in which the phase $\varphi(x)$ varies with the position x.

In this setting, the physical system must be interpreted as a system of particles. If a particle at position x, the unitary dynamics U_φ shifts its phase by $\varphi(x)$. In the Hilbert space of one particle, let us denote by $|x;m\rangle$ the simultaneous eigenstate of position x and internal state m, where the inner product of two eigenstates is defined as

$$\langle x;m|x';m'\rangle = \delta(x-x')\delta_{mm'}. \tag{5.1}$$

We note that $|x;m\rangle$ is not an physical state since it is unnormalizable. Informally, the norm $|x;m\rangle$ can be regarded as a vector of length $\frac{1}{\sqrt{dx}}$, where dx represents the infinitesimal length. Furthermore, we consider two internal states[1] "+" and "−". The particles in the + state are phase-shifted by the unitary dynamics, while those in the − state remain unchanged. Then, the action of the unitary dynamics U_φ can formally be written as

$$U_\varphi|x;+\rangle = e^{i\varphi(x)}|x;+\rangle, \qquad U_\varphi|x;-\rangle = |x;-\rangle. \tag{5.2}$$

The resource of quantum metrology is the number N of particles that are used for the probe. The *standard quantum limit* (SQL) refers to the fundamental bound on the estimation error in the *classical scheme*, which precludes entanglement between the particles. We allow interparticle entanglement in the *quantum scheme*, on the other hand, which will lead to the *Heisenberg limit* on the estimation error.

To simplify discussions, we consider a one-dimensional system confined in a finite interval $0 \le x \le L$ with the boundary condition $\varphi(x+L) = \varphi(x)$. In Sec. 5.4, we briefly discuss phase function estimation in other types of systems such as the infinite length limit $L \to \infty$ and systems with higher dimensions.

5.1.2 Error measure and regularity

As we have explained in Sec. 3.1.1, function estimation involves the choice of an appropriate regularity. This is also true even with quantum metrology. In fact, function estimation involves infinite degrees of freedom, which causes the quantum Cramér–Rao (CR) bound to diverge when seen as a multiparameter estimation problem.

[1] We do not exclude the possibility of other degrees of freedom about the particle. For example, the exchange symmetry of the particles may or may not exist.

We also note that there are multiple ways to measure an estimation error. Hence, we must choose the regularity and the error measure that are appropriate for quantum metrology.

It is convenient to introduce some function norms. For example, if we employ the *uniform norm* $\|\cdot\|_\infty$ defined as[2]

$$\|u\|_\infty = \sup_x |u(x)|, \tag{5.3}$$

the statement of Th. 3.2 can be written as follows:

$$\|f^{(m)}\|_\infty \leq M \implies \|f^* - f\|_\infty \leq Ch^m M. \tag{5.4}$$

In this formula, the premise (left of $\Rightarrow$) represents the regularity and the conclusion (right of $\Rightarrow$) provides an error bound.

In quantum metrology, on the other hand, it is more suitable to use the L^2 norm $\|\cdot\|_2$:

$$\|u\|_2 = \left[\int_0^L \frac{dx}{L} |u(x)|^2\right]^{1/2}. \tag{5.5}$$

In fact, we find that L^2 norm is compatible to the Euclidean distance in multi-parameter estimation, and therefore we can employ the quantum CR inequality to give a suitable error bounds. Moreover, an L^2 norm naturally appears as the metrics of a Hilbert space, which is directly connected to the quantum mechanics.

A candidate of the error measure is therefore the *mean-square error*:

$$\mathrm{MSE}(\varphi^*, \varphi) = \|\varphi^* - \varphi\|_2^2 = \int_0^L \frac{dx}{L} |\varphi^*(x) - \varphi(x)|^2. \tag{5.6}$$

Finally, the estimation error δ is given as the stochastic average over the estimator φ^*:

$$\delta^2 = \langle \mathrm{MSE}(\varphi^*, \varphi) \rangle. \tag{5.7}$$

Since the error is measured by using the L^2 norm, the regularity of the function should also be described by using the L^2 norm of the mth derivative (see Eq. (3.5)):

$$\|\varphi^{(m)}\|_2 \leq M. \tag{5.8}$$

As we have explained in Sec. 3.1.3, it is more desirable to take a broad class of functions in consideration. In that sense, we generalize Eq. (5.8) to an L^2 counterpart of the class $C^{m,\sigma}$ regularity (Def. 3.4), so that the degrees of smoothness is continuously parametrized by $q = m + \sigma$. The definition of the σ-Hölder continuity in Def. 3.3 can be rewritten as

$$\sup_{\varepsilon \neq 0} \frac{\|\Delta_\varepsilon f\|_\infty}{|\varepsilon|^\sigma} \leq M, \tag{5.9}$$

[2]More strictly, we need to ignore values in a set of measure zero when taking the supremum in Eq. (5.3), which is called the essential supremum.

where we have employed a *finite difference* denoted by Δ_ε:

$$\Delta_\varepsilon f(x) = f(x+\varepsilon) - f(x). \tag{5.10}$$

Therefore, the L^2 counterpart of the Hölder continuity is given as follows:

Definition 5.1 A function f is said to be *weakly σ-Hölder continuous* if there exists some constant $M \geq 0$ such that

$$\sup_{\varepsilon \neq 0} \frac{\|\Delta_\varepsilon f\|_2}{|\varepsilon|^\sigma} \leq M, \tag{5.11}$$

or more explicitly,

$$\sup_{\varepsilon \neq 0} \int_0^L \frac{dx}{L} \frac{|f(x+\varepsilon) - f(x)|^2}{|\varepsilon|^{2\sigma}} \leq M^2. \tag{5.12}$$

The minimum of such M is called the *weak σ-Hölder norm* of f.

Using the definition above, we define the L^2 counterpart of the class $C^{m,\sigma}$ regularity as follows:

Definition 5.2 Let us set $q = m + \sigma$ with an integer $m \geq 0$ and a number $0 < \sigma \leq 1$. When a function φ is m-time differentiable, and the mth derivative $\varphi^{(m)}$ is weakly σ-Hölder continuous with norm M, we say that the function φ is *(q, M)-regular*. In other words, the function φ is (q, M)-regular if and only if

$$\sup_{\varepsilon \neq 0} \int_0^L \frac{dx}{L} \frac{|\varphi^{(m)}(x+\varepsilon) - f^{(m)}(x)|^2}{|\varepsilon|^{2\sigma}} \leq M^2. \tag{5.13}$$

Although the (q, M)-regularity is complicated in appearance, it is indeed suitable to handle function estimation in quantum metrology. To summarize, the goal of the function phase estimation is to evaluate the estimation error defined in (5.7) under the assumption of the (q, M)-regularity.

5.2 Methods of function estimation

Before entering the general theory, let us begin with specific methods of phase function estimation. We present two methods with significantly different process of estimation. We call them the *position-state method* and the *wavenumber-state method*, since the probe states used in these methods are localized in position and wavenumber, respectively. Moreover, the position-state method employs the *local linear smoothing* (LLS), a classical algorithm introduced in Sec. 3.1, while the wavenumber-state method uses the *quantum state tomography* (QST), the quantum estimation technique introduced in Sec. 2.1.4. Despite these significant differences, the error bounds become equivalent up to a constant factor between both methods, may it be in the classical scheme or quantum one.

5.2.1 Position-state method

The first method uses the quantum version of the LLS. We take n_1 sample points $0 \leq x_1 < x_2 < \cdots < x_{n_1} < L$ and estimate each phase $\varphi(x_j)$ by using n_2 particles, so that $N = n_1 n_2$ particles are used in total. These sample points are distributed with the equal spacing $h = L/n_1$.

We call this method the *position-state method* because the probe states are localized in the position:

$$|q\rangle \propto \frac{1}{\sqrt{2}}(|x_j;+\rangle + |x_j;-\rangle). \tag{5.14}$$

We note that the completely localized state in position written in Eq. (5.14) is not a physical state. Rather, a physical quantum state is approximately described by a linear superposition state over a finite width $l > 0$, say,

$$|q\rangle = \frac{1}{\sqrt{2l}} \int_{x_i - l/2}^{x_i + l/2} dx(|x;+\rangle + |x;-\rangle). \tag{5.15}$$

With sufficiently small $l \ll h$, we may assume that the probe state is given by Eq. (5.14), which can be used to measure the phase $\theta_j = \varphi(x_j)$ at the sample point x_j. Let us assume n_2 separable particles in this probe state, and θ_j^* be the optimal unbiased estimator for θ_j. Then, the problem is equivalent to the estimation of a single phase, where the estimation error is given by the SQL in Sec. 2.2.1:

$$\langle |\theta_j^* - \theta_j|^2 \rangle = {n_2}^{-1}. \tag{5.16}$$

We consider an LLS kernel β of degree m with support s. Regarding the periodic boundary condition of $[0, L]$, the estimator φ^* for the entire function is calculated as

$$\varphi^*(x) = \sum_{j=1}^{n_1} \theta_j^* \sum_{d \in \mathbb{Z}} \beta\left(\frac{x - x_j + Ld}{h}\right). \tag{5.17}$$

We note that the summation over d is essentially finite. In fact, when $n_1 > s$, the value $\beta(\frac{x-x_j+Ld}{h})$ becomes nonzero for at most one instance of $d \in \mathbb{Z}$ per every j. The stochastic average of the estimator φ^* is

$$\varphi_{\mathrm{LLS}}(x) = \sum_{j=1}^{n_1} \theta_j \sum_{d \in \mathbb{Z}} \beta\left(\frac{x - x_j + Ld}{h}\right). \tag{5.18}$$

Then, by a straightforward calculation, we can decompose the squared error δ^2 into two terms:

$$\delta^2 = \langle \mathrm{MSE}(\varphi^*, \varphi) \rangle = \delta_{\mathrm{det}}^2 + \delta_{\mathrm{stat}}^2, \tag{5.19}$$

$$\delta_{\mathrm{det}}^2 := \mathrm{MSE}(\varphi_{\mathrm{LLS}}, \varphi), \quad \delta_{\mathrm{stat}}^2 := \langle \mathrm{MSE}(\varphi^*, \varphi_{\mathrm{LLS}}) \rangle, \tag{5.20}$$

where $\delta^2_{\rm det}$ originates from the LLS and $\delta^2_{\rm stat}$ is due to the stochastic nature of the measurement. In this sense, Eq. (5.19) is regarded as a decomposition into the bias and the variance similar to Eq. (2.5).

First, we calculate the statistical term $\delta^2_{\rm stat}$. As long as $n_1 > s$, this term is found to be independent of n_1. In fact,

$$\delta^2_{\rm stat} = \langle {\rm MSE}(\varphi^*, \varphi_{\rm LLS})\rangle = \left\langle \int_0^L \frac{dx}{L} |\varphi^*(x) - \varphi_{\rm LLS}(x)|^2 \right\rangle \tag{5.21}$$

$$= \sum_{j=1}^{n_1} \langle |\theta_j^* - \theta_j|^2 \rangle \int_0^L \frac{dx}{L} \Big| \sum_{d\in\mathbb{Z}} \beta\Big(\frac{x - x_j + Ld}{h}\Big)\Big|^2 \tag{5.22}$$

$$= n_1 n_2^{-1} \int_{-s/2}^{s/2} \frac{hdz}{L} |\beta(z)|^2 = \|\beta\|_2^2 n_2^{-1}. \tag{5.23}$$

Next, we evaluate the deterministic term $\delta^2_{\rm det}$. In principle, the scaling of this error is the same as the one derived in Th. 3.5; however, the derivation is rather complicated because of the difference in the error measure.

Proposition 5.3 *Suppose that β is an LLS kernel of degree m with support s. Then, for any (q, M)-regular function φ (cf. Def. 5.2), the function $\varphi_{\rm LLS}$ in Eq.* (5.18) *satisfies*

$$\sqrt{{\rm MSE}(\varphi_{\rm LLS}, \varphi)} \le C h^q M, \tag{5.24}$$

where $q = m + \sigma$ and $C = \sqrt{\frac{2}{2\sigma+1}} \frac{s^{q+1}}{2^q (m-1)!} \sup_z |\beta(z)|$.

The proof of this proposition will be given at the end of this subsection.

So far, we have obtained the two terms of error:

$$\delta^2_{\rm stat} = \|\beta\|_2^2 n_2^{-1}, \qquad \delta^2_{\rm det} = C^2 h^{2q} M^2 = C^2 L^{2q} n_1^{-2q} M^2. \tag{5.25}$$

Noting that $N = n_1 n_2$, there exists a trade-off relation between the bias and the variance. The optimal balance between n_1 and n_2 is

$$n_1 = O\big((L^{2q} M^2 N)^{1/(2q+1)}\big), \qquad n_2 = O\big((L^{-2q} M^{-2} N^{2q})^{1/(2q+1)}\big), \tag{5.26}$$

where we attain the SQL by the position-state method in the classical scheme:

$$\delta_{\rm SQL} = O\big((L M^{1/q} N^{-1})^{q/(2q+1)}\big). \tag{5.27}$$

The corresponding Heisenberg limit is obtained if we consider n_2 entangled particles instead of separable ones. The optimal state is the generalized *Greenberger–Horne–Zeilinger state* (GHZ state):

$$|q\rangle \propto \frac{1}{\sqrt{2}} \big(|x_j; +\rangle^{\otimes n_2} + |x_j; -\rangle^{\otimes n_2}\big), \tag{5.28}$$

where we achieve $\langle|\theta_j^* - \theta_j|^2\rangle = {n_2}^{-2}$ instead of Eq. (5.16). This alters the exponent of n_2 in Eq. (5.25) accordingly:

$$\delta_{\mathrm{stat}}^2 = \|\beta\|_2^2 n_2^{-2}, \qquad \delta_{\mathrm{det}}^2 = C^2 L^{2q} {n_1}^{-2q} M^2. \tag{5.29}$$

which results in the optimal error

$$\delta_{\mathrm{HL}} = O\big((LM^{1/q}N^{-1})^{q/(q+1)}\big). \tag{5.30}$$

Proof of Prop. 5.3. We can use the same argument as the proof of Th. 3.5 to derive

$$\varphi_{\mathrm{LLS}}(x) - \varphi(x) = \sum_{j=1}^{n_1} \sum_{d\in\mathbb{Z}} \beta\Big(\frac{x - x_j + Ld}{h}\Big) R_x(x_j - Ld), \tag{5.31}$$

where $R_x(x')$ is the residual of the Taylor expansion:

$$R_x(x') = \int_x^{x'} \frac{(x'-t)^{m-1}}{(m-1)!}[\varphi^{(m)}(t) - \varphi^{(m)}(x)]dt. \tag{5.32}$$

When $x \le x' \le x + sh/2$, this residual term can be bounded from above as

$$\begin{aligned}
\big|R_x(x')\big|^2 &\le (x'-x)\int_x^{x'} \Big|\frac{(x'-t)^{m-1}}{(m-1)!}[\varphi^{(m)}(t) - \varphi^{(m)}(x)]\Big|^2 dt \\
&\le \frac{(sh/2)^{2m-1}}{[(m-1)!]^2}\int_x^{x'} \big|\varphi^{(m)}(t) - \varphi^{(m)}(x)\big|^2 dt \\
&\le \frac{(sh/2)^{2m-1}}{[(m-1)!]^2}\int_{x-sh/2}^{x+sh/2} \big|\varphi^{(m)}(t) - \varphi^{(m)}(x)\big|^2 dt \\
&\le \frac{(sh/2)^{2m-1}}{[(m-1)!]^2}\int_{-sh/2}^{sh/2} \big|\Delta_\varepsilon \varphi^{(m)}(x)\big|^2 d\varepsilon,
\end{aligned} \tag{5.33}$$

where the first inequality results from the Cauchy–Schwartz inequality. The same bound also holds for $x - sh/2 \le x' \le x$.

Let us set $B = \sup_z|\beta(z)|$. We restrict the pairs (j,d) in the summation (5.31) such that the value $\beta(\frac{x-x_j+Ld}{h})$ is nonzero. This condition is satisfied only if $|x - x_j + Ld| \le sh/2$, and there exist at most s instances of such (j,d). Hence,

$$\begin{aligned}
\big|\varphi_{\mathrm{LLS}}(x) - \varphi(x)\big|^2 &= \Big|\sum_{j,d} \beta\Big(\frac{x - x_j + Ld}{h}\Big) R_x(x_j - Ld)\Big|^2 \\
&\le \sum_{j,d}\Big|\beta\Big(\frac{x - x_j + Ld}{h}\Big)\Big|^2 \cdot \sum_{j,d}\Big|R_x(x_j - Ld)\Big|^2 \\
&\le sB^2 \cdot s\frac{(sh/2)^{2m-1}}{[(m-1)!]^2}\int_{-sh/2}^{sh/2} \big|\Delta_\varepsilon \varphi^{(m)}(x)\big|^2 d\varepsilon.
\end{aligned} \tag{5.34}$$

By integrating both hands over $0 \le x \le L$, we obtain

$$\begin{aligned}\mathrm{MSE}(\varphi_{\mathrm{LLS}}, \varphi) &= \int_0^L \frac{dx}{L} |\varphi_{\mathrm{LLS}}(x) - \varphi(x)|^2 \\ &\le s^2 B^2 \frac{(sh/2)^{2m-1}}{[(m-1)!]^2} \int_{-sh/2}^{sh/2} M^2 \varepsilon^{2\sigma} d\varepsilon \\ &\le 2s^2 B^2 M^2 \frac{(sh/2)^{2m+2\sigma}}{(2\sigma+1)[(m-1)!]^2} = C^2 M^2. \end{aligned}$$ □

5.2.2 A caveat: the phase periodicity problem

In the estimation of a phase function, however, we need to consider the 2π-periodicity of the phase. For this reason, we measure the distance between two phases θ and θ' by ignoring the difference of an integer multiple of 2π:

$$[\theta' - \theta]_{2\pi} = \min_{n \in \mathbb{Z}} |\theta' - \theta - 2\pi n|. \tag{5.35}$$

On the estimation of a single phase θ_j, we only need to replace the absolute value with $[\cdot]_{2\pi}$:

$$\langle [\theta_j^* - \theta_j]_{2\pi}^2 \rangle = O(n_2^{-1}) \tag{5.36}$$

for separable probe states and

$$\langle [\theta_j^* - \theta_j]_{2\pi}^2 \rangle = O(n_2^{-2}) \tag{5.37}$$

for entangled probe states [81].

For $m \ge 1$, however, the phase periodicity problem have a serious effect on the LLS. Although each phase θ_j can be estimated modulo 2π, the entire function $\varphi^*(x)$ in Eq. (5.17) cannot be determined even modulo 2π, since the coefficients $\beta(\frac{x-x_j}{h})$ is not an integer in general. This does not apply to the case of $m = 0$, since the kernel $\beta(z)$ can always be an integer for this case.

In order to circumvent this problem, we require an additional regularity on the phase function φ. One such regularity is a uniform bound on the first derivative:

$$\|\varphi'\|_\infty \le M_1, \tag{5.38}$$

in which case $|\theta_{j+1} - \theta_j| \le \pi/2$ holds for sufficiently large n_1. Then, except for a rare event with probability exponentially small in n_2, we can estimate the relative angle $\theta_{j+1}^* - \theta_j^*$ within the error of $\pi/2$. Therefore, we can calculate the estimator without being annoyed by the phase periodicity.

We note that, even with Eq. (5.38), the estimator $\varphi^*(x)$ is still determined modulo 2π. Therefore, a pertinent error measure can be defined by replacing the absolute value in the mean-square error (5.6) with $[\cdot]_{2\pi}$:

$$\mathrm{MSPE}(\varphi^*, \varphi) := \int_0^L \frac{dx}{L} [\varphi^*(x) - \varphi(x)]_{2\pi}^2, \tag{5.39}$$

which is called the *mean-square periodic error* in Ref. [82]. Then, the estimation error δ is described by

$$\delta^2 = \langle \mathrm{MSPE}(\varphi^*, \varphi) \rangle, \tag{5.40}$$

which are bounded by the SQL (5.27) or the Heisenberg limit (5.30) in their respective schemes.

A yet simpler solution is to restrict the phase function within a finite range, e.g., $|\varphi(x)| \leq \pi/2$, in which case the phase periodicity problem is no longer exists, and the estimation error may be given by Eq. (5.7).

5.2.3 Wavenumber-state method

In this section, we introduce another method of function estimation, which turns out to be only applicable for $0 < q \leq 1$, i.e. $m = 0$. The one-particle probe state is set to

$$|q\rangle = \frac{1}{\sqrt{2L}} \int_0^L dx (|x;+\rangle + |x;-\rangle), \tag{5.41}$$

which is completely delocalized in position as opposed to Eq. (5.14). In fact, this state is the eigenstate for the wavenumber $k = 0$, whence we call this method the *wavenumber-state method*. We note that all N particles should be in this identical state in the classical scheme.

After the unitary evolution by U_φ, this state will evolve into

$$|q_\varphi\rangle = \frac{1}{\sqrt{2L}} \int_0^L dx (e^{i\varphi(x)}|x;+\rangle + |x;-\rangle). \tag{5.42}$$

Therefore, if we can completely identify the wavefunction of $|q_\varphi\rangle$, we can retrieve the target function φ. Of course, the quantum state $|q_\varphi\rangle$ can be accessed only through a finite number of measurements, which restricts the amount of information available. The quantum state $|q_\varphi\rangle$ can be approximately reconstructed via the QST introduced in Sec. 2.1.4. However, the error bound is known to diverge when the dimension of the Hilbert space is infinite, which is the case with the function estimation.

Therefore, in the wavenumber-state method, we take the following strategy:

- We choose an appropriate subspace $\mathcal{H}$ of the Hilbert space of the particle, whose dimension $D = \dim \mathcal{H}$ is finite.
- We perform the projection measurement with the POVM generated by $\{P, I - P\}$, where P denotes the projection operator onto $\mathcal{H}$.
- We discard the probe state if the measurement outcome corresponds to $I-P$. This postselection yields the quantum state $|\tilde{q}_\varphi\rangle \propto P|q_\varphi\rangle \in \mathcal{H}$.
- We perform the QST on $|\tilde{q}_\varphi\rangle$, and the estimator φ^* is retrieved from the reconstructed state $|q^*\rangle$ so that $|q_{\varphi^*}\rangle$ is the closest to the state $|q^*\rangle$.

The estimation error by this method can be bounded by the quantum infidelity $\mathcal{I}(|q^*\rangle, |q_\varphi\rangle)$ between the reconstructed state $|q^*\rangle$ and the original probe state $|q_\varphi\rangle$, where the quantum infidelity is generally defined as

$$\mathcal{I}(|q\rangle, |q'\rangle) = 1 - |\langle q|q'\rangle|^2. \tag{5.43}$$

In fact, the mean-square periodic error in Eq. (5.39) can be bounded from above by the infidelity. First, the inequality

$$\mathcal{I}(|\varphi\rangle, |q_{\varphi^*}\rangle) \le 2\Big[\mathcal{I}(|q^*\rangle, |q_\varphi\rangle) + \mathcal{I}(|q^*\rangle, |q_{\varphi^*}\rangle)\Big] \le 4\mathcal{I}(|q^*\rangle, |q_\varphi\rangle) \tag{5.44}$$

follows from the definition of the estimator φ^*, thus we have

$$\begin{aligned}
4\mathcal{I}(|q^*\rangle, |q_\varphi\rangle) &\ge \mathcal{I}(|q_{\varphi^*}\rangle, |q_\varphi\rangle) \\
&= 1 - \left|\frac{1}{2L}\left[\int_0^L dx e^{-i\varphi^*(x)} e^{i\varphi(x)} + \int_0^L dx \cdot 1\right]\right| \\
&\ge 1 - \frac{1}{2L}\int_0^L dx |e^{i[\varphi(x)-\varphi^*(x)]} + 1| \\
&= \frac{1}{L}\int_0^L dx \left(1 - \cos\frac{1}{2}[\varphi(x) - \varphi^*(x)]\right) \\
&\ge \frac{1}{\pi^2 L}\int_0^L dx [\varphi(x) - \varphi^*(x)]_{2\pi}^2 = \frac{1}{\pi^2}\mathrm{MSPE}(\varphi^*, \varphi).
\end{aligned} \tag{5.45}$$

Therefore, the estimation error can be decomposed into two parts:

$$\delta^2 \le 4\pi^2 \langle \mathcal{I}(|q^*\rangle, |q_\varphi\rangle)\rangle \le 8\pi^2(\delta_{\mathrm{PS}}^2 + \delta_{\mathrm{QST}}^2), \tag{5.46}$$

$$\delta_{\mathrm{PS}}^2 := \mathcal{I}(|\tilde{q}_\varphi\rangle, |q_\varphi\rangle), \qquad \delta_{\mathrm{QST}}^2 := \langle \mathcal{I}(|q^*\rangle, |\tilde{q}_\varphi\rangle)\rangle. \tag{5.47}$$

On the right-hand side of Eq. (5.46), δ_{PS} is a deterministic value arising from the postselection, while δ_{QST} is a stochastic value owing to the QST. In this sense, Eq. (5.46) may be interpreted as the bias–variance decomposition similar to (5.19).

The estimation error depends on the choice of the Hilbert space $\mathcal{H}$. We will see that the estimation error can be efficiently bounded when $\mathcal{H}$ consists of the quantum state with a small wavenumber.

First, we express this state in terms of the basis consisting of wavenumber eigenstates:

$$|q_\varphi\rangle = \frac{1}{\sqrt{2}}\left(\sum_{k\in\mathbb{Z}} u_k |e_k\rangle + |f_0\rangle\right), \tag{5.48}$$

$$|f_0\rangle = \int_0^L \frac{dx}{\sqrt{L}} |x; -\rangle, \qquad |e_k\rangle = \int_0^L \frac{dx}{\sqrt{L}} e^{2\pi i k x/L} |x; +\rangle. \tag{5.49}$$

The coefficients u_k are determined by the Fourier transform of the function

$$u(x) := e^{i\varphi(x)} = \sum_{k\in\mathbb{Z}} u_k e^{2\pi i k x/L}. \tag{5.50}$$

Since $|\Delta_\varepsilon u(x)| = |e^{i\varphi(x+\varepsilon)-i\varphi(x)}| \le |\varphi(x+\varepsilon)-\varphi(x)| = |\Delta_\varepsilon \varphi(x)|$, we see that the regularity of the phase function φ is inherited to the function u:[3]

$$\sup_{\varepsilon\neq 0} \frac{\|\Delta_\varepsilon u\|_2}{|\varepsilon|^\sigma} \le \sup_{\varepsilon\neq 0} \frac{\|\Delta_\varepsilon \varphi\|_2}{|\varepsilon|^\sigma} \le M. \tag{5.51}$$

Next, we use the following theorem:

Theorem 5.4 *Let us set* $0 < q \le 1$. *We consider an arbitrary function* $u : [0, L] \to \mathbb{C}$ *with the periodic boundary condition, and define its Fourier transform* $\{u_k\}_{k\in\mathbb{Z}}$ *by*

$$u(x) := \sum_{k\in\mathbb{Z}} u_k e^{2\pi i k x/L}. \tag{5.52}$$

Then, there exists a real number $C_q > 0$ *such that if* u *is* (q, M)*-regular, the high-wavenumber components of* $\{u_k\}$ *can be bounded as*

$$\sum_{k=K}^{\infty} (|u_k|^2 + |u_{-k}|^2) \le C_q M^2 \left(\frac{L}{K}\right)^{2q} \tag{5.53}$$

for an arbitrary integer $K > 0$.

The proof of this theorem is given in Appendix B, which is based on Ref. [83].

Now, let $\mathcal{H}$ be the Hilbert space spanned by $|f_0\rangle, |e_{-K+1}\rangle, |e_{-K+2}\rangle, \ldots, |e_{K-1}\rangle$. This Hilbert space has a finite dimension $\dim \mathcal{H} = 2K$, and contains quantum states with wavenumbers less than $2\pi K/L$. Moreover, we can use Th. 5.4 to bound the probability p of failure in the postselection:

$$p = \|(1-P)|q_\varphi\rangle\|^2 = \sum_{k=K} \frac{1}{2}(|u_k|^2 + |u_{-k}|^2) \le \frac{1}{2} C_q M^2 \left(\frac{L}{K}\right)^{2q}, \tag{5.54}$$

indicating that p is negligible for sufficiently large K. In fact, the postselection infidelity is also equal to p:

$$\begin{aligned} \delta_{\rm PS}^2 = \mathcal{I}(|\tilde{q}_\varphi\rangle, |q_\varphi\rangle) &= 1 - |\langle q_\varphi|\tilde{q}_\varphi\rangle|^2 \\ &= 1 - \left|\langle q_\varphi| \frac{P|q_\varphi\rangle}{\|P|q_\varphi\rangle\|}\right|^2 = 1 - \|P|q_\varphi\rangle\|^2 \\ &= p = O(K^{-2q} L^{2q} M^2). \end{aligned} \tag{5.55}$$

On the other hand, the QST infidelity $\delta_{\rm QST}^2$ depends on the dimension $2K$ of the Hilbert space and the number N of particles used in the estimation[4]. The optimal fidelity follows from Eq. (2.33):

$$\delta_{\rm QST}^2 = \langle \mathcal{I}(|q^*\rangle, |\tilde{q}_\varphi\rangle)\rangle = \frac{2K-1}{N+2K} = O(KN^{-1}), \tag{5.56}$$

[3] the similar relation does *not* hold for $m \ge 1$, which is why we have restricted q to $0 < q \le 1$ for the wavenumber-state method.

[4] Since postselection may fail, $(1-p)N$ particles are actually available for QST on average. This does not affect the error evaluation as p remains small for sufficiently large N.

where we note that $K \ll N$ is required for small infidelity.

Finally, the optimal error $\delta^2 = \langle \mathrm{MSPE}(\varphi^*, \varphi) \rangle$ is obtained by the trade-off between Eqs. (5.55) and (5.56). This is attained by setting the cutoff wavenumber K to $K = O((L^{2q} M^{-2} N)^{1/(2q+1)})$. In this case, when

$$\delta_{\mathrm{SQL}} = O((L M^{1/q} N^{-1})^{q/(2q+1)}) \tag{5.57}$$

is obtained.

5.2.4 Amplified estimation

In the quantum scheme, the wavenumber-state method is not so straightforward as the position-state method. In order to exploit n_2-particle entanglement for $n_2 > 1$, the wavenumber eigenstate in Eq. (5.41) should be replaced with the multipartite *Einstein–Podolsky–Rosen state* (EPR state):

$$|q(n_2)\rangle \propto \frac{1}{\sqrt{2L}} \int_0^L dx (|x;+\rangle^{\otimes n_2} + |x;-\rangle^{\otimes n_2}). \tag{5.58}$$

This state will evolve into

$$|q_\varphi(n_2)\rangle \propto \frac{1}{\sqrt{2L}} \int_0^L dx (e^{in\varphi(x)} |x;+\rangle^{\otimes n_2} + |x;-\rangle^{\otimes n_2}) \tag{5.59}$$

by the unitary evolution $U_\varphi^{\otimes n_2}$. By comparing Eq. (5.59) with Eq. (5.42), we find that the amplified phase function $\psi(x) = n_2 \varphi(x)$ can be estimated by performing the postselective measurement and the QST on the probe state (5.59). Since φ is a (q, M)-regular function, the amplified function ψ is found to be $(q, n_2 M)$-regular.

By preparing n_1 copies of this probe state, we can obtain the estimator ψ^* for the amplified function ψ. Therefore, we expect that the estimator φ^* for the original phase function be obtained by $\varphi^*(x) = \frac{1}{n_2} \psi^*(x)$. We note that this expectation is not entirely correct; $\psi^*(x)$ can be calculated only modulo 2π in the wavenumber-state method, leading to the ambiguity of $\varphi^*(x)$ by an integer multiple of $2\pi/n_2$. We will show the how to remove this ambiguity problem later.

For the amplified estimator ψ^*, the postselection and QST infidelities can be calculated similarly to Eqs. (5.55) and (5.56):

$$\delta_{\mathrm{PS}} = O(K^{-q} L^q n_2 M), \qquad \delta_{\mathrm{QST}} = O\left(\left(1 + \frac{n_1}{K} \right)^{-1/2} \right). \tag{5.60}$$

We note that, in order to remove the phase ambiguity in φ^*, the original error in ψ^* must be below some constant, say $\frac{\pi}{3}$. This implies that the number n_1 must be comparable with K, in which case we may simply write $\delta_{\mathrm{QST}} = O(K^{1/2} n_1^{-1/2})$.

Now, we suppose that $\varphi^*(x)$ can be calculated without ambiguity. For the estimator φ^* of the original function φ based on $\varphi^*(x) = \frac{1}{n_2} \psi^*(x)$, each component of the error should be multiplied by n_2^{-1}:

$$\delta_{\mathrm{PS}} = O(K^{-q} L^q M), \quad \delta_{\mathrm{QST}} = O(K^{1/2} n_1^{-1/2} n_2^{-1}) = O(K^{1/2} n_1^{1/2} N^{-1}), \tag{5.61}$$

where $N = n_1 n_2$ is the total number of particles. Equation (5.61) indicates that n_1 should be as small as possible, which yields $n_1 = O(K)$. Therefore, the optimal error is obtained as the trade-off relation between

$$\delta_{\mathrm{PS}} = O(K^{-q} L^q M) \quad \text{and} \quad \delta_{\mathrm{QST}} = O(K N^{-1}). \tag{5.62}$$

The terms δ_{PS} and δ_{QST} become the same order of magnitude when K is set to $O\big((L^q M N)^{1/(q+1)}\big)$, which leads to the Heisenberg limit:

$$\delta_{\mathrm{HL}} = O\big((L M^{1/q} N^{-1})^{q/(q+1)}\big). \tag{5.63}$$

Because $n_1 = O(K)$ and $n_1 n_2 = N$, the parameters n_1 and n_2 should be chosen so that

$$n_1 = O\big((L^q M N)^{1/(q+1)}\big), \quad n_2 = O\big((L^{-q} M^{-1} N^q)^{1/(q+1)}\big). \tag{5.64}$$

Now, we explain how to remove the ambiguity concerning the estimator φ^*. This can be done by an idea similar to the hierarchical estimation in Sec. 4.1.

We choose an integer $J > 0$ such that $4c_1 2^J = (L^{-q} M^{-1} N^q)^{1/(q+1)}$ for sufficiently large c_1, and consider a sequence of estimators ψ_j^* $(j = 0, 1, \ldots, J)$ by setting the parameters

$$\begin{cases} K = c_2 (L^q M N)^{1/(q+1)}, \\ n_1 = c_1 (J + 1 - j)(L^q M N)^{1/(q+1)} = \frac{1}{4}(J + 1 - j) 2^{-J} N, \\ n_2 = 2^j. \end{cases} \tag{5.65}$$

Then, each ψ_j^* becomes an estimator for the amplified function $\psi_j(x) = 2^j \varphi(x)$, and the final estimator $j = J$ corresponds to the Heisenberg limit in Eq. (5.63). Moreover, this is consistent with the number N of particles; in fact, the total number of particles needed in the estimation is

$$\sum_{j=0}^{J} n_1 n_2 = \sum_{j=0}^{J} \frac{1}{4}(J + 1 - j) 2^{j-J} N = [1 - 2^{-J}(J + 1)] N, \tag{5.66}$$

and does not exceed N.

The expected error of the estimator is $\langle \mathrm{MSPE}(\psi_j^*, \psi_j) \rangle = O(1)$, which can be arbitrarily small by setting sufficiently large c_1 and c_2. Furthermore, the factor $J + 1 - j$ in n_1 can be utilized for the Chernoff bound (Th. 4.1), in which the probability of a significant failure in QST becomes exponentially small in $J + 1 - j$. In fact, for sufficiently large c_1, we have

$$\mathbf{P}\left[\mathrm{MSPE}(\psi_j^*, \psi_j) \geq \left(\frac{\pi}{3}\right)^2\right] \leq 8^{-(J-j)}, \tag{5.67}$$

where $\mathbf{P}$ denotes the probability of the event. Hence, if we define a subset X_n of the interval $[0, L]$ by

$$I_n = \{0 \leq x \leq L \mid [\psi_j^*(x) - \psi_j(x)]_{2\pi} \geq \pi/3\}, \tag{5.68}$$

its Lebesgue measure $|I_n|$ is at most $8^{-(J-j)}L$ on average.

Finally, we define the estimator $\varphi^*(x)$ for every x so that

$$[\psi_j^*(x) - 2^j\varphi^*(x)]_{2\pi} < \frac{\pi}{3}, \qquad \forall\, j = 1, \ldots, m \tag{5.69}$$

holds for as large m as possible. Such $\varphi^*(x)$ can be taken as long as $x \notin I_0 \cup I_1 \cup \cdots \cup I_m$ because of Eq. (5.68). Once Eq. (5.69) is satisfied, the phase ambiguity is resolved up to modulo $\frac{2\pi}{2^m}$ and the value $[\varphi^*(x) - \varphi(x)]_{2\pi}$ is at most $\frac{2\pi}{3\cdot 2^m}$.

Therefore, the error of the estimator φ^* can be calculated as

$$\begin{aligned}
\delta^2 &= \left\langle \int_0^L \frac{dx}{L}[\varphi^*(x) - \varphi(x)]_{2\pi}^2 \right\rangle \\
&\le \left\langle \sum_{j=0}^{J-1} \frac{|I_j|}{L}\left(\frac{2\pi}{3\cdot 2^m}\right)^2 + \left(\frac{2\pi}{3\cdot 2^J}\right)^2 \right\rangle \\
&\le \sum_{j=0}^{J} \frac{4\pi^2}{9\cdot 4^j}\cdot 8^{(j-J)} = \frac{8\pi^2}{9\cdot 4^J} = \frac{128\pi^2 c_1^2}{9}(L^{-q}M^{-1}N^q)^{2/(q+1)},
\end{aligned} \tag{5.70}$$

which is the Heisenberg limit with an additional factor.

The SQL and the Heisenberg limit for the wavenumber-state method have been obtained in Eqs. (5.57) and (5.70). These are exactly the same order of magnitude as the error bounds for the position-state method (Eqs. (5.27) and (5.30)). In fact, the trade-off relations in the position- and wavenumber-state methods can be mapped to each other by setting the parameter K to the number n_1 of the sample points or vice versa.

This relation resembles the Nyquist–Shannon sampling theorem in signal detection theory. This theorem states that low-pass filtering with cutoff wavenumber f is informationally equivalent to sampling with the sampling rate $2f$. Since the cutoff wavenumber is $2\pi K/L$ and the sampling rate is n_1/L in our setting, it can be intuitively understood from the sampling theorem that the parameters K and n_1 are equivalent up to a constant factor.

5.3 Theoretical lower bounds

So far, we have presented two complementary methods of function estimation: the position-state method and the wavenumber-state method. In both methods, however, our analysis leads to the same order of the SQL and the Heisenberg limit

$$\delta_{\mathrm{SQL}} = O\big((LM^{1/q}N^{-1})^{q/(2q+1)}\big), \qquad \delta_{\mathrm{HL}} = O\big((LM^{1/q}N^{-1})^{q/(q+1)}\big). \tag{5.71}$$

In the $q \to \infty$ limit, the exponents in Eq. (5.71) become identical to those in the error bounds of parameter estimation:

$$\delta_{\mathrm{SQL}} = O(N^{-1/2}), \qquad \delta_{\mathrm{HL}} = O(N^{-1}). \tag{5.72}$$

For a finite q, however, the error bounds of function estimation decrease slower than the error bounds of parameter estimation. This is due to the bias in function estimation, which is inevitable because we cannot fully explore the infinite degrees of freedom within finite time. As a result, we are forced to spend some of the resource to decrease the bias, leading to the slower decrease. This behavior is also known in other problems of classical function estimation, e.g., density kernel estimation [84].

In fact, Eq. (5.71) give error bounds regardless of the method of estimation, and hence the position- and wavenumber-state methods both provide optimal error bounds up to a constant factor. In this section, we prove these rigorous error bounds by using the analysis of multiparameter quantum metrology.

5.3.1 Reduction to multiparameter estimation

Since we are interested in a lower bound on the estimation error, we may replace the phase function estimation problem with an easier one. More formally, we may consider an unknown parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$ such that any method of estimating phase function φ can be used for the estimation of $\boldsymbol{\theta}$, and that

$$\langle \mathrm{MSE}(\varphi^*, \varphi) \rangle \leq \delta^2 \implies \langle \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \rangle \leq (c\delta)^2 \tag{5.73}$$

holds for some constant $c > 0$.

Following Sec. 2.3.2, we consider estimation of a Hamiltonian $H_{\boldsymbol{\theta}} = \sum_{j=1}^{2K} \theta_j X_j$ with

$$X_k = -\int_0^L dx \sin\frac{2\pi kx}{L} |x;+\rangle\langle x;+|, \tag{5.74}$$

$$X_{K+k} = -\int_0^L dx \cos\frac{2\pi kx}{L} |x;+\rangle\langle x;+|. \tag{5.75}$$

By defining

$$\varphi_{\boldsymbol{\theta}}(x) = \sum_{k=1}^{K} \left(\theta_k \sin\frac{2\pi kx}{L} + \theta_{K+k} \cos\frac{2\pi kx}{L} \right), \tag{5.76}$$

we can write the Hamiltonian $H_{\boldsymbol{\theta}}$ and its evolution over unit time as

$$H_{\boldsymbol{\theta}} = \sum_{j=1}^{2K} \theta_j X_j = -\int_0^L dx \varphi_{\boldsymbol{\theta}}(x) |x;+\rangle\langle x;+|, \tag{5.77}$$

$$e^{-iH_{\boldsymbol{\theta}}} = \int_0^L dx \left[e^{i\varphi_{\boldsymbol{\theta}}(x)} |x;+\rangle\langle x;+| + |x;-\rangle\langle x;-| \right] \tag{5.78}$$

$$= U_{\varphi_{\boldsymbol{\theta}}}. \tag{5.79}$$

Therefore, if we can somehow estimate the phase function φ, we can use the same method for estimating $\boldsymbol{\theta}$ via the relation (5.76). Furthermore, the estimation error of $\boldsymbol{\theta}$ coincides with that of φ:

$$\mathrm{MSE}(\varphi_{\boldsymbol{\theta}^*}, \varphi_{\boldsymbol{\theta}}) = \|\varphi_{\boldsymbol{\theta}^* - \boldsymbol{\theta}}\|_2^2 = \frac{1}{2}\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2, \tag{5.80}$$

which follows from Perseval's equality.

We note that the function $\varphi_{\boldsymbol{\theta}}$ must satisfy the regularity condition. This means that the range of the parameter vector $\boldsymbol{\theta}$ must be restricted as well. In particular, we need to fix an energy scale E as the upper bound on the Euclid norm of $\boldsymbol{\theta}$:

$$\|\boldsymbol{\theta}\| \leq E, \tag{5.81}$$

which can be determined by the following theorem:

Theorem 5.5 *Let us set $q = m + \sigma$ with an integer $m \geq 0$ and a real number $0 < \sigma \leq 1$. We consider an arbitrary m-time differentiable function $u : [0, L] \to \mathbb{C}$ with the periodic boundary condition, and define its Fourier transform $\{u_k\}_{k\in\mathbb{Z}}$ by*

$$u(x) := \sum_{k\in\mathbb{Z}} u_k e^{2\pi i k x/L}. \tag{5.82}$$

Then, there exists a real number $C'_q > 0$ such that if

$$\sum_{k=1}^{\infty} \left(\frac{|k|}{L}\right)^{2q} (|u_k|^2 + |u_{-k}|^2) \leq C'_q M^2 \tag{5.83}$$

holds for $M > 0$, the function u is (q, M)-regular.

We give the proof of this theorem in Appendix B. Since the Fourier coefficients of $u(x) = \varphi_{\boldsymbol{\theta}}(x)$ are

$$u_k = \begin{cases} \frac{1}{2}(\theta_{K+k} - i\theta_k) & (k = 1, \ldots, K); \\ \frac{1}{2}(\theta_{K+|k|} + i\theta_{|k|}) & (k = -K, \ldots, -1); \\ 0 & (\text{otherwise}), \end{cases} \tag{5.84}$$

a sufficient condition for $\varphi_{\boldsymbol{\theta}}$ to be (q, M)-regular is

$$\sum_{k=1}^{K} \left(\frac{k}{L}\right)^{2q} \frac{|\theta_k|^2 + |\theta_{K+k}|^2}{2} \leq C'_q M^2, \tag{5.85}$$

or yet more simply,

$$\left(\frac{K}{L}\right)^{2q} \frac{\|\boldsymbol{\theta}\|_2^2}{2} \leq C'_q M^2. \tag{5.86}$$

Therefore, the (q, M)-regularity of the function $\varphi_{\boldsymbol{\theta}}$ is ensured by choosing the energy scale E of the order $O(K^{-q} L^q M)$.

5.3.2 Standard quantum limit

We derive the SQL of the multiparameter metrology for $\boldsymbol{\theta} \in \mathbb{R}^{2K}$. As explained in Sec. 2.3.3, a lower bound δ of an unbiased estimator $\boldsymbol{\theta}^*$ of a parameter vector

$\boldsymbol{\theta} \in \mathbb{R}^m$ is given by the quantum CR inequality. According to Eq. (2.78), we have

$$\delta \geq (\mathrm{tr}[J_{\boldsymbol{\theta}}^{-1}])^{1/2} \geq \frac{m}{(\mathrm{tr}\, J_{\boldsymbol{\theta}})^{1/2}}, \tag{5.87}$$

where $J_{\boldsymbol{\theta}}$ is the *quantum Fisher inforamtion* (QFI) matrix of the probe state. Therefore, a uniform bound on the error of an unbiased estimator is given as

$$\delta_{\mathrm{UUB}} = \inf_{\|\boldsymbol{\theta}\| \leq E} \frac{m}{(\mathrm{tr}\, J_{\boldsymbol{\theta}})^{1/2}}, \tag{5.88}$$

where UUB denotes the "Uniform Unbiased Bound."

On the other hand, such an uniform upper bound cannot be given on the error of an possibly biased estimator; the error is possibly zero at a specific instance of $\boldsymbol{\theta}$ (See an example in Sec. 2.1.1). Instead, there exists an error bound on the worst-case error δ defined by

$$\delta^2 = \sup_{\|\boldsymbol{\theta}\| \leq E} \langle \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \rangle. \tag{5.89}$$

In fact, we have the following inequality:

Proposition 5.6 *Let us define* $\delta_{\mathrm{WBB}} > 0$ *by*

$$\frac{1}{\delta_{\mathrm{WBB}}} = \frac{1}{\delta_{\mathrm{UUB}}} + \frac{1}{E}. \tag{5.90}$$

Then, the worst-case error δ *in Eq.* (5.89) *satisfies* $\delta \geq \delta_{\mathrm{WBB}}$.

The proof is at the end of this subsection. Proposition 5.6 means that we can find some instance of $\boldsymbol{\theta}$ in the region $\|\boldsymbol{\theta}\| \leq E$ at which the estimation error exceeds δ_{WBB}. Here, WBB denotes the "Worst-case Biased Bound."

Now, we calculate δ_{UUB} first. The sum of the square of the matrices in Eqs. (5.74) and (5.75) can be calculated as

$$\mathbf{X} := \sum_{j=1}^{2K} X_j^2 = \sum_{j=1}^{K} (X_k^2 + X_{K+k}^2) = K \int_0^L dx |x;+\rangle\langle x;+|, \tag{5.91}$$

whose operator norm is K. Let $J_{1,\boldsymbol{\theta}}$ be the QFI matrix of a one-particle probe state. With evolution time $T = 1$, the trace of $J_{1,\boldsymbol{\theta}}$ can be bounded from above by Prop. 2.5:

$$\mathrm{tr}\, J_{1,\boldsymbol{\theta}} \leq 4KT^2 = 4K. \tag{5.92}$$

For the QFI matrix $J_{\boldsymbol{\theta}}$ of an N-particle probe state, we have an upper bound $\mathrm{tr}\, J_{\boldsymbol{\theta}} \leq 4NK$, which is independent of the parameter $\boldsymbol{\theta}$. Therefore, the uniform bound δ_{UUB} in Eq. (5.88) is bounded from below as

$$\delta_{\mathrm{UUB}} \geq \inf_{\|\boldsymbol{\theta}\| \leq E} \frac{2K}{(\mathrm{tr}\, J_{\boldsymbol{\theta}})^{1/2}} \geq \left(\frac{K}{N}\right)^{1/2}. \tag{5.93}$$

Combining this fact with $E = O(K^{-q}L^q M)$, the worst-case error in (5.90) is found to be maximal when E and δ_{UUB} are comparable to each other, i.e., $K = O((L^{2q}M^{-2}N)^{1/(2q+1)})$ and

$$\delta_{\mathrm{SQL}} = O((LM^{1/q}N^{-1})^{q/(2q+1)}). \tag{5.94}$$

This coincides with the SQL in Sec. 5.2.

Proof of Prop. 5.6. Let us denote by $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ the stochastic average of a possibly biased estimator $\boldsymbol{\theta}^*$. The variance $V_{\boldsymbol{\theta}}$ is bounded from below as

$$\operatorname{tr} V_{\boldsymbol{\theta}} \geq \operatorname{tr}[D_{\boldsymbol{\theta}} J_{\boldsymbol{\theta}}^{-1} D_{\boldsymbol{\theta}}^{\mathrm{T}}] \geq \frac{(\operatorname{tr} D_{\boldsymbol{\theta}})^2}{(\operatorname{tr} J_{\boldsymbol{\theta}})} \geq m^{-2}(\operatorname{tr} D_{\boldsymbol{\theta}})^2 \delta_{\mathrm{UUB}}^2, \tag{5.95}$$

where the matrix $D_{\boldsymbol{\theta}}$ is defined by $\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}$. The first inequality in Eq. (5.95) is due to Prop. 2.3, while the second and the third inequalities follow from the Cauchy–Schwartz inequality and Eq. (5.88), respectively. Hence, the estimation error is bounded from below as

$$\left\langle \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \right\rangle = \|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 + \operatorname{tr} V_{\boldsymbol{\theta}} \geq \|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 + m^{-2}(\operatorname{tr} D_{\boldsymbol{\theta}})^2 \delta_{\mathrm{UUB}}^2, \tag{5.96}$$

where the matrix $D_{\boldsymbol{\theta}}$ is defined by $\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\mathrm{T}}$.

Let us assume the contrary of this proposition, i.e.,

$$\sup_{\|\boldsymbol{\theta}\| \leq E} \left\langle \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \right\rangle < \delta_{\mathrm{WBB}}^2. \tag{5.97}$$

Then, Eq. (5.96) suggests that both of the following inequalities must hold:

$$\sup_{\|\boldsymbol{\theta}\| \leq E} \|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\| < \delta_{\mathrm{WBB}} = \frac{E\delta_{\mathrm{UBB}}}{\delta_{\mathrm{UBB}} + E}, \tag{5.98}$$

$$\sup_{\|\boldsymbol{\theta}\| \leq E} m^{-1} \operatorname{tr} D_{\boldsymbol{\theta}} < \frac{\delta_{\mathrm{WBB}}}{\delta_{\mathrm{UBB}}} = \frac{E}{\delta_{\mathrm{UBB}} + E}. \tag{5.99}$$

Now, we consider the integral of $m^{-1} \operatorname{tr} D_{\boldsymbol{\theta}}$ over the region $\|\boldsymbol{\theta}\| \leq E$:

$$\mathcal{A} = \int_{\|\boldsymbol{\theta}\| \leq E} m^{-1} \operatorname{tr} D_{\boldsymbol{\theta}} d\boldsymbol{\theta} \tag{5.100}$$

If we denote by V the volume of an m-dimensional hyperball with unit radius, we obtain

$$\mathcal{A} = VE^m \cdot \sup_{\|\boldsymbol{\theta}\| \leq E} m^{-1} \operatorname{tr} D_{\boldsymbol{\theta}} < \frac{VE^{m+1}}{\delta_{\mathrm{UBB}} + E} \tag{5.101}$$

from Eq. (5.99). On the other hand, we may apply the divergence theorem to represent $\mathcal{A}$ by the integral over the $(m-1)$ dimensional hypersphere $\|\boldsymbol{\theta}\| = E$ with area mVE^{m-1}:

$$\mathcal{A} = \int_{\|\boldsymbol{\theta}\| \leq E} m^{-1} (\nabla_{\boldsymbol{\theta}} \cdot \boldsymbol{\mu}_{\boldsymbol{\theta}}) d\boldsymbol{\theta} = \int_{\|\boldsymbol{\theta}\| = E} m^{-1} (\boldsymbol{n}_{\boldsymbol{\theta}} \cdot \boldsymbol{\mu}_{\boldsymbol{\theta}}) d\boldsymbol{S}_{\boldsymbol{\theta}}, \tag{5.102}$$

where $d\boldsymbol{S}_{\boldsymbol{\theta}}$ is the infinitesimal area and $\boldsymbol{n}_{\boldsymbol{\theta}}$ is the normal vector of the hypersphere. Furthermore, from Eq. (5.98) we obtain

$$\begin{aligned}\boldsymbol{n}_{\boldsymbol{\theta}} \cdot \boldsymbol{\mu}_{\boldsymbol{\theta}} &= \boldsymbol{n}_{\boldsymbol{\theta}} \cdot \boldsymbol{\theta} + \boldsymbol{n}_{\boldsymbol{\theta}} \cdot (\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &\geq \|\boldsymbol{\theta}\| - \|\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\theta}\| > E - \frac{E\delta_{\mathrm{UBB}}}{\delta_{\mathrm{UBB}} + E}. \end{aligned} \tag{5.103}$$

Hence,

$$\mathcal{A} > mVE^{m-1} \cdot m^{-1}\left(E - \frac{E\delta_{\mathrm{UBB}}}{\delta_{\mathrm{UBB}} + E}\right) = \frac{VE^{m+1}}{\delta_{\mathrm{UBB}} + E}. \tag{5.104}$$

The proposition is proved since Eqs. (5.101) and (5.104) lead to contradiction. □

5.3.3 Heisenberg limit

To derive the Heisenberg limit, we assume a simple picture on the initial probe state. The first assumption is that the N-particle probe state splits into n_1 separate groups such that each group contains n_2 entangled particles ($N = n_1 n_2$). Within each group, the probe state undergoes unitary evolution by $U_{\varphi}^{\otimes n_2}$. As we have seen in Sec. 2.3.1, the n_2-particle Hilbert space can be divided into invariant sectors due to the symmetry under the permutation group. Upon each sector, the unitary gate behaves as U_{φ}^{n} with an integer $0 \leq n \leq n_2$. Therefore, the estimation in each sector is equivalent to the estimation of an amplified phase function $\psi(x) = n\varphi(x)$, with the maximum amplification $n = n_2$ is attained when the particles are in a completely symmetric state. Hence, the second assumption is that, the probe state is in a symmetric sector where the unitary evolution acts as the phase shift by $\psi(x) = n\varphi(x)$.

Now, let us derive the Heisenberg limit in the aforementioned picture; we hypothesize that the Heisenberg limit will be the same for a more general class of entangled states. We begin with the SQL for the estimation error of the amplified function $\psi = n\varphi$, where M and N in Eq. (5.94) are replaced by nM and n_1, respectively:

$$\delta_{\mathrm{SQL},n} \geq O\big((Ln^{1/q}M^{1/q}n_1^{-1})^{q/(2q+1)}\big). \tag{5.105}$$

We can calculate the estimator of the original function φ only if the error $\delta_{\mathrm{SQL},n}$ does not exceed a certain constant in the order of unity. Therefore, n must be small enough that

$$n \leq O(L^{-q}M^{-1}n_1^{q}), \tag{5.106}$$

in which case the estimation error of the original function can be bounded from below as

$$\delta \geq n^{-1}\delta_{\mathrm{SQL},n} \geq O\big((LM^{1/q}n^{-2}n_1^{-1})^{q/(2q+1)}\big). \tag{5.107}$$

Noting that the total number of particles is $N = n_1 n_2 \geq n_1 n$, we obtain

$$\begin{aligned} n^2 n_1 &= [(nn_1)^{2q+1}(nn_1^{-q})]^{1/(q+1)} \\ &\leq [N^{2q+1} \cdot O(L^{-q}M^{-1})]^{1/(q+1)} = O\big((L^{-q}M^{-1}N^{2q+1})^{1/(q+1)}\big), \end{aligned} \tag{5.108}$$

where we have used Eq. (5.106). By substituting Eq. (5.108) into Eq. (5.107), we obtain a lower bound on the estimation error

$$\delta_{\mathrm{HL}} = O\big((LM^{1/q}N^{-1})^{q/(q+1)}\big), \tag{5.109}$$

which is of the same order of magnitude as the Heisenberg limit in Sec. 5.2.

In fact, the wavenumber K in this section plays the same role as the wavenumber K in the wavenumber-state method (Sec. 5.2.3); if we replace E and δ_{UUB} with δ_{PS} and δ_{QST}, respectively, the trade-off relation (5.90) corresponds to Eq. (5.46). Actually, the wavenumber-state method is designed with intention to asymptotically satisfy the quantum CR inequality in the theoretical analysis, although this attempt is successful only for the case of $0 < q \leq 1$.

5.4 Some remarks on the phase function estimation

5.4.1 Long-range limit and Gaussian process

First, we consider the limit in which the domain of the function L tends to infinity. We claim that in our notation, the constant M in the (q, M)-regularity and the estimation error δ are both intensive quantities while the number N of particles is extensive.

In fact, given a (q, M)-regular function φ and its estimator φ^* with estimation error δ, we may regard them as periodic functions over $[0, nL]$ for any positive integer n. However, even when the domain is extended to $[0, nL]$, the function φ is also (q, M)-regular and the estimation error of φ^* is δ. This is due to our definition of the (q, M)-regularity (Def. 5.2) and the mean-square (periodic) error (Eqs. (5.6) and (5.39)), in which the integral over $[0, L]$ is divided by L.

Let $\mathcal{N} = N/L$ be the number of particles per unit length of the domain, so that all concerning quantities become intensive. Then, the SQL and the Heisenberg limit can be described in a way independent of L:

$$\delta_{\mathrm{SQL}} = O\big((M^{1/q}\mathcal{N}^{-1})^{q/(2q+1)}\big), \qquad \delta_{\mathrm{HL}} = O\big((M^{1/q}\mathcal{N}^{-1})^{q/(q+1)}\big). \tag{5.110}$$

These results are natural in the sense that the number of particles required for the estimation is in proportion to L.

The $L \to \infty$ limit also corresponds to a generalization of the preceding studies by Berry *et al.* [39–41, 43, 68, 85] of quantum metrology for Gaussian signals. They have applied the filtering theory (Sec. 3.2) and quantum-optical measurements to a signal $\{s(t)\}_{t\in\mathbb{R}}$ with the power spectrum $I_s(\omega) \sim |\omega|^{-p}$. As a result, they have obtained the SQL and the Heisenberg limit:

$$\delta_{\mathrm{SQL}} = O\big(\mathcal{N}^{-(p-1)/2p}\big), \qquad \delta_{\mathrm{HL}} = O\big(\mathcal{N}^{-(2p-1)/(2p+1)}\big). \tag{5.111}$$

By setting $p = 2q+1$, we find that the exponents in Eq. (5.111) becomes the same as those in Eq. (5.110). In fact, the signal $s(t)$ satisfies the condition of (q, M)-regularity with probability one unless p is an odd number. When $1 < p < 3$, for

example, Eq. (3.41) implies the stochastic σ-Hölder continuity:

$$\langle |\Delta_\varepsilon s(t)|^2 \rangle = \langle |s(t+\varepsilon) - s(t)|^2 \rangle \leq C|\varepsilon|^{p-1} = C|\varepsilon|^{2q} \tag{5.112}$$

for every $t \in \mathbb{R}$. Therefore, the long-time average $\|\Delta_\varepsilon s\|_2^2 = \int_0^T \frac{dt}{T} |\Delta_\varepsilon s(t)|^2$ converges to $C|\varepsilon|^{2q}$, and thus becomes almost surely (q, M)-regular for $M > C^{1/2}$.

5.4.2 Extension to higher dimensions

For the final remark of this chapter, we consider a phase function $\varphi(\boldsymbol{x})$ defined over a D-dimensional box $\boldsymbol{x} \in [0, L]^D$ with the periodic boundary condition. The error measure and the regularity of the function estimation for the one-dimensional case can be generalized in a straightforward way. In fact, the error can be measured on the basis of the mean-square error

$$\mathrm{MSE}(\varphi^*, \varphi) = \int \frac{d^D \boldsymbol{x}}{L^D} |\varphi^*(\boldsymbol{x}) - \varphi(\boldsymbol{x})|^2. \tag{5.113}$$

For $q = m + \sigma$, the phase function $\varphi(\boldsymbol{x})$ is (q, M)-regular if every mth partial derivative f of φ satisfies the weak σ-Hölder continuity:

$$\sup_{\varepsilon \neq \boldsymbol{0}} \int \frac{d^D \boldsymbol{x}}{L^D} \frac{|f(\boldsymbol{x} + \boldsymbol{\varepsilon}) - f(\boldsymbol{x})|^2}{\|\boldsymbol{\varepsilon}\|^{2\sigma}} \leq M^2. \tag{5.114}$$

We conjecture that the SQL and the Heisenberg limit for D-dimensional phase functions are given by

$$\delta_{\mathrm{SQL}} = O\big((L^D M^{D/q} N^{-1})^{q/(2q+D)}\big), \quad \delta_{\mathrm{HL}} = O\big((L^D M^{D/q} N^{-1})^{q/(q+D)}\big). \tag{5.115}$$

In fact, we expect that the position-state method is applicable to the multidimensional case, in which the n_1^D sample points must be taken from the box $[0, L]^D$ and the total number of particles is $N = n_1^D n_2$. If we suppose that the optimal error can be obtained from the trade-off relation in Eqs. (5.25) and (5.29), we can obtain Eq. (5.115).

We also expect that the wavenumber-state method is applicable as well. If this is the case, the wavenumber becomes a D-dimensional integer vector, and the Hilbert space $\mathcal{H}$ satisfies $\dim \mathcal{H} = O(K^D)$. Therefore, the QST error must be generalized to $\delta_{\mathrm{QST}} = O(K^{D/2} N^{-1/2})$ in Eq. (5.56) and $\delta_{\mathrm{QST}} = O(K^{D/2} n_1^{-1/2})$ in Eq. (5.60), which also yields Eq. (5.115).

Chapter 6

Quantum edge detection

In this chapter, we propose a practical application of quantum metrology for the estimation of continuous data. In particular, we provide a physical setup of quantum measurement that can efficiently detect edges of a phase function. We also calculate the error bounds of the measurement, where we show that the error can be significantly reduced by quantum correlation. We note that we basically consider probe states with distinguishable particles, (e.g. photons labeled with different time) whose wavefunctions are Gaussian, while we also provide preliminary results on non-Gaussian wavefunctions and indistinguishable photons.

This means that the signal-to-noise ratio can be significantly reduced by use of an appropriate quantum probes. Since edges mark anomalous structures in functions, we argue that quantum correlations in edge detection may help us efficiently discover new phenomena.

Edge detection is not only of practical importance but also of theoretical interest. We show that in the quantum edge detection, error can be bounded by the uncertainty relation between position and momentum, which is one of the fundamental characteristics in quantum mechanics. Moreover, we apply the theory of function estimation in Chapter 5 to the quantum edge detection to demonstrate that the quantum scheme reaches the Heisenberg limit only in the region where the number of photon is scarce and the lengthscale is sufficiently small.

6.1 Physical Setup

6.1.1 Edges and wavelets

First, we mathematically define edges by using the concept of wavelets following Refs. [50, 51]. In a one-dimensional function $\varphi(x)$, edges appear as a large change of its value when viewed with a fixed lengthscale s. To discard unwanted fluctuations at lengthscales shorter than s, we first consider smoothing the function by the convolution:

$$\varphi_s(x) = \varphi(x) * \rho_s(x) = \int_{-\infty}^{\infty} dx\varphi(y)\rho_s(x-y). \tag{6.1}$$

For the rest of this section, the limits of the integral are set to be $\pm\infty$ unless otherwise specified.

The function ρ_s is called a *smoothing function* at scale s. We require that the smoothing function ρ_s is a probability density function with mean 0 and variance s^2. The family of scaling functions $\{\rho_s\}_{s>0}$ is obtained from rescaling the unit-scale function ρ_1:

$$\rho_s(x) = s^{-1}\rho_1(s^{-1}x). \tag{6.2}$$

A classic example of smoothing function is a Gaussian function:

$$\rho_s(x) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{x^2}{2s^2}}. \tag{6.3}$$

After obtaining the smoothed function φ_s, edges are located as the positions at which φ_s changes more rapidly than elsewhere. This can be found as the extrema of the first derivative of φ_s:

$$\varphi_s'(x) = \rho_s'(x) * \varphi(x), \tag{6.4}$$

or the zeroes of the second derivative:

$$\varphi_s''(x) = \rho_s''(x) * \varphi(x). \tag{6.5}$$

These derivatives can be directly calculated by convolving the function φ with ρ_s' or ρ_s'' without referring to the smoothed function φ_s.

Here, we focus on the first derivative ρ_s'. A family of functions $\{\psi_s\}_{s>0}$ can be defined by

$$\psi_s(x) = c_s\rho_s'(x), \tag{6.6}$$

where $c_s > 0$ is some normalization constant depending on s. Such ψ_s is called a *wavelet function* at scale s, and the result of convolution

$$W_s[\varphi](x) = \varphi(x) * \psi_s(x), \tag{6.7}$$

is referred to as the *wavelet transform* of φ. Therefore, the edges of scale s can be detected if we can find the extrema of the wavelet transform $W_s[\varphi]$.

As a side note, the second derivatives $\rho_s''(x)$ of the smoothing function are also often used as wavelets, especially in computational signal analysis. In this case, the edges correspond to the zeroes of the wavelet transform, which are easier to handle on computers than the extrema.

6.1.2 Measurement of wavelet transform

Let us consider the phase function $\varphi(x)$ in a unitary evolution U_φ:

$$U_\varphi|x\rangle = e^{i\varphi(x)}|x\rangle. \tag{6.8}$$

In this chapter, the particles in this systems are considered to be photons, which are the most commonly used in imaging. Furthermore, we do not introduce internal states as we did in Chapter 5, since the internal state $|x;-\rangle$ in Eq. (5.2) is

not necessary for the purpose of this chapter. The goal is to estimate the wavelet transform $W_s[\varphi](x)$ by quantum measurements, where the wavelet function ψ_s can be written in terms of the smoothing function ρ_s as in Eq. (6.6).

We set the wavefunction $q(x)$ of a one-photon probe state as follows:

$$q(x) = \sqrt{\rho_s(x_0 - x)}. \tag{6.9}$$

The wavefunction of the probe state $|q_\varphi\rangle = U_\varphi|q\rangle$ after the evolution is

$$q_\varphi(x) = e^{i\varphi(x)}q(x). \tag{6.10}$$

Now, we measure the momentum $p^* = p^*(x_0)$ of a probe state $|q_\varphi\rangle$. The expected momentum $\langle p^*\rangle$ can be straightforwardly calculated as

$$\begin{aligned}\langle p^*\rangle &= -i\hbar \int dx \overline{q_\varphi(x)} q'_\varphi(x) \\ &= -i\hbar \int dx \overline{q(x)}[i\varphi'(x)q(x) + q'(x)] \\ &= -i\hbar \int dx \overline{q(x)} q'(x) + \hbar \int dx \varphi'(x)|q(x)|^2. \end{aligned} \tag{6.11}$$

Here, the first term in the last line in Eq. (6.11) vanishes since $\overline{q(x)}q'(x) = -\frac{1}{2}\rho'_s(x_0 - x)$ and the smoothing function $\rho_s(x)$ vanishes at $x \to \pm\infty$. Hence, the expected momentum of the output probe is

$$\langle p^*\rangle = \hbar(\rho_s * \varphi')(x_0) = \hbar(\rho'_s * \varphi)(x_0) = \frac{\hbar}{c_s}(\psi_s * \varphi)(x_0), \tag{6.12}$$

which is in proportion to the wavelet transform $W_s[\varphi](x_0)$.

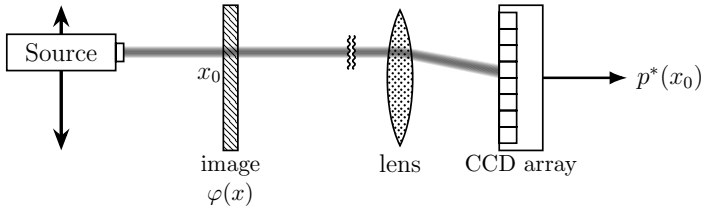
In general, we can prove the following proposition:

Proposition 6.1 (Wavelet measurement) *If the initial probe state is chosen such that (i) the position distribution is $\rho_s(x_0 - x)$ and (ii) the expected momentum is 0, the expected momentum after the unitary evolution by U_φ is in proportion to the wavelet transform $W_s[\varphi](x_0)$.*

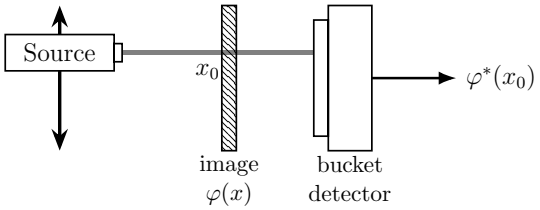
This can be easily confirmed from Eq. (6.11). Although a pure probe state is preferred for an accurate measurement, this proposition holds even for a mixed probe state. Furthermore, multiple photons with quantum correlation may also be used as the probe state, which we will show in Sec. 6.2.3.

6.1.3 Wavelet measurement by imaging techniques

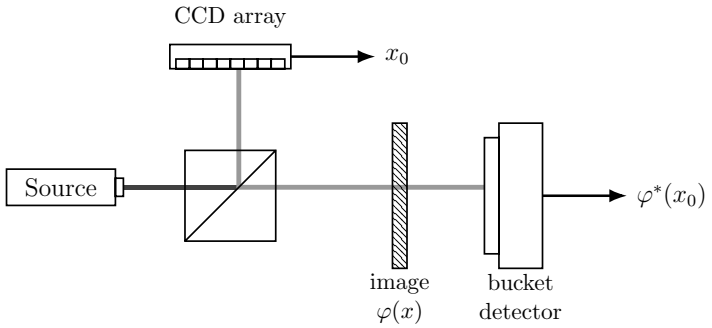
We illustrate in Fig. 6.1 (a) the physical implementation of the wavelet measurement, which is similar to imaging by a scanner in Fig. 6.1 (b). In both cases, the beam source is moved to various positions x_0, where the measurement outcome corresponds to the function value at x_0.



(a) Wavelet measurement



(b) Imaging by a scanner



(c) Ghost imaging

Figure 6.1: Optical settings explained in Sec. 6.1.3. **(a)** Wavelet measurement. The beam is localized at various positions x_0 and the momentum $p^* = p^*(x_0)$ is measured by the Fraunhofer diffraction. **(b)** Imaging by a scanner, on which the wavelet measurement is based. The output signal collected by the bucket detector yields an estimator of the function value $\varphi(x_0)$. **(c)** A more sophisticated setup called *ghost imaging*, where the source is split into two correlated beams and one of them is measured by a CCD array. The estimator of the function φ can be obtained by synthesizing the outputs of both the CCD array and the bucket detector.

There are two differences between the wavelet measurement in Fig. 6.1 (a) and the imaging in Fig. 6.1 (b). One of them is relatively minor; the beam in the wavelet measurement is considered to have a finite beamwidth s, while it is regarded as a zero-width beam in the ordinary measurement. The other, major difference is the measurement device; in the ordinary imaging, we employ a bucket detector that measures the intensity of the beam, from which the original function $\varphi(x_0)$ can be estimated. In the wavelet measurement, we need to resolve the momentum of the beam to obtain the wavelet coefficient $W_s[\varphi](x_0)$. A simple way to measure the momentum is to condense the beam with a lens at a sufficiently large distance and measure the position by a CCD array. The detected position corresponds to the momentum by virtue of the Fraunhofer diffraction.

For better spatial resolution, we may incorporate *ghost imaging*, an imaging technique to localize the probe beam at an accurate position [52, 86], which has also been utilized in the experiment of quantum metrology [53]. In ghost imaging, we use a pair of correlated photons, which may be classical correlation [87] or quantum correlation [86]. As illustrated in Fig. 6.1 (c), one of the pair is injected into the physical image and detected by the bucket detector, while the other photon is captured by a scanning detector such as a CCD array. Owing to the correlated positions of the two photons, the probe beam through the physical image is conditioned by the output of the CCD array, which results in a localized beam. In this way, one can combine the intensity resolution of the bucket detector and the spatial resolution of the CCD array to obtain an accurate image. We expect that the same conditional probe beam can be used in the wavelet measurement.

6.2 Estimation Error of Wavelet Measurement

6.2.1 Probability density of the measured momentum

In this section, we derive the estimation error of the measured momentum p^*, which characterizes the precision of the wavelet measurement. We begin by the case in which the initial state consists of only one photon, and obtain the error by directly calculating the probability density of p^* by means of the Wigner distribution. First, we denote the Fourier transformation $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$ as

$$\mathcal{F}_{p\to x}[f(p)] = \int dp e^{-ipx/\hbar} f(p), \tag{6.13}$$

$$\mathcal{F}^{-1}_{x\to p}[g(x)] = \int \frac{dx}{2\pi\hbar} e^{ipx/\hbar} g(x). \tag{6.14}$$

Then, the Wigner function W_q associated with the wavefunction $q(x)$ is

$$W_q(x,p) = \mathcal{F}^{-1}_{y\to p}\left[q\left(x-\frac{y}{2}\right)\overline{q}\left(x+\frac{y}{2}\right)\right]. \tag{6.15}$$

Now, we focus on the Wigner function associated with the wavefunction $q_\varphi(x) = e^{i\varphi(x)}q(x)$ after the unitary evolution. Here, we recall the convolution theorem:

The product and the convolution exchange their roles via the Fourier transform. To be more specific, we have

$$\mathcal{F}_{p\to x}[f(p) * g(p)] = \mathcal{F}_{p\to x}[f(p)]\mathcal{F}_{p\to x}[g(p)], \tag{6.16}$$

where $*$ is the convolution with respect to the variable p.

Then, starting from the definition, we obtain

$$\begin{aligned} W_{q_\varphi}(x,p) &= \mathcal{F}^{-1}_{y\to p}\left[e^{i\varphi(x-\frac{y}{2})}q\left(x-\frac{y}{2}\right)e^{-i\varphi(x+\frac{y}{2})}\overline{q}\left(x+\frac{y}{2}\right)\right] \\ &= \mathcal{F}^{-1}_{y\to p}\left[e^{i[\varphi(x-\frac{y}{2})-\varphi(x+\frac{y}{2})]}q\left(x-\frac{y}{2}\right)\overline{q}\left(x+\frac{y}{2}\right)\right] \\ &= \mathcal{F}^{-1}_{y\to p}\left[e^{i[\varphi(x-\frac{y}{2})-\varphi(x+\frac{y}{2})]}\right] * \mathcal{F}^{-1}_{y\to p}\left[q\left(x-\frac{y}{2}\right)\overline{q}\left(x+\frac{y}{2}\right)\right] \\ &= W_\varphi(p|x) * W_q(x,p), \end{aligned} \tag{6.17}$$

where we have defined

$$W_\varphi(p|x) = \mathcal{F}^{-1}_{y\to p}\left[e^{i[\varphi(x-\frac{y}{2})-\varphi(x+\frac{y}{2})]}\right]. \tag{6.18}$$

Therefore, the momentum p^* of the wavefunction $q_\varphi(x)$ is distributed according to the density

$$P(p^*) = \int dx W_{q_\varphi}(x,p^*) = \iint dxdp W_q(x,p^*-p)W_\varphi(p|x). \tag{6.19}$$

In this way, the probability density of p^* is expressed in an integral form, in which the dependencies on the initial wavefunction $q(x)$ and the phase function $\varphi(x)$ are separated from each other.

The problem becomes even simpler if the Wigner function $W_q(x,p)$ is a direct product of the position and the momentum distributions. As far as pure states are concerned, this happens only when $q(x)$ is Gaussian [88]. To be more specific, if we set

$$q(x) = \sqrt{\rho_s(x_0-x)}, \qquad \rho_s(x) = \frac{1}{\sqrt{2\pi}s}e^{-x^2/2s^2}, \tag{6.20}$$

the corresponding Wigner function is

$$W_q(x,p) = \frac{1}{2\pi\hbar}\exp\left[-\frac{(x-x_0)^2}{2s^2}\right]\exp\left[-\frac{s^2p^2}{2\hbar^2}\right] = \rho_s(x_0-x)\rho_{\hbar/2s}(p). \tag{6.21}$$

Then, the probability density $P(p^*)$ can be expressed as

$$P(p^*) = \iint dxdp\rho_{\hbar/2s}(p^*-p)W_\varphi(p|x)\rho_s(x_0-x). \tag{6.22}$$

In Eq. (6.22), the measured momentum p^* appears to be the outcome of the composition of three stochastic processes: The first process $x \mapsto x_0$ and the last

process $p \mapsto p^*$ are the addition of Gaussian noises with standard deviation s and $\hbar/2s$, respectively. The second process can be regarded as a Markovian process $x \mapsto p$ with the transition probability density $W_\varphi(p|x)$.

The second process $x \mapsto p$, however, is not a stochastic process in the true sense. Although the function $W_\varphi(p|x)$ is normalized with respect to p as

$$\int dp W_\varphi(p|x) = \lim_{y\to 0} e^{i[\varphi(x-\frac{y}{2})-\varphi(x+\frac{y}{2})]} = 1, \tag{6.23}$$

the nonnegativity $W_\varphi(p|x) \geq 0$ does not hold unless the phase function $\varphi(x)$ is constant. In this respect, we call the mapping $x \mapsto p$ a *quasi-Markovian* process with the transition quasiprobability density $W_\varphi(p|x)$, which is not necessarily, and in general is not, positive.

6.2.2 Error bound for one-photon probe states

Let us recall the probability density of the measured momentum p^* in Eq. (6.19):

$$P(p^*) = \iint dx dp W_q(x, p^* - p) W_\varphi(p|x). \tag{6.24}$$

The statistical error δ of the measured momentum p^* is

$$\begin{aligned}
\delta^2 &= \int dp^* (p^* - \langle p^* \rangle)^2 P(p^*) \\
&= \iiint (p^* - \langle p^* \rangle)^2 dp^* dx dp W_q(x, p^* - p) W_\varphi(p|x) \\
&= \iiint (p' + p - \langle p^* \rangle)^2 dp' dx dp W_q(x, p') W_\varphi(p|x).
\end{aligned} \tag{6.25}$$

Here, we encounter the first and the second moments of the quasi-stochastic variable p subject to the quasiprobability density $W_\varphi(p|x)$. In fact, these moments can be analytically calculated:

$$\int p dp W_\varphi(p|x) = \lim_{y\to 0} \left(i\hbar \frac{\partial}{\partial y} e^{i[\varphi(x-\frac{y}{2})-\varphi(x+\frac{y}{2})]} \right) = \hbar \varphi'(x); \tag{6.26}$$

$$\int p^2 dp W_\varphi(p|x) = \lim_{y\to 0} \left(-\hbar^2 \frac{\partial^2}{\partial y^2} e^{i[\varphi(x-\frac{y}{2})-\varphi(x+\frac{y}{2})]} \right) = [\hbar \varphi'(x)]^2. \tag{6.27}$$

This implies that the variance of p is zero, which marks the oddness of the quasi-stochastic variable. Hence, Eq. (6.25) becomes

$$\begin{aligned}
\delta^2 &= \iint (p' + \hbar\varphi'(x) - \langle p^* \rangle)^2 dp' dx W_q(x, p') \\
&= \iint (p + \hbar\varphi'(x) - \langle p^* \rangle)^2 dx dp W_q(x, p).
\end{aligned} \tag{6.28}$$

With a Gaussian probe in Eq. (6.20), p^* is regarded as the output of a chain of quasi-Markovian processes: $x_0 \mapsto x \mapsto p \mapsto p^*$, where the second process $x \mapsto p$

does not contribute to the statistical error. Hence, the error can be decomposed into two parts:

$$\delta^2 = E_p^2 + E_x^2, \tag{6.29}$$

$$E_p^2 = \int p^2 dp \rho_{\hbar/2s}(p) = \frac{\hbar^2}{4s^2}, \tag{6.30}$$

$$E_x^2 = \int [\hbar\varphi'(x) - \langle p^* \rangle]^2 dx \rho_s(x - x_0). \tag{6.31}$$

Here, E_x is originated from the process $x_0 \mapsto x$ and E_p is originated from the process $p \mapsto p^*$; we call these terms the *x-error* and the *p-error*, respectively.

Furthermore, if we assume that the phase function φ is sufficiently smooth and the linear approximation $\hbar\varphi'(x) = \varphi'(x_0) + (x - x_0)\varphi''(x_0)$ is valid, we also obtain the x-error as

$$E_x^2 \approx \int [\hbar\varphi''(x_0)]^2 (x_0 - x)^2 dx \rho_s(x - x_0) = [\hbar\varphi''(x_0)s]^2. \tag{6.32}$$

Therefore, an overall error is given by the sum of Eqs. (6.30) and (6.32), and the magnitude of the second derivative $|\varphi''(x_0)|$ determines which of them is the dominant term[1]. Intuitively, the second derivative $\varphi''(x_0)$ should be small near the edges, since an edge at x_0 corresponds to a zero of the smoothed second derivative $\varphi''(x) * \rho_s(x)$. Hence, we consider that the p-error dominates over the x-error in precisely located edges, i.e. when we search for the extrema of the wavelet transform. This idea will be numerically tested in Sec. 6.2.4.

In the previous section, the error of the wavelet measurement for a single-photon probe has been calculated. In this section, we extend this analysis to a general multiple-photon source, where the presence of quantum correlation conditionally reduces the estimation error.

6.2.3 Wavelet measurement with distinguishable N photons

The wavelet measurement so far explained can be naturally extended to multi-photon probe states in a natural way. We consider the case in which the probe state consists of N photons that are always distinguishable (e.g. when they are labeled with distinct time or frequency). In Sec. 6.4.2, we show preliminary analysis on indistinguishable photons, where similar results are derived.

For distinguishable N photons, the initial state can be written in term of the joint wavefunction $q(\boldsymbol{x}) = q(x_1, \ldots, x_N)$.

We require that each measured momentum p_j^* corresponds to the same wavelet transform:

$$\langle p_j^* \rangle = \frac{\hbar}{c_s}(\psi_s * \varphi)(x_0) = \frac{\hbar}{c_s} W_s[\varphi](x_0). \tag{6.33}$$

[1] We note that s is given as the wavelet scale and fixed.

By Prop. 6.1, this is satisfied if the initial expected momenta are all zero and the marginal distribution of the position x_j coincides with ρ_s:

$$\int_{\mathbb{R}^{N-1}} d\boldsymbol{x}_{\neq j} |q(x_1, \ldots, x_N)|^2 = \rho_s(x_0 - x_j), \tag{6.34}$$

where we denote by $d\boldsymbol{x}_{\neq j}$ the integration over $x_1, \ldots, x_N$ save for x_j. Then, the ultimate estimator p^* can be obtained as their arithmetic mean:

$$p^* = \frac{p_1^* + \cdots + p_N^*}{N} = \frac{1}{N}\mathbf{1}^{\mathrm{T}}\boldsymbol{p}^*, \qquad \mathbf{1} = (1, \ldots, 1). \tag{6.35}$$

The joint probability of the measured momenta $\boldsymbol{p}^* = (p_1^*, \ldots, p_N^*)$ can be represented similarly to Eq. (6.19):

$$P(\boldsymbol{p}^*) = \iint_{\mathbb{R}^N \times \mathbb{R}^N} d\boldsymbol{x} d\boldsymbol{p} W_q^{(n)}(\boldsymbol{x}, \boldsymbol{p}^* - \boldsymbol{p}) W_\varphi^{(n)}(\boldsymbol{p}|\boldsymbol{x}), \tag{6.36}$$

where the Wigner functions are extended to their joint versions:

$$W_q^{(n)}(\boldsymbol{x}, \boldsymbol{p}) = \mathcal{F}_{\boldsymbol{y}\to\boldsymbol{p}}^{-1}\left[q\left(\boldsymbol{x} - \frac{\boldsymbol{y}}{2}\right)\overline{q}\left(\boldsymbol{x} + \frac{\boldsymbol{y}}{2}\right)\right], \tag{6.37}$$

$$W_\varphi^{(n)}(\boldsymbol{p}|\boldsymbol{x}) = W_\varphi(p_1|x_1) \cdots W_\varphi(p_N|x_N). \tag{6.38}$$

We consider the case in which the joint wavefunction is an N-variate Gaussian function:

$$q(\boldsymbol{x}) = \sqrt{\rho(\boldsymbol{x}_0 - \boldsymbol{x}; C_x)}, \tag{6.39}$$

where $\boldsymbol{x}_0 = (x_0, \ldots, x_0)$ and $\rho(\boldsymbol{x}; C)$ is the probability density of the N-variate Gaussian distribution with the covariance matrix C:

$$\rho(\boldsymbol{x}; C) = [(2\pi)^N \det C]^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}} C^{-1} \boldsymbol{x}\right]. \tag{6.40}$$

We investigate the optimal error of the estimator p^* in Eq. (6.35) in the *classical scheme* and the *quantum scheme*: In the former case, photons are uncorrelated and the covariance matrices C_x and C_p are both diagonal; in the latter case, any correlated probe with positive covariance matrices is possible.

Equation (6.34) corresponds to the requirement that the diagonal entries in C_x be all s^2. Furthermore, since the uncertainty relation is satisfied for the Gaussian wavefunction, the covariance matrix for the momenta is equal to $C_p = \frac{\hbar^2}{4} C_x^{-1}$. The joint Wigner function for the Gaussian wavefunction is simply the product of position and momentum distribution:

$$W_q^{(n)}(\boldsymbol{x}, \boldsymbol{p}) = \rho(\boldsymbol{x}_0 - \boldsymbol{x}; C_x)\rho(\boldsymbol{p}; C_p). \tag{6.41}$$

Hence, the joint distribution in Eq. (6.36) can be written as

$$P(\boldsymbol{p}^*) = \iint_{\mathbb{R}^N \times \mathbb{R}^N} d\boldsymbol{x} d\boldsymbol{p} \rho(\boldsymbol{p}^* - \boldsymbol{p}; C_p) W_\varphi^{(n)}(\boldsymbol{p}|\boldsymbol{x}) \rho(\boldsymbol{x}_0 - \boldsymbol{x}; C_x) \tag{6.42}$$

This can be regarded a chain of quasi-Markovian processes $\boldsymbol{x}_0 \mapsto \boldsymbol{x} \mapsto \boldsymbol{p} \mapsto \boldsymbol{p}^*$, where the Gaussian noises in $\boldsymbol{x}_0 \mapsto \boldsymbol{x}$ and $\boldsymbol{p} \mapsto \boldsymbol{p}^*$ are correlated in the quantum scheme.

The estimation error of the measured momenta $\boldsymbol{p}^*$ is given by the covariance matrix $[V]_{jk} = \langle (p_j^* - \langle p^* \rangle)(p_k^* - \langle p^* \rangle) \rangle$, which can also be decomposed into two parts originated from the momentum variance and position variance of the initial probe. A similar argument to the single-photon case leads to

$$V = V_x + V_p, \tag{6.43}$$

$$[V_p]_{jk} = \int p_j p_k d\boldsymbol{p} \rho(\boldsymbol{p}; C_p) = [C_p]_{jk}, \tag{6.44}$$

$$\begin{aligned}[V_x]_{jk} &= \int [\hbar\varphi'(x_j) - \langle p^* \rangle][\hbar\varphi'(x_k) - \langle p^* \rangle] d\boldsymbol{x} \rho(\boldsymbol{x} - \boldsymbol{x}_0; C_x) \\ &\approx \hbar^2 |\varphi''(x_0)|^2 [C_x]_{jk}.\end{aligned} \tag{6.45}$$

Recalling that the estimator p^* is the arithmetic mean of the N momenta (Eq. (6.35)), the estimation error δ can be expressed as

$$\delta^2 = \left\langle (p^* - \langle p^* \rangle)^2 \right\rangle = \frac{1}{N^2} \mathbf{1}^{\mathrm{T}} V \mathbf{1} = E_x^2 + E_p^2, \tag{6.46}$$

where the x-error E_x and the p-error E_p are defined by

$$\begin{aligned}E_x^2 &= \frac{1}{N^2} \mathbf{1}^{\mathrm{T}} V_x \mathbf{1} \approx \frac{\hbar^2 |\varphi''(x_0)|^2}{N^2} \mathbf{1}^{\mathrm{T}} C_x \mathbf{1}, \\ E_p^2 &= \frac{1}{N^2} \mathbf{1}^{\mathrm{T}} V_p \mathbf{1} = \frac{1}{N^2} \mathbf{1}^{\mathrm{T}} C_p \mathbf{1}.\end{aligned} \tag{6.47}$$

This implies that the estimation error can be suppressed by negative off-diagonal entries in the covariance matrices. However, this effect cannot be simultaneously exploited for both positions and momenta; in fact, the uncertainty relation $C_p = \frac{\hbar^2}{4} C_x^{-1}$ implies

$$(\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}) \cdot (\mathbf{1}^{\mathrm{T}} C_p \mathbf{1}) \geq (\mathbf{1}^{\mathrm{T}} C_p^{1/2} C_x^{1/2} \mathbf{1})^2 = \frac{\hbar^2 N^2}{4}. \tag{6.48}$$

Now, we consider how small the p-error can be, assuming that $|\varphi''(x_0)|$ is sufficiently small. In the classical scheme, the only possibility is $C_x = s^2 I$ and $C_p = \frac{\hbar^2}{4s^2} I$. The p-error is

$$E_p^2 = \frac{\hbar^2}{4s^2 N}, \tag{6.49}$$

which exhibits the *standard quantum limit* (SQL) $E_p = O(N^{-1/2})$. This result is also evident from the p-error for a single photon in Eq. (6.30), since the statistical variance is reduced by a factor of N^{-1} by taking the arithmetic mean of N i.i.d. estimators.

In the quantum scheme, the maximum correlation in the x-error is found to be

$$\mathbf{1}^{\mathrm{T}} C_x \mathbf{1} = \mathrm{tr}[\mathbf{1}\mathbf{1}^{\mathrm{T}} C_x] \leq \|\mathbf{1}\mathbf{1}^{\mathrm{T}}\| (\mathrm{tr}\, C_x) \leq N^2 s^2, \tag{6.50}$$

in which case the minimum p-error is obtained from Eq. (6.48):

$$E_p^2 = \frac{1}{N^2} \mathbf{1}^{\mathrm{T}} C_p \mathbf{1} \geq \frac{1}{N^2} \cdot \frac{\hbar^2 N^2}{4N^2 s^2} = \frac{\hbar}{4s^2 N^2}. \tag{6.51}$$

In fact, if we set the off-diagonal elements to rs^2 for some $0 \leq r < 1$, namely

$$[C_x]_{jk} = s^2[(1-r)\delta_{jk} + r], \tag{6.52}$$

we can calculate the momentum covariance matrix C_p and the p-error E_p:

$$[C_p]_{jk} = \frac{\hbar^2}{4s^2(1-r)} \left[\delta_{jk} - \frac{r}{(1-r) + Nr} \right], \quad E_p^2 = \frac{\hbar^2}{4s^2[(1-r) + Nr]}. \tag{6.53}$$

We note that Eq. (6.51) becomes an equality in the limit of $r \to 1$. Therefore, as long as we ignore the x-error, the Heisenberg limit $\delta = O(N^{-1})$ of the parameter estimation is attainable.

On the other hand, the effect of x-error becomes a serious problem as the p-error approaches the Heisenberg limit. In fact, Eq. (6.48) suggests that the Heisenberg-limited p-error $E_p = O(N^{-1})$ forces the x-error to be of the order of unity, which is much worse than the SQL. In Sec. 6.3.1, we perform a more rigorous analysis on the Heisenberg limit with the x-error taken into consideration.

6.2.4 Numerical calculation

Here, we numerically investigate the influence of the x-error for the wavelet detection of a smooth function $\varphi(x)$ $(0 \leq x \leq 10)$, which is displayed in Fig. 6.2 along with its wavelet transforms. We assume that the input probe has a multivariate Gaussian wavefunction, where the position covariance matrix C_x can be expressed as in Eq. (6.52). Therefore, we have three parameters in the setting: the number N of photons, the wavelet scale s, and the quantum correlation coefficient r.

The naïve calculation of the probability density of the momenta involves $2N$-dimensional integration, which becomes intractable even with small N, say, $N = 5$. This combinatorial explosion can be circumvented by the Monte-Carlo method, especially because the Gaussian wavefunction is considered in our case. However, a large number of samples are required if we are to calculate the probability density of the N-dimensional vector $\boldsymbol{p}^*$. Rather than that, we directly calculate the mean momentum $p^* = \frac{p_1^* + \cdots + p_N^*}{N}$ by using the characteristic function, i.e., the Fourier transform of the probability density:

$$\chi_{\boldsymbol{p}^*}(\boldsymbol{y}) = \mathcal{F}_{\boldsymbol{p}^* \to \boldsymbol{y}}[P(\boldsymbol{p}^*)]. \tag{6.54}$$

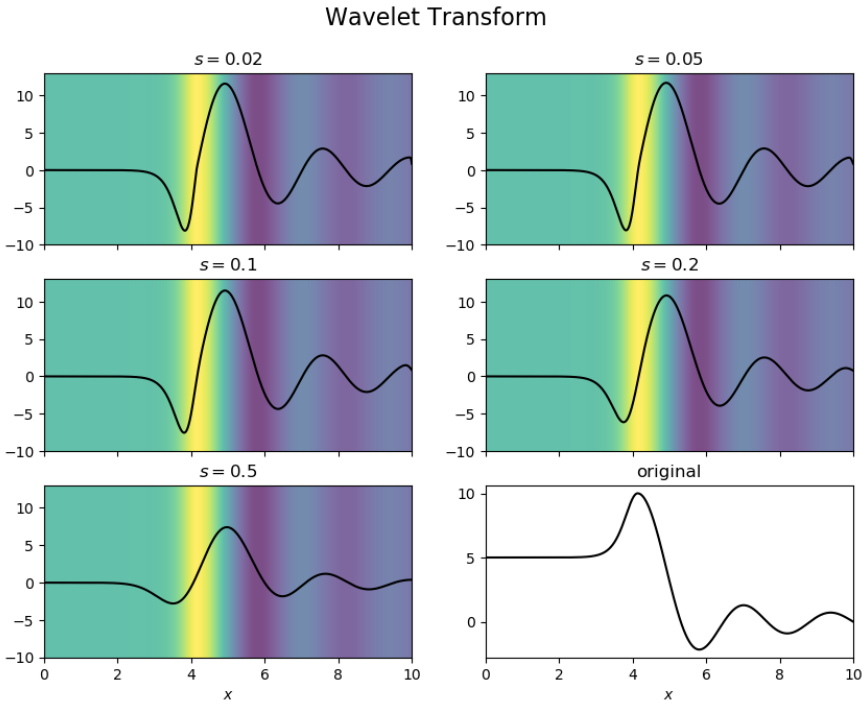


Figure 6.2: A smooth function employed in the numerical calculation and its wavelet transforms. The original function is displayed in the bottom-right pane as well as indicated in the background color of the other panes. In the wavelet transforms, a large maximum emerges around $x = 5$ regardless of the wavelet scale.

Starting from (6.42), we obtain

$$\begin{aligned}\chi_{\boldsymbol{p}^*}(\boldsymbol{y}) &= \int_{\mathbb{R}^N} d\boldsymbol{x}\rho(\boldsymbol{x}_0-\boldsymbol{x};C_x)\mathcal{F}_{\boldsymbol{p}\to\boldsymbol{y}}[\rho(\boldsymbol{p};C_p)*W_\varphi^{(n)}(\boldsymbol{p}|\boldsymbol{x})]\\ &= \int_{\mathbb{R}^N} d\boldsymbol{x}\rho(\boldsymbol{x}_0-\boldsymbol{x};C_x)\mathcal{F}_{\boldsymbol{p}\to\boldsymbol{y}}[\rho(\boldsymbol{p};C_p)]\mathcal{F}_{\boldsymbol{p}\to\boldsymbol{y}}[W_\varphi^{(n)}(\boldsymbol{p}|\boldsymbol{x})]\\ &= \int_{\mathbb{R}^N} d\boldsymbol{x}\rho(\boldsymbol{x}_0-\boldsymbol{x};C_x)e^{-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}C_x\boldsymbol{y}}\prod_{j=1}^{N}e^{i[\varphi(x_j-\frac{1}{2}y_j)-\varphi(x_j+\frac{1}{2}y_j)]}. \qquad (6.55)\end{aligned}$$

Since the mean momentum p^* is linearly dependent on $\boldsymbol{p}^*$, the characteristic function of p^* can be calculated from Eq. (6.55) as

$$\begin{aligned}\chi_{p^*}(y) &= \chi_{\boldsymbol{p}^*}\left(\frac{y}{N},\cdots,\frac{y}{N}\right)\\ &= \int_{\mathbb{R}^N} d\boldsymbol{x}\rho(\boldsymbol{x}_0-\boldsymbol{x};C_x)e^{-\frac{y^2}{2}(\mathbf{1}^{\mathrm{T}}C_x\mathbf{1})}\prod_{j=1}^{N}e^{i[\varphi(x_j-\frac{y}{2N})-\varphi(x_j+\frac{y}{2N})]}\\ &= \int_{\mathbb{R}^N} d\boldsymbol{x}\rho(\boldsymbol{x};C_x)e^{-\frac{y^2}{2N^2}(\mathbf{1}^{\mathrm{T}}C_x\mathbf{1})}\prod_{j=1}^{N}e^{i[\varphi(x_0-x_j-\frac{y}{2N})-\varphi(x_0-x_j+\frac{y}{2N})]}, \qquad (6.56)\end{aligned}$$

which can be calculated by the Monte Carlo method on the N-variate Gaussian variable $\boldsymbol{x}$. The probability density $P(p^*)$ can be calculated by taking the inverse Fourier transform of the characteristic function.

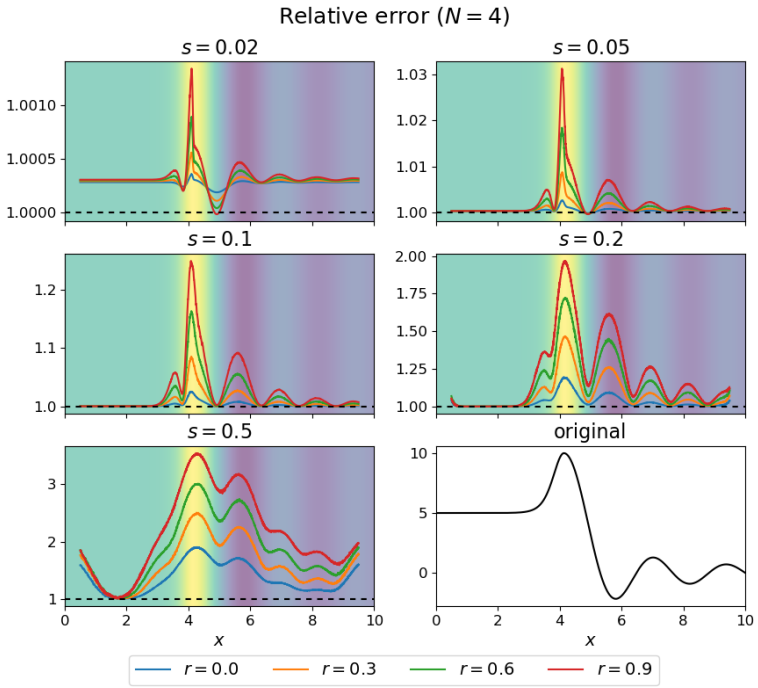
We have calculated the probability density for positions x_0 away from the edge points[2]. The product in Eq. (6.56) can be efficiently calculated for multiple instances of x_0 by a vector processor.

The *relative error* δ/E_p is obtained by taking the ratio of the statistical error δ to the theoretical p-error E_p in Eq. (6.53), so that the x-error is negligible if the relative error is close to the unity. The relative error is plotted in Fig. 6.3 for various configurations of (N,s,r). This figure indicates the conditions with which the x-error becomes significant:

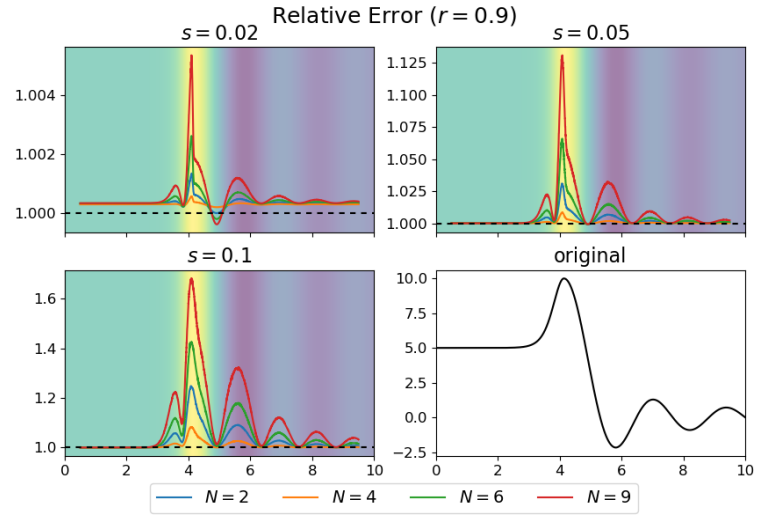
- The x-error is negligible where $|\varphi''(x)|$ is very small, which can be typically observed in the region $0\le x\le 3$ and around $x=5$. We particularly note that the edge around $x=5$ can be efficiently detected without suffering from a large x-error. Conversely, the x-error becomes significant where $|\varphi''(x)|$ is large, which is mainly observed in the peaks of the original function around, e.g. $x=4$ or $x=6$.

- Where the x-error is significant, the relative error increases with the number n of photons and/or the quantum correlation r. This is indicated by the fact that the x-error becomes large as the p-error approaches the Heisenberg limit.

[2]We have also checked that the bias of the mean momentum was negligible compared with the variance, indicating that the number of samples used for the Monte Carlo method is sufficient.



(a)



(b)

Figure 6.3: Position dependence of relative errors δ/E_p, where **(a)** the wavelet scale s and the quantum correlation r are varied while the number of photon is fixed to $n = 4$, and **(b)** the wavelet scale s and the number n of photons are varied while the quantum correlation is fixed to $r = 0.9$. The original function $\varphi(x)$ is displayed in the bottom-right panes as well as indicated by the background color in the other panes. The dashed line indicates the unity; the relative error tends to be on this line in the region where the x-error is negligible.

- As the wavelet scale s increases, the x-error begins to dominated in a wider region, resulting in a notably large error as a whole. For such a large scale as $s = 0.5$, not only the relative error δ/E_p but the absolute error δ itself becomes large by increasing the quantum correlation. This suggests that the Heisenberg limit can be observed only when the wavelet scale s is relatively small, depending on the magnitude of the quantum correlation.

6.3 Function Analysis on Wavelet Measurement

6.3.1 Regularity and the x-error

In Sec. 6.2, we have investigated the effect of quantum correlation on the precision of the edge detection both analytically and numerically. We have found that the statistical error can be decomposed into two parts: the x-error and the p-error. The p-error is subject to the SQL $E_p = O(N^{-1/2})$ for the classical scheme and the Heisenberg limit $E_p = O(N^{-1})$ for the quantum scheme. On the other hand, we have also found the drawback of the wavelet measurement: the x-error behaves complemetarily with the p-error, and the x-error can grow intolerably large as the p-error approaches to the Heisenberg limit. Therefore, we need to ensure how far the quantum enhancement can be exploited in the wavelet measurement by using the functional analysis similar to Chap. 5.

For a single-photon probe, the x-error E_x can be calculated from Eq. (6.31):

$$E_x^2 = R(x_0) := \int [\hbar\varphi'(x) - \langle p^* \rangle]^2 dx \rho_s(x_0 - x), \tag{6.57}$$

which involves the nonlinear transformation of the Gaussian distribution, making the analytical calculation intractable in general. However, by replacing $\langle p^* \rangle$ with $\hbar\varphi'(x_0)$ in Eq. (6.57), we obtain an upper bound on $R(x_0)$:

$$\begin{aligned} R(x_0) &\leq \hbar^2 \int [\varphi'(x) - \varphi'(x_0)]^2 dx \rho_s(x_0 - x) \\ &= \hbar^2 \int [\varphi'(x_0 + \varepsilon) - \varphi'(x_0)]^2 d\varepsilon \rho_s(\varepsilon). \end{aligned} \tag{6.58}$$

Hence, the x-error depends on the magnitude of the difference $|\varphi'(x_0 + \varepsilon) - \varphi'(x_0)|$ when ε is typically of the order of s. This can be quantified by recourse to an appropriate regularity on the function $\varphi(x)$.

Let us consider the function $\varphi(x)$ over an arbitrary interval $0 \leq x \leq L$ under the periodic boundary condition. The appropriate regularity of the function φ is $(1 + \sigma, M)$-regularity for some $0 < \sigma \leq 1$ (See Def. 5.2):

$$\sup_{\varepsilon \neq 0} \int_0^L \frac{dx}{L} \frac{|\varphi'(x + \varepsilon) - \varphi'(x)|^2}{|\varepsilon|^{2\sigma}} \leq M^2, \tag{6.59}$$

where we obtain

$$\int_0^L \frac{dx_0}{L} R(x_0) \leq \hbar^2 \int_0^L \frac{dx}{L} \int d\varepsilon \left|\varphi'(x_0+\varepsilon) - \varphi'(x_0)\right|^2 \rho_s(\varepsilon)$$
$$\leq \hbar^2 \int d\varepsilon M^2 |\varepsilon|^{2\sigma} \rho_s(\varepsilon) = \frac{\Gamma(\sigma+\frac{1}{2})}{\sqrt{\pi}} \hbar^2 M^2 s^{2\sigma}. \tag{6.60}$$

We consider the squared x-error $\overline{E_x^2}$ averaged over the interval $0 \leq x \leq L$. By combining Eqs. (6.57) and (6.60), we find that this average is bounded from above as $\overline{E_x^2} \leq \hbar^2 C^2 M^2 s^{2\sigma}$, where $C = \pi^{-1/4}[\Gamma(\sigma+\frac{1}{2})]^{1/2}$.

For the multi-photon probe with position covariance matrix C_x, the x-error can be written as

$$E_x^2 = \int \left[\hbar \frac{\varphi'(x_1) + \cdots + \varphi'(x_N)}{N} - \langle p^* \rangle\right]^2 dx \rho(\boldsymbol{x}_0 - \boldsymbol{x}; C_x) \tag{6.61}$$

instead of Eq. (6.57). Here, we assume the following approximation:

$$E_x^2 \approx \frac{\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}}{s^2 N^2} R(x_0), \tag{6.62}$$

where $R(x_0)$ is defined in Eq. (6.57). In fact, this approximation can be justified when C_x is a block-diagonal matrix whose nonzero entries are all close to s^2:

$$C_x = s^2 \begin{pmatrix} 1 & 1-\epsilon & \cdots & 1-\epsilon & & & & \\ 1-\epsilon & 1 & \ddots & \vdots & & & 0 & \\ \vdots & \ddots & \ddots & 1-\epsilon & & & & \\ 1-\epsilon & \cdots & 1-\epsilon & 1 & & & & \\ & & & & \ddots & & & \\ & & & & & 1 & 1-\epsilon & \cdots & 1-\epsilon \\ & & & & & 1-\epsilon & 1 & \ddots & \vdots \\ & 0 & & & & \vdots & \ddots & \ddots & 1-\epsilon \\ & & & & & 1-\epsilon & \cdots & 1-\epsilon & 1 \end{pmatrix}. \tag{6.63}$$

In particular, we obtain the the equality $E_x^2 = \frac{\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}}{s^2 N^2} R(x_0)$ when $x_1, \ldots, x_N$ are mutually independent (i.e., $C_x = s^2 I$) or when $x_1, \ldots, x_N$ is completely correlated (i.e., $C_x = s^2 \mathbf{1}\mathbf{1}^{\mathrm{T}}$). We note that assuming such a special matrix as in Eq. (6.62) does not affect the error bound except for an additional factor of 2; see Appendix D for the rigorous proof.

With the approximation (6.62) in mind, the *root mean square* (RMS) x-error over the interval $0 \leq x \leq L$ is evaluated as

$$\overline{E_x^2} := \int_0^L \frac{dx_0}{L} E_x^2 \lesssim \hbar^2 C^2 M^2 \frac{\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}}{N^2 s^{2-2\sigma}}. \tag{6.64}$$

On the other hand, the p-error in Eq. (6.49) is uniform regardless of the function φ or the position x_0:

$$E_p^2 = \frac{\mathbf{1}^{\mathrm{T}} C_p \mathbf{1}}{N^2}. \tag{6.65}$$

Therefore, the upper bound on the estimation error is

$$\delta^2 = E_p^2 + \overline{E_x^2} \lesssim \frac{1}{N^2}\left(\hbar^2 C^2 M^2 \frac{\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}}{s^{2-2\sigma}} + \mathbf{1}^{\mathrm{T}} C_p \mathbf{1}\right). \tag{6.66}$$

In the classical scheme, the two terms $\mathbf{1}^{\mathrm{T}} C_x \mathbf{1} = s^2 N$ and $\mathbf{1}^{\mathrm{T}} C_p \mathbf{1} = \frac{\hbar^2}{4s^2} N$ are fixed. Therefore, the error of the wavelet measurement is

$$\delta^2 \lesssim \frac{\hbar^2}{4s^2 N}(1 + 4C^2 M^2 s^{2+2\sigma}), \tag{6.67}$$

from which we see that the error is dominated by E_p only if the scale is sufficiently small such that $s^{1+\sigma} \leq 2CM$. Regardless of the scale s, the N-dependency of the error is the same as the SQL for parameter estimation: $\delta = O(N^{-1/2})$.

In the quantum scheme, on the other hand, the two terms $\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}$ and $\mathbf{1}^{\mathrm{T}} C_p \mathbf{1}$ are free variables while they are subject to the constraints (6.48) and (6.50). Suppose that the equality in Eq. (6.50) holds. Then, we have

$$\delta^2 \lesssim \frac{\hbar^2}{4s^2}\left(\frac{1}{N^2} + 4C^2 M^2 s^{2+2\sigma}\right). \tag{6.68}$$

Hence, the Heisenberg limit $\delta^2 \approx \frac{\hbar^2}{4s^2 N^2}$ derived in Sec. 6.2 holds only if $s \leq s_{\mathrm{cr}}$, where the critical scale s_{cr} is given by $s_{\mathrm{cr}}^{1+\sigma} = 2CMN^{-1}$. Since s_{cr} becomes smaller as the number N of photons gets larger, the Heisenberg limit $\delta = O(N^{-1})$ does not hold in the limit of $N \to \infty$. This qualitatively explains the numerical results that the statistical error deteriorates from the Heisenberg limit for large N.

When $s \geq s_{\mathrm{cr}}$, we need to consider the tradeoff relation between $\overline{E_x^2}$ and E_p^2 to derive the optimal error. In other words, the quantum correlation must be adjusted so that the x-error and the p-error are comparable to each other. By using Eq. (6.48), we obtain the optimal error

$$\delta^2 \approx \hbar^2 \frac{CM}{N s^{1-\sigma}}. \tag{6.69}$$

Therefore, the error eventually becomes the SQL $\delta = O(N^{-1})$, while they are better than Eq. (6.67) by a constant factor.

6.3.2 Edge detection and function estimation

We have seen the existence of the SQL $\delta = (N^{-1/2})$ and the Heisenberg limit $\delta = O(N^{-1})$ in the wavelet measurement for a fixed wavelet scale s. In this subsection, we consider the edge detection for multiple scales s and establish the results consistent with the function estimation in Chap. 5.

In order to locate the edges in the function $\varphi(x)$, one needs to perform the wavelet measurements for various locations x_0 and search for the extrema. No general strategy for searching extrema exists, although some useful heuristics are known such as Canny's algorithm [46] and the kernel method [89].

Here, we consider the brute-force search where all possible locations x_0 are searched. Noting that an edge of scale s should be located with precision at most $O(s)$, we need to check $O(s^{-1})$ locations per unit length. Hence, the total number of photons consumed per unit length can be written as

$$\mathcal{N} = O(s^{-1}N). \tag{6.70}$$

Mallat and Zhong [50, 51] have shown that the entire function φ can be restored from the locations x_0 and strengths $W_s[\varphi](x_0)$ of edges, where the lengthscale s is taken geometrically:

$$s = 2^j s_0 \quad (j \in \mathbb{Z}). \tag{6.71}$$

Let us recall that the wavelet function ψ_s is defined as $\psi_s(x) = c_s\rho_s'(x)$ with the smoothing function ρ_s and a constant $c_s > 0$. For the measured momentum p^*, the wavelet transform can be calculated as

$$W_s[\varphi](x_0) = \frac{c_s}{\hbar}\langle p^*\rangle. \tag{6.72}$$

For the purpose of the next theorem, we need to set $c_s = s$. Here, we apply results from Ref. [50, 51], while we omit mathematically rigorous assumptions and formulations:

Proposition 6.2 *The following statements can be found in Refs. [50, 51]:*

(1) *There exists an algorithm to reconstruct the function $\varphi(x)$ within the mean-square error δ^2 from the wavelet strengths $W_s[\varphi](x_0)$ estimated with accuracy $O(\delta)$, where x_0 is taken over all edges of scale s and the scale $s = 2^j s_0$ is taken for every $j \in \mathbb{Z}$.*

(2) *When the function φ is $(1+\sigma, M)$-regular, the wavelets of scale s for contributes to the mean-square error by at most $O(M^2 s^{2(1+\sigma)})$ in the function reconstruction.*

By using the statement above, the function $\varphi(x)$ can be reconstructed as follows: We perform the edge detection for scales $s = 2^j s_0$ for integer $j \geq 0$ and reconstruct the function $\varphi(x)$ according to the algorithm described in (1). Here, we ignore the edges with a scale less than s_0, which contributes little to the RMS error owing to (2).

Let us neglect the x-error, in which case the momentum p^* can be measured with statistical error

$$\delta^2 = \frac{\hbar^2}{4s^2 N^X}, \tag{6.73}$$

where $X = 1$ for the classical scheme and $X = 2$ for the quantum scheme. Since p^* is linearly related to the wavelet transform $W_s[\varphi](x_0)$ by (6.72) with $c_s = s$, the statistical error for the wavelet strength $W_s[\varphi](x_0)$ for a single edge x_0 is

$$\delta^2 = \left(\frac{s}{\hbar}\right)^2 \frac{\hbar^2}{4s^2 N^X} = \frac{4}{N^X}. \tag{6.74}$$

Therefore, the number $\mathcal{N}$ of photons per unit length is of the order $O(s^{-1}N) = O(s^{-1}\delta^{-2/X})$. Owing to the nature of the geometric series, the summation for $s = s_0, 2s_0, 2^2 s_0, \ldots$ becomes of the order

$$\mathcal{N} = O(s_0^{-1}\delta^{-2/X}). \tag{6.75}$$

On the other hand, the lack of edge detection with scale less than s_0 contributes to the RMS error at most $O(M^2 s_0^{2(1+\sigma)})$, which must be at most comparable to δ^2. Therefore, the lengthscale s_0 must be chosen such that

$$\delta = O(M s_0^{1+\sigma}). \tag{6.76}$$

Combining this equation with (6.75), we can establish the scaling relation between $\mathcal{N}$, δ, and M:

$$\delta = O\big((M\mathcal{N}^{1+\sigma})^{-\frac{1}{(X/2)(1+\sigma)+1}}\big) = \begin{cases} O\big((M\mathcal{N}^{1+\sigma})^{-\frac{1}{2(1+\sigma)+1}}\big) & \text{(classical);} \\ O\big((M\mathcal{N}^{1+\sigma})^{-\frac{1}{(1+\sigma)+1}}\big) & \text{(quantum),} \end{cases} \tag{6.77}$$

which coincides with the SQL and the Heisenberg limit of the function estimation in Eq. (5.110) with $q = 1 + \sigma$. This fact indicates that the wavelet measurement is an efficient way to extract the information of the entire function $\varphi(x)$, which can be performed separately on each lengthscale s.

In a practical situation, however, the wavelet measurement is hindered by the x-error as we have seen in the previous subsection, making the function restoration algorithm less efficient than the direct methods we have described in Chap 5.

6.4 Wavelet detection with more general states

6.4.1 Possibility of non-Gaussian probes

As we have seen in the previous sections, the x-error stands in the way of the efficient wavelet measurement, without which the Heisenberg limit $\delta = O(N^{-1})$ can be unconditionally attained. For the last part of this chapter, we discuss whether the statistical error in the wavelet measurement can be reduced by introducing a non-Gaussian probe.

We consider the single photon detection for brevity, although the analysis is parallel for the multiple photon detection. We return to Eq. (6.25):

$$\delta^2 = \iint \big(p + \hbar\varphi'(x) - \langle p^* \rangle\big)^2 dx dp W_q(x,p), \tag{6.78}$$

which is the variance of the operator $\hat{p} + \hbar\varphi(\hat{x})$. Since the variance of $\hat{x}$ is s^2, the Heisenberg uncertainty relation due to the commutation relation

$$[\hat{x}, \hat{p} + \hbar\varphi(\hat{x})] = i\hbar \tag{6.79}$$

bounds the statistical error from below:

$$\delta \geq \frac{\hbar}{2s}, \tag{6.80}$$

which coincides with the bound on the p-error $E_p 1$ in the Gaussian case. This suggests that, for a general probe state, the x-error can be eliminated and only the p-error remains.

This is confirmed by setting the probe wavefunction to

$$q(x) \propto e^{i[p_0 x - \varphi_0(x)]} e^{-\frac{(x_0 - x)^2}{4s^2}}, \tag{6.81}$$

where p_0 is chosen so that the initial expected momentum is zero. If the unknown function $\varphi(x)$ coincides with $\varphi_0(x)$, it can be straightforwardly checked that the momentum p^* is subject to the Gaussian distribution with variance $\frac{\hbar^2}{4s^2}$, attaining the equality in Eq. (6.80).

This strategy cannot be applied in the real situation, of course, as we do not know the function $\varphi(x)$ in question in the first place. In fact, the statistical error cannot be reduced from the Gaussian case if the function $\varphi(x)$ is unknown. To see this, let us consider an alternate case in which the phase function is $-\varphi(x)$. The statistical error for this case can be written as

$$\delta^2 = \iint \left(p - \hbar\varphi'(x) + \langle p^* \rangle\right)^2 dx dp W_q(x, p), \tag{6.82}$$

where $\langle p^* \rangle$ indicates the expected value when the phase function is $\varphi(x)$ rather than $-\varphi(x)$. By taking the mean between Eqs. (6.78) and (6.82), we obtain

$$\delta^2 = \iint p^2 dx dp W_q(x, p) + \iint \left(\hbar\varphi'(x) - \langle p^* \rangle\right)^2 dx dp W_q(x, p). \tag{6.83}$$

In the right-hand side of Eq. (6.83) the first term is the variance of p and the second term is the variance of $\varphi'(x)$. These terms can be regarded as the p-error and the x-error, whence δ^2 is bounded by the same tradeoff relation as the Gaussian case.

It is possible, though, to modify the probe state on the basis of some previous measurements on $\varphi(x)$. This approach is outside the scope of our study, however, since this requires more than the measurement on the wavelet transform $W_s[\varphi](x_0)$.

6.4.2 Preliminary results on indistinguishable photons

So far, the situation was limited to the case in which the probe beam consists of a definite number of distinguishable photon states. In order to apply our theory to

more general photon states such as coherent states and multimode squeezed state, we rewrite the system in the second-quantized formulation.

Let us denote by $\hat{a}_p$ the annihilation operator of a photon with momentum p. Then, the annihilation operator $\hat{a}_x$ of a photon localized at position x can formally be defined as

$$\hat{a}_x = \frac{1}{\sqrt{2\pi\hbar}} \int e^{ipx/\hbar} \hat{a}_p. \tag{6.84}$$

The unitary evolution U_φ of the system in Eq. (6.8) can be written as $U_\varphi = e^{-iH_\varphi}$, where H_φ is defined as

$$H_\varphi = -\int dy \varphi(y) \hat{a}_y^\dagger \hat{a}_y. \tag{6.85}$$

Owing to the commutation relation $[H_\varphi, \hat{a}_x] = \varphi(x)\hat{a}_x$, the unitary evolution $\hat{U}_\varphi$ can be expressed in the Heisenberg picture as

$$U_\varphi^\dagger \hat{a}_x U_\varphi = e^{i\varphi(x)} \hat{a}_x, \tag{6.86}$$

which corresponds to the Schrödinger picture of U_φ in Eq. (6.8).

In the second quantization, we can describe the wavelet measurement using a probe state with an indefinite photon number. For example, the probe state can be a coherent state $|\alpha\rangle$, where $\alpha(x)$ is a square-integrable function, which is defined as a simultaneous eigenstate of the annihilation operators: $\hat{a}_x|\alpha\rangle = \alpha(x)|\alpha\rangle$. A coherent state can be obtained as $|\alpha\rangle = D(\alpha)|0\rangle$ from the vacuum $|0\rangle$, where we have used the displacement operator

$$D(\alpha) = \exp\left[\int dx \alpha(x) \hat{a}_x^\dagger - \text{h.c.}\right]. \tag{6.87}$$

By Eq. (6.86), we see that the unitary operator U_φ lets a coherent state $|\alpha\rangle$ evolve into another coherent state $|\alpha_\varphi\rangle$, $\alpha_\varphi(x) = e^{i\varphi(x)}\alpha(x)$.

Now, we consider a general probe state $|q_\varphi\rangle = U_\varphi|q\rangle$ for the wavelet measurement. We first derive the precision bound on the basis of one-photon detection, which yields a similar result to Sec. 6.2. Instead of the probability distribution $P(p^*)$, we consider the photon flux $I(p^*)$ [90, 91] with respect to the measured momentum p^*:

$$I(p^*) = \langle q_\varphi | \hat{a}_{p^*}^\dagger \hat{a}_{p^*} | q_\varphi \rangle. \tag{6.88}$$

In other words, $I(p^*)dp$ corresponds to the number of photons whose momenta fall into an infinitesimal interval $[p^*, p^* + dp]$.

The number density $I(p^*)$ can be calculated in the same way as we did in Sec. 6.2.1. We may define the Wigner function $W_q(x,p)$ by

$$W_q(x,p) = \langle q|\hat{I}(x,p)|q\rangle, \tag{6.89}$$

$$\hat{I}(x,p) = \mathcal{F}_{y\to p}^{-1}[\hat{a}_{x+\frac{y}{2}}^\dagger \hat{a}_{x-\frac{y}{2}}] = \int \frac{dy}{2\pi\hbar} e^{ipy/\hbar} \hat{a}_{x+\frac{y}{2}}^\dagger \hat{a}_{x-\frac{y}{2}}. \tag{6.90}$$

In fact, integration of $W_q(x,p)$ by one variable yields

$$\int dp W_q(x,p) = \langle q|\hat{a}_x^\dagger \hat{a}_x|q\rangle =: I(x); \tag{6.91}$$

$$\int dx W_q(x,p) = \langle q|\hat{a}_p^\dagger \hat{a}_p|q\rangle =: I(p), \tag{6.92}$$

which correspond to the photon fluxes with respect to position and momentum, respectively. The Wigner functions $W_{q_\varphi}(x,p)$ and $W_q(x,p)$ are related in the same way as Eq. (6.17):

$$\begin{aligned} W_{q_\varphi}(x,p) &= \mathcal{F}^{-1}_{y\to p}\big[\langle q_\varphi|\hat{a}^\dagger_{x+\frac{y}{2}}\hat{a}_{x-\frac{y}{2}}|q_\varphi\rangle\big] \\ &= \mathcal{F}^{-1}_{y\to p}\big[\langle q|U_\varphi^\dagger \hat{a}^\dagger_{x+\frac{y}{2}}\hat{a}_{x-\frac{y}{2}}U_\varphi|q\rangle\big] \\ &= \mathcal{F}^{-1}_{y\to p}\big[e^{i[\varphi(x-\frac{y}{2})-\varphi(x+\frac{y}{2})]}\langle q|\hat{a}^\dagger_{x+\frac{y}{2}}\hat{a}_{x-\frac{y}{2}}|q\rangle\big] \\ &= W_{q_\varphi}(x,p) = W_q(x,p) * W_\varphi(p|x). \end{aligned} \tag{6.93}$$

Hence, the photon flux in Eq. (6.88) can be calculated as

$$I(p^*) = \int dx W_{q_\varphi}(x,p^*) = \iint dx dp W_q(x,p^*-p) W_\varphi(p|x), \tag{6.94}$$

which can be treated in the same way as the probability density in Eq. (6.19) if we normalize it by the total photon number $I = \int dp^* I(p^*)$. In particular, the statistical error derived in this section holds when the Winger function $W_q(x,p)$ of the probe state is Gaussian, e.g. the coherent state $|\alpha\rangle$ with $\alpha(x) \propto e^{-(x_0-x)^2/4s^2}$.

Next, we will investigate the effect of correlated photons on the basis of N-photon detection. The photon flux with respect to the simultaneously detected momenta $\boldsymbol{p}^* = (p_1^*, \dots, p_N^*)$ can be described as

$$I(p_1^*, \dots, p_N^*) = \langle q_\varphi|\hat{a}^\dagger_{p_1}\cdots\hat{a}^\dagger_{p_N}\hat{a}_{p_N}\cdots\hat{a}_{p_1}|q_\varphi\rangle. \tag{6.95}$$

When the photons are indistinguishable (i.e. commutation relations $[a^\dagger_{p_i}, a_{p_j}] = \delta(p_i - p_j)$ holds even for different i and j), the event of detection is invariant under the permutation of the momenta $(p_1^*, \dots, p_N^*)$. For example, the detection of two photons of momenta (p_1^*, p_2^*) contributes not only to the flux $I(p_1^*, p_2^*)$ but also to the flux $I(p_2^*, p_1^*)$.[3] The photon flux can also be rewritten as

$$I(p_1^*, \dots, p_N^*) = I(p_1^*)\cdots I(p_N^*) g^{(N)}(p_1^*, \dots, p_N^*), \tag{6.96}$$

where $g^{(N)}$ is known as the correlation function of order N [91]:

$$g^{(N)}(p_1, \dots, p_N) = \frac{\langle q_\varphi|\hat{a}^\dagger_{p_1}\cdots\hat{a}^\dagger_{p_N}\hat{a}_{p_N}\cdots\hat{a}_{p_1}|q_\varphi\rangle}{\langle q_\varphi|\hat{a}^\dagger_{p_1}\hat{a}_{p_1}|q_\varphi\rangle\cdots\langle q_\varphi|\hat{a}^\dagger_{p_N}\hat{a}_{p_N}|q_\varphi\rangle}. \tag{6.97}$$

[3] This means that every N-photon detection contributes to the total flux over $N!$ times, since the momenta $p_1^*, \dots, p_N^*$ can be regarded as distinct. It is meaningless to ask for individual values with non-unique momenta, such as $I(p, p, \dots)$, because the photon flux yields the photon number only when it is integrated over $\boldsymbol{p}^*$.

The intercorrelation of photons is indicated by $g^{(N)} \neq 1$, which may alter the statistical error for simultaneous N-photon detection.

Similarly to the analysis in Sec. 6.3.1, we can extend Eq. (6.94) into the multi-dimensional version:

$$I(\boldsymbol{p}^*) = \iint dxdp W_q^{(N)}(\boldsymbol{x}, \boldsymbol{p}^* - \boldsymbol{p}) W_\varphi^{(N)}(\boldsymbol{p}|\boldsymbol{x}). \tag{6.98}$$

Here, the extended Wigner function $W_q^{(N)}(\boldsymbol{x}, \boldsymbol{p})$ is defined as

$$W_q^{(N)}(\boldsymbol{x}, \boldsymbol{p}) = \langle q| {:}\hat{I}(x_1, p_1) \cdots \hat{I}(x_N, p_N){:} |q\rangle, \tag{6.99}$$

where the products between colons are so-called *normally ordered*, i.e. rearranged such that the creation operators always comes before the annihilation operators. Hence, the statistical error with simultaneous detection of N photons can be calculated in the same way as in Sec. 6.2.3.

Since $g^{(N)} = 1$ holds for a coherent state $|\alpha\rangle$, the measured momenta of N photons are mutually independent and we obtain the same SQL as Eq. (6.49). To obtain a different result, we consider the multimode squeezing of the coherent state [91–93]:

$$|\alpha, \zeta\rangle = S(\zeta) D(\alpha) |0\rangle, \tag{6.100}$$

where $\zeta(y_1, y_2)$ is a symmetric square-integrable function and the squeezing operator $S(\zeta)$ is defined as

$$S(\zeta) = \exp \frac{1}{2} \left[\iint dy_1 dy_2 \zeta(y_1, y_2) \hat{a}_{y_1}^\dagger \hat{a}_{y_2}^\dagger - \text{h.c.} \right]. \tag{6.101}$$

Such squeezing can be introduced by parametric down-conversion [94–96]. When the squeezing is weak, the correlation function of order 2 can be approximated as [97, 98]:

$$g^{(2)}(p_1, p_2) \approx 1 + 2 \operatorname{Re} \frac{\tilde{\zeta}(p_1, p_2)}{\tilde{\alpha}(p_1) \tilde{\alpha}(p_2)}, \tag{6.102}$$

where $\tilde{\alpha}$ and $\tilde{\zeta}$ are the Fourier transforms of α and ζ. Therefore, the covariance matrix C_p of the momenta can be changed by the squeezing function ζ, which may be used to the break the SQL. It is left as an open problem how far the error can be reduced by using squeezed beams, which requires calculation beyond weak squeezing.

Chapter 7

Summary and Outlook

7.1 Summary

In this thesis, we have developed the theory of quantum estimation on continuous data.

In Chapter 2, we have reviewed quantum information theory and quantum metrology, where we have explained two different types of quantum limits, the *standard quantum limit* (SQL) and the Heisenberg limit. In quantum estimation theory, the estimation error δ has a lower bound called the quantum Cramér–Rao (CR) bound, which is obtained by geometrically evaluating the local structure of the quantum state space in terms of the Fisher information metric. In the classical scheme, the probe states are assumed to have no correlation, and the quantum CR bound yields a lower bound called the SQL: $\delta \geq O(N^{-1/2})$, where N indicates the amount of resource. This can be regarded as a classical error bound in the sense that it is consistent with the central limit theorem. The SQL can be surpassed, however, when nonclassical correlation in the measurement is introduced. The quantum CR bound for the general case is the Heisenberg limit: $\delta \geq O(N^{-1})$, and the equality $\delta = O(N^{-1})$ can be achieved by, e.g., a generalized *Greenberger–Horne–Zeilinger state* (GHZ state).

In Chapter 3, we have reviewed the existing studies of function estimation and signal estimation, where we make use of functional analysis. We have introduced the notion of the Hölder condition, which quantifies the smoothness of a given function by some parameter q. The Hölder condition can be used to estimate a function supported on a continuous spacetime, which is called the *local linear smoothing* (LLS). This estimation method uses a point-spread function called the "kernel" to reduce discrete observations to a continuous function, which best approximates the function when the width of the function is set to be optimal. As a result, the approximation error is bounded from above by a quantity depending on q, which indicates that smoother function can be approximated more efficiently.

For the sake of completeness, we have also reviewed the related work on signal estimation. There, stochastic signals was shown to be estimated by use of filtering, which is based on the power-spectrum analysis of a stochastic process. This technique roughly corresponds to the LLS in that we need to set the optimal timescale of smoothing according to the shape of the power spectrum, which represents the typical smoothness of the signal.

In Chapter 4, we have presented a comprehensive study on a multiparameter quantum metrology, which extends the previous studies in the author's master thesis. On the basis of the quantum limits upon linear Hamiltonian models derived in the master thesis, the sequential and parallel schemes are compared with each other. After a series of operator-algebraic arithmetics, we have shown that the asymptotic costs of estimation for both schemes are equivalent. This implies that quantum metrology can be parallelized by a constant factor of overhead regardless of the size of the Hamiltonian model.

In Chapter 5, we have derived the theoretical limits on function estimation incorporated with quantum metrology. We consider an unknown phase function $\varphi(x)$ in a quantum system, whose dynamics is described by the phase factor $e^{-i\varphi(x)}$ depending on the position x. In the estimation of the phase function $\varphi(x)$, the accuracy is measured by the root-mean-square error. The theoretical limits are derived by reducing the problem to that of the multiparameter estimation. Under the assumption of a bounded q-Hölder condition, a lower bound on the error is found to be $O(N^{-q/(2q+1)})$ with classical scheme, while quantum scheme allows the error of the order of $O(N^{-q/(q+1)})$. These error scalings can be stated as the SQL and the Heisenberg limit on function estimation. While their dependences on N are weaker than the counterparts on parameter estimation, the latter can be reproduced from the former by formally taking the smooth limit $q \to \infty$.

We have also investigated the methodology to attain the quantum limits. Notably, one may choose either position- or momentum-localized states for the probes: The local linear smoothing can be used for the localized states; the postselected quantum tomography can be used for the momentum-localized states. Both strategies lead to the SQL or the Heisenberg limit, except that the estimation method by momentum-localized states is ensured only for $0 < q \leq 1$.

In Chapter 6, we have applied the theory of function estimation to the edge detection. Continuing with the quantum system with a phase function $\varphi(x)$, we consider a problem of estimating the wavelet transform of the function φ, which can be used to locate the edges of a given length scale. We have proposed a modified version of ghost imaging as a physical implementation of the wavelet measurement. In this strategy, one uses a wave packet as a probe state, and the wavelet transform can be measured by detecting the momentum of the output state. We have shown that the probability density function of the momentum is described by a chain of quasi-Markovian processes, no matter whether the probe state consists of a single photon or multiple photons. Furthermore, the estimation error can be decomposed into the momentum variance ("p-error") and the transformed position variance ("x-error"), whence we could analytically evaluate the error bounds from the position-momentum uncertainty relation. The Heisenberg-limited edge detection is found to be performed by a Gaussian beam with negative quantum correlations in the momentum space.

We have also demonstrated the consistency with the function estimation. It is known that the entire function $\varphi(x)$ can be restored by measuring the location and strength of each edge with various length scales s. As long as we

ignore the x-errors, the estimation error in restoring the function is subject to the SQL $O(N^{-q/(2q+1)})$ and the Heisenberg limit $O(N^{-q/(2q+1)})$, depending on whether or not the quantum correlation is permitted. This suggests that the wavelet measurement is an efficient way to extract the information from a given function.

7.2 Outlook

This thesis rigorously shows a potential impact of quantum mechanics on the estimation of continuous data by analytically calculating the estimation error. In particular, we have derived the optimal error in the estimation of an unknown function encoded in a quantum system, which depends on the smoothness of the function to be estimated.

Some problems are open for discussion, however, when we consider practical applications of our studies. First, there exists many techniques for continuous data analysis, such as signal separation [44] or model prediction [99]. How should we apply these techniques to continuous data encoded in quantum systems? One naïve way is to estimate the entire function as described in Chapter 5, and then conduct classical algorithms for further analysis. On the other hand, it would be more efficient if we can perform the continuous data analysis via some special quantum measurement, circumventing the function estimation itself. The edge detection disucussed in Chapter 6 is one such example, since it detects the wavelet coefficient rather than the entire function. However, it still involves a brute-force search for finding the extrema, which is essentially as costly as the function estimation.

This leads us to the second perspective beyond this thesis: the possibility to combine rigorous quantum metrology and heuristic algorithms in edge detection. For example, kernel regression [69, 89, 100] will be useful for quantum metrology because it is a common optimization algorithm to reduce the number of costly measurements. For two-dimensional images, quantum metrology may be combined with the Canny algorithm [46], which is based on the qualitative assumption that edge extends in some one-dimensional region.

The third perspective is for the theoretical interest. We have seen that the quantum limits on function estimation can be reached by various probes: In Chapter 5, we have seen the optimality of position- and momentum-localized probes. Wave packets can be regarded as lying between these extremes, but they are still quantum-limited as seen in Chapter 6. These facts suggest that almost any type of probe states can be used for the optimal estimation, no matter how they are localized in the position or momentum space. However, the general theory about this methodology is yet to be known.

Acknowledgement

First of all, let me show my greatest gratitude to my supervisor, Prof. Masahito Ueda. He is among the most ardent researchers that I have met ever. However he is occupied with many positions, he has always been eager to advise and have conversation with me. He has also provided me with an opportunity to communicate with physicists across fields who concerns the theory of deep learning. Thanks to his pertinent advice and suggestions, I could grow my rough ideas into a research topic of much interest and write up this thesis.

I would also like to thank Prof. Masato Koashi. In addition to the helpful discussions at quarterly assessment as my ALPS sub-supervisor, I have a few times consulted him as an expert in quantum information theory. From the deep and broad knowledge in both theory and experiment, he has suggested how to extend my research in both the fundamental and practical directions.

I appreciate Dr. Dominic Berry very much for letting me stay in his group in Macquarie University for three month in 2018. We have had discussion about the cutting-edge techniques in quantum metrology and signal detection, from which I have learned how to push my work forward and how to stress its importance. I have also had a good time hearing lectures and seminars from the students and other visitors in his group.

Numerous thanks go to all members and former members in Ueda group, with whom I have been very happy to work together. I am also proud to support Yuto Murashita, Gong Zongping, Ryusuke Hamazaki, and Kohei Kawabata as a co-author in their respective work. In particular, I would like to show my gratitude to Ryusuke, who have encouraged me through my doctoral course when I am in and out of condition.

I profoundly appreciate all the researchers who have kindly had discussion with me, including Dr. Tomohiro Shitara, Dr. Yuto Ashida, Prof. Fuyuhiko Tanaka, Dr. Ikko Hamamura, Prof. Keisuke Fujii, and Dr. Yuichiro Matsuzaki.

Finally, I acknowledge financial supports from Grant-in-Aid for JSPS Fellows (KAKENHI Grant No. JP19J11303) and the Advanced Leading graduate course for Photon Science “ALPS” by JSPS.

Appendix A

Trace in the Symmetric Tensor-product Space

Here, we show Proposition 4.5 in Chapter 4 by proposing an algorithm to calculate

$$\mathrm{tr}[P_{\mathrm{sym}}\{X\}_r^m], \qquad \{X\}_r = \sum_{i=1}^{r} I \otimes \cdots \otimes \underbrace{X}_{i\mathrm{th}} \otimes \cdots \otimes I. \tag{A.1}$$

in terms of $\mathrm{tr}\, X, \ldots, \mathrm{tr}\, X^m$ for general m. This requires a few preparatory steps concerning combinatorial mathematics.

A.1 Preparation: Bell Polynomials

Definition A.1 A sequence of positive numbers $\lambda = (\lambda_1, \ldots, \lambda_h)$ is said to be a *partition* of n if $\lambda_1 \geq \cdots \geq \lambda_h \geq 1$ and $\lambda_1 + \cdots + \lambda_h = n$. The number h is said to be the *height* of λ. Given a partition λ of n, the corresponding monomial $p^\lambda(x_1, x_2, \ldots)$ can be defined by

$$p^\lambda(x_1, \ldots, x_{\lambda_1}) = x_{\lambda_1} \cdots x_{\lambda_h}. \tag{A.2}$$

We may equivalently write this as

$$p^\lambda(x_1, \ldots, x_{\lambda_1}) = x_1^{j_1} \cdots x_{\lambda_1}^{j_{\lambda_1}}, \tag{A.3}$$

where j_k is the number of k contained in the sequence λ.

Example A.2 $\lambda = (4, 2, 2, 2, 1, 1)$ is a partition of 12 with height 6. The corresponding monomial is $p^\lambda(x_1, x_2, x_3, x_4) = x_1^2 x_2^3 x_4$.

We consider partitioning a set $S = \{1, \ldots, n\}$ into h subsets so that the sizes of the subsets are $\lambda_1, \ldots, \lambda_h$. We note that subsets of the same size is mutually indistinguishable. If we denote by b_λ the number of the ways of such partitioning, it can be calculated as

$$b_\lambda = \frac{n!}{\lambda_1! \cdots \lambda_h!} \cdot \frac{1}{j_1! \cdots j_{\lambda_1}!}. \tag{A.4}$$

Here, the integers $j_1, \ldots, j_{\lambda_1}$ are the same as the ones defined in Def. A.1.

Example A.3 We consider the partition $\lambda = (2, 2, 1)$ of 5. The number of ways to partition the set $\{1, 2, 3, 4, 5\}$ into subsets of size $2, 2, 1$ is

$$\frac{5!}{2!2!1!} \cdots \frac{1}{1!2!} = 15, \tag{A.5}$$

which can be directly enumerated as

$$\begin{array}{lll}
\{\{1,2\},\{3,4\},\{5\}\}, & \{\{1,3\},\{2,4\},\{5\}\}, & \{\{1,4\},\{2,3\},\{5\}\}, \\
\{\{1,2\},\{3,5\},\{4\}\}, & \{\{1,3\},\{2,5\},\{4\}\}, & \{\{1,5\},\{2,3\},\{4\}\}, \\
\{\{1,2\},\{4,5\},\{3\}\}, & \{\{1,4\},\{2,5\},\{3\}\}, & \{\{1,5\},\{2,4\},\{3\}\}, \\
\{\{1,3\},\{4,5\},\{2\}\}, & \{\{1,4\},\{3,5\},\{2\}\}, & \{\{1,5\},\{3,4\},\{2\}\}, \\
\{\{2,3\},\{4,5\},\{1\}\}, & \{\{2,4\},\{3,5\},\{1\}\}, & \{\{2,5\},\{3,4\},\{1\}\}.
\end{array}$$

Definition A.4 The *partial Bell polynomials* $B_{n,h}(x_1, \ldots, x_{n-k+1})$ are defined as

$$B_{n,h}(x_1, \ldots, x_{n-k+1}) = \sum_{\lambda \vdash (n,h)} b_\lambda p^\lambda(x_1, \ldots, x_{\lambda_1}), \tag{A.6}$$

where we denote by $\lambda \vdash (n, h)$ that λ is the partition of n with height h. The *total Bell polynomials* $B_n(x_1, \ldots, x_n)$ are defined as

$$B_n(x_1, \ldots, x_n) = \sum_{h=1}^{n} B_{n,h}(x_1, \ldots, x_{n-k+1}) = \sum_{\lambda \vdash n} b_\lambda p^\lambda(x_1, \ldots, x_{\lambda_1}), \tag{A.7}$$

where $\lambda \vdash n$ means that λ is the partition of n with any height.

Example A.5 Let us consider $B_{5,3}$. There are two partitions of 5 with height 3: one is $\lambda = (2, 2, 1)$, which have appeared in Example A.3, and the other is $\mu = (3, 1, 1)$. Hence, the partial Bell polynomial is

$$B_{5,3}(x_1, x_2, x_3) = b_\lambda p^\lambda(x_1, x_2) + b_\mu p^\mu(x_1, x_2, x_3) = 15x_1x_2^2 + 10x_1^2x_3. \tag{A.8}$$

The total Bell polynomial B_5 can be obtained by considering all 7 partitions:

$$B_5(x_1, x_2, x_3, x_4, x_5) = x_1^5 + 10x_1^3x_2 + 15x_1x_2^2 + 10x_1^2x_3 + 10x_2x_3 + 5x_1x_4 + x_5. \tag{A.9}$$

The Bell polynomials are useful in several situations, including one in which we calculate higher derivative of a composite function. This is known as Faà di Bruno's formula:

Lemma A.6 (Faà di Bruno's formula [101, 102]) *If $f(x)$ and $g(x)$ are both n-time differentiable functions of x, the n-th derivative of the composite function $f(g(x))$ is given by*

$$\frac{d^n}{dx^n} f(g(x)) = \sum_{h=1}^{n} B_{n,h}\Big(g'(x), \ldots, g^{(n-h+1)}(x)\Big) f^{(h)}(g(x)). \tag{A.10}$$

In particular, for $f(x) = e^x$,

$$\frac{d^n}{dx^n} e^{g(x)} = B_n\Big(g'(x), \ldots, g^{(n)}(x)\Big) e^{g(x)}. \tag{A.11}$$

For later convenience, let us define the homogeneity and the weighted homogeneity:

Definition A.7 A polynomial $P(x_1, x_2, \ldots, x_m)$ is *homogeneous polynomial* of degree h if and only if

$$P(tx_1, tx_2, \ldots, tx_m) = t^h P(x_1, x_2, \ldots, x_m). \tag{A.12}$$

Similarly, a polynomial $P(x_1, x_2, \ldots, x_m)$ is *weighted-homogeneous polynomial* of degree n if and only if

$$P(tx_1, t^2x_2, \ldots, t^m x_m) = t^n P(x_1, x_2, \ldots, x_m). \tag{A.13}$$

It can straightforwardly be seen that a partial Bell polynomial $B_{n,h}$ is homogeneous of degree h as well as weighted-homogeneous of degree n. More generally, the following proposition holds:

Proposition A.8 *A polynomial $P(x_1, \ldots, x_m)$ is homogeneous of degree h and weighted-homogeneous of degree n and if and only if it can be written in the form*

$$P(x_1, \ldots, x_m) = \sum_{\lambda \vdash (n,h)} a_\lambda p^\lambda(x_1, \ldots, x_{\lambda_1}) \tag{A.14}$$

for some constants a_λ.

A.2 Derivation of the trace

We consider a system of free particles subject to the Bose–Einstein statistics. We regard the operator X on the d-dimensional Hilbert space $\mathcal{H}_S = \mathbb{C}^d$ as the Hamiltonian for a system one particle. If there are N particles in the system, the corresponding Hamiltonian is represented by

$$P_{\mathrm{sym}}\{X\}_N, \tag{A.15}$$

where P_{sym} denotes the projection onto the symmetric Fock space. The partition function of this system is

$$Z(N, \beta) = \mathrm{tr}\big[e^{-\beta P_{\mathrm{sym}}\{X\}_N}\big] = \sum_{m=0}^{\infty} \frac{(-\beta)^m}{m!} \mathrm{tr}\big[P_{\mathrm{sym}}\{X\}_N^m\big]. \tag{A.16}$$

On the other hand, if the system is in equilibrium with a particle bath with chemical potential μ, the system is described by the grand partition function

$$\Xi(\mu, \beta) = \sum_{N=0}^{\infty} e^{\mu\beta N} Z(N, \beta) = \sum_{m=0}^{\infty} \sum_{N=0}^{\infty} e^{\mu\beta N} \frac{(-\beta)^m}{m!} \mathrm{tr}\big[P_{\mathrm{sym}}\{X\}_N^m\big]. \tag{A.17}$$

By taking $t = -\beta$ and $s = e^{\beta\mu}$, we may write

$$\Xi(s,t) = \sum_{m=0}^{\infty}\sum_{N=0}^{\infty} s^N \frac{t^m}{m!}\operatorname{tr}\bigl[P_{\mathrm{sym}}\{X\}_N^m\bigr]. \tag{A.18}$$

Hence, the desired trace (A.1) can be obtained by expanding the grand partition function $\Xi(s,t)$ in the powers of s and t. In particular, the terms in proportion to $\frac{t^m}{m!}$ can be extracted by the differentiation

$$\frac{\partial^n}{\partial t^n}\Xi(s,t=0) = \sum_{N=0}^{\infty} s^N \operatorname{tr}\bigl[P_{\mathrm{sym}}\{X\}_N^m\bigr]. \tag{A.19}$$

If we denote the eigenvalues of X by $\epsilon_1,\dots,\epsilon_d$, the grand partition function can also be written as

$$\Xi(s,t) = \prod_{j=1}^{d}\frac{1}{1-e^{\beta(\mu-\epsilon_j)}} = \prod_{j=1}^{d}\frac{1}{1-se^{t\epsilon_j}}. \tag{A.20}$$

Hence, the logarithm $\xi(s,t) = \log\Xi(s,t)$ can be expanded as

$$\begin{aligned}\xi(s,t) &= \sum_{j=1}^{d} -\log(1-se^{t\epsilon_j}) = \operatorname{tr}\bigl[-\log(1-se^{tX})\bigr]\\ &= \operatorname{tr}\Bigl[\sum_{j=1}^{\infty}\frac{(se^{tX})^j}{j}\Bigr] = \sum_{j=1}^{\infty}\frac{s^j}{j}\operatorname{tr}[e^{jtX}].\end{aligned} \tag{A.21}$$

Extracting the terms in proportion to $\frac{t^k}{k!}$ yields

$$\begin{aligned}\frac{\partial^k}{\partial t^k}\xi(s,t=0) &= \sum_{j=1}^{\infty}\frac{s^j}{j}\operatorname{tr}[(jX)^k]\\ &= \sum_{j=1}^{\infty} j^{k-1}s^j\operatorname{tr}[X^k] = \frac{sA_{k-1}(s)}{(1-s)^k}\operatorname{tr}[X^k],\end{aligned} \tag{A.22}$$

where we have introduced the following polynomials.

Definition A.9 (Eulerian polynomials [103, 104]) The *Eulerian polynomials* $A_k(x)$ $(k = 0, 1, \dots)$ is defined by the formal power series

$$A_k(x) = (1-x)^{k+1}\sum_{j=1}^{\infty} j^k x^{j-1}. \tag{A.23}$$

In fact, $A_k(x)$ becomes a polynomial with finite degree; the degree of $A_k(x)$ is $k-1$ for $k \geq 1$.

The fact that $A_k(x)$ are finite-degree polynomials can be confirmed as follows. For $k \geq 1$, we have

$$\begin{aligned}\frac{d}{dx}[xA_{k-1}(x)] &= \frac{d}{dx}\left[(1-x)^k \sum_{j=1}^{\infty} j^{k-1}x^j\right] \\ &= -k(1-x)^{k-1}\sum_{j=1}^{\infty} j^{k-1}x^j + (1-x)^k \sum_{j=1}^{\infty} j^k x^{j-1} \\ &= \frac{A_k(x) - kxA_{k-1}(x)}{1-x},\end{aligned} \tag{A.24}$$

from which we obtain the recurrence relation

$$A_k(x) = (1-x)\frac{d}{dx}[xA_{k-1}(x)] + kxA_{k-1}(x). \tag{A.25}$$

Therefore, starting from

$$A_0(x) = (1-x)\sum_{j=1}^{\infty} x^{j-1} = 1, \tag{A.26}$$

we can recursively compute the Eulerian polynomials:

$$\begin{gathered}A_0(x) = A_1(x) = 1, \quad A_2(x) = x+1, \quad A_3(x) = x^2 + 4x + 1, \\ A_4(x) = x^3 + 11x^2 + 11x + 1, \quad A_5(x) = x^4 + 26x^3 + 66x^2 + 26x + 1,\end{gathered}$$

and so on. In particular, it can be shown by induction that the highest-degree term of $A_k(x)$ is x^{k-1} for $k \geq 1$.

Let us return to Eq. (A.22). If we apply Faà di Bruno's formula (Lem. A.6) to the grand partition function $\Xi(s,t) = e^{\xi(s,t)}$, we obtain

$$\begin{aligned}\frac{\partial^m}{\partial t^m}\Xi(s,t=0) &= \frac{\partial^m}{\partial t^m} e^{\xi(s,t=0)} \\ &= e^{\xi(s,t=0)} B_m\left(\frac{\partial}{\partial t}e^{\xi(s,t=0)}, \dots, \frac{\partial^m}{\partial t^m}e^{\xi(s,t=0)}\right) \\ &= e^{-d\log(1-s)} B_m\left(\frac{sA_0(s)}{(1-s)}\operatorname{tr} X, \dots, \frac{sA_{m-1}(s)}{(1-s)^m}\operatorname{tr} X^m\right) \\ &= \frac{1}{(1-s)^d} B_m\left(\frac{sA_0(s)}{(1-s)}\operatorname{tr} X, \dots, \frac{sA_{m-1}(s)}{(1-s)^m}\operatorname{tr} X^m\right).\end{aligned} \tag{A.27}$$

By using the fact that the Bell polynomial B_m is weighted-homogeneous of degree m, we obtain

$$\begin{aligned}\frac{\partial^m}{\partial t^m}\Xi(s,t=0) &= \frac{1}{(1-s)^{m+d}} B_m\big(sA_0(s)\operatorname{tr} X, \dots, sA_{m-1}(s)\operatorname{tr} X^m\big) \\ &= \frac{1}{(1-s)^{m+d}} \sum_{l=1}^{m} s^l S_{m,l}\big(\operatorname{tr} X, \dots, \operatorname{tr} X^m\big),\end{aligned} \tag{A.28}$$

where we define the polynomials $S_{m,l}(x_1, \ldots, x_m)$ by

$$B_m(sA_0(s)x_1, \ldots, sA_{m-1}(s)x_m) = \sum_{l=1}^{m} s^l S_{m,l}(x_1, \ldots, x_m). \tag{A.29}$$

The desired trace (A.1) is the coefficient of s^N in the last line of Eq. (A.28). We recall that the fraction $\frac{1}{(1-s)^{m+d}}$ can be represented by a power series of s as

$$\frac{1}{(1-s)^{m+d}} = \sum_{l=0}^{\infty} \binom{m+d+l-1}{m+d-1} s^l, \tag{A.30}$$

where $\binom{n}{m}$ is the binomial coefficient. Therefore, the desired trace can be identified as

$$\operatorname{tr}\bigl[P_{\mathrm{sym}}\{X\}_N^m\bigr] = \sum_{l=1}^{m} \frac{(m+d+N-l-1)!}{(m+d-1)!(N-l)!} S_{m,l}(\operatorname{tr} X, \ldots, \operatorname{tr} X^m), \tag{A.31}$$

which can be systematically calculated from Eq. (A.29) by using the Bell and the Eulerian polynomials.

Now, let us substitute $m = 4$ to prove Prop. 4.5. The corresponding Bell polynomial is

$$B_4(x_1, x_2, x_3, x_4) = x_1^4 + 4x_3x_1 + 3x_2^2 + 6x_2x_1^2 + x_1^4 \tag{A.32}$$

and Eq. (A.29) becomes

$$\begin{aligned} &B_4(sx_1, sx_2, s(s+1)x_3, s(s^2+4s+1)x_4) \\ &= sx_4 + s^2(3x_2 + 4x_1x_3 + 4x_4) + s^3(6x_1^2x_2 + 4x_1x_3 + x_4) + s^4x_1^4, \end{aligned} \tag{A.33}$$

from which the polynomials $S_{4,l}(x_1, x_2, x_3, x_4)$ $(l = 1, 2, 3, 4)$ can be identified.

Now, the trace in question can be calculated as

$$\operatorname{tr}\bigl[P_{\mathrm{sym}}\{X\}_r^4\bigr] = \sum_{l=1}^{4} \binom{r+d+3-l}{d+3} S_{4,l}(\operatorname{tr} X, \operatorname{tr} X^2, \operatorname{tr} X^3, \operatorname{tr} X^4). \tag{A.34}$$

Noting that $\operatorname{tr} X = 0$ is assumed in Prop. 4.5, this equation can be expanded as

$$\begin{aligned} \operatorname{tr}\bigl[P_{\mathrm{sym}}\{X\}_r^4\bigr] = &\binom{r+d+2}{d+3} \operatorname{tr} X^4 \\ &+ \binom{r+d+1}{d+3} [3(\operatorname{tr} X^2)^2 + 4 \operatorname{tr} X^4] \\ &+ \binom{r+d}{d+3} \operatorname{tr} X^4. \end{aligned} \tag{A.35}$$

If we divide both sides by $D = \binom{r+d-1}{d-1}$, this equation becomes

$$\begin{aligned}\frac{1}{D}\operatorname{tr}\bigl[P_{\mathrm{sym}}\{X\}_r^4\bigr] &= \frac{(r+d+2)(r+d+1)(r+d)r}{d(d+1)(d+2)(d+3)}\operatorname{tr} X^4 \\ &\quad + \frac{(r+d+1)(r+d)r(r-1)}{d(d+1)(d+2)(d+3)}[3(\operatorname{tr} X^2)^2 + 4\operatorname{tr} X^4] \\ &\quad + \frac{(r+d)r(r-1)(r-2)}{d(d+1)(d+2)(d+3)}\operatorname{tr} X^4, \end{aligned} \tag{A.36}$$

which yields Prop. 4.5.

Appendix B

Fourier Transform of Regular Functions

Here, we prove two theorems 5.4 and 5.5 that have appeared in Chapter 5. Both theorems provide conditions for a function $u(x)$ to be (q, M)-regular on the basis of its Fourier coefficients, where Th. 5.4 yields a necessary condition and Th. 5.5 yields a sufficient condition. Before proving these theorems, we present a lemma that characterizes the behavior of the Fourier coefficients of a (q, M)-regular function:

Lemma B.1 *Let us set $q = m + \sigma$ with an integer $m \geq 0$ and a real number $0 < \sigma \leq 1$. We consider an m-time differentiable function $u : [0, L] \to \mathbb{C}$ with the periodic boundary condition, and define its Fourier transform $\{u_k\}_{k\in\mathbb{Z}}$ by*

$$u(x) := \sum_{k\in\mathbb{Z}} u_k e^{2\pi ikx/L}. \tag{B.1}$$

Then, the function u is (q, M)-regular if and only if the inequality

$$\sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2)\left(\frac{k}{L}\right)^{2m} \sin^2 \frac{\pi k\varepsilon}{L} \leq \frac{M^2|\varepsilon|^{2\sigma}}{(2\pi)^{2m}} \tag{B.2}$$

holds for an arbitrary ε.

Proof. The Fourier transform of the mth derivative $u^{(m)}$ is calculated as

$$u^{(m)}(x) = \sum_{k\in\mathbb{Z}} \left(\frac{2\pi ik}{L}\right)^m u_k e^{2\pi ikx/L} \tag{B.3}$$

and therefore

$$\begin{aligned}\Delta_\varepsilon u(x) = u(x+\varepsilon) - u(x) &= \sum_{k\in\mathbb{Z}} \left(\frac{2\pi ik}{L}\right)^m u_k [e^{2\pi ik(x+\varepsilon)/L} - e^{2\pi ikx/L}] \\ &= \sum_{k\in\mathbb{Z}} \left(\frac{2\pi ik}{L}\right)^m u_k (e^{2\pi ik\varepsilon/L} - 1) e^{2\pi ikx/L}. \end{aligned} \tag{B.4}$$

Hence, by the Perceval inequality, we have

$$\begin{aligned}\|\Delta_\varepsilon u\|_2^2 &= \int_0^L \frac{dx}{L}|u(x+\varepsilon)-u(x)|^2 \\ &= \sum_{k\in\mathbb{Z}}\left|\left(\frac{2\pi ik}{L}\right)^m u_k(e^{2\pi ik\varepsilon/L}-1)\right|^2 \\ &= \sum_{k\in\mathbb{Z}}|u_k|^2\left(\frac{2\pi k}{L}\right)^{2m}\sin^2\frac{\pi k\varepsilon}{L} \\ &= \sum_{k=1}^{\infty}(|u_k|^2+|u_{-k}|^2)\left(\frac{2\pi k}{L}\right)^{2m}\sin^2\frac{\pi k\varepsilon}{L}. \end{aligned} \tag{B.5}$$

We recall that the (q, M)-regularity is defined by the inequality

$$\sup_{\varepsilon\neq 0}\frac{\|\Delta_\varepsilon u^{(m)}\|_2}{|\varepsilon|^\sigma} \le M, \tag{B.6}$$

which is equivalent to

$$\forall\varepsilon \quad \|\Delta_\varepsilon u^{(m)}\|_2^2 \le M^2|\varepsilon|^{2\sigma}. \tag{B.7}$$

The desired inequality Eq. (B.2) is obtained by substituting Eq. (B.5) into the left-hand side of Eq. (B.7). □

Now, we prove the two theorems in Chapter 5.

Theorem B.2 (Theorem 5.5 in the main text) *Let us set $q = m + \sigma$ with an integer $m \ge 0$ and a real number $0 < \sigma \le 1$. We consider an arbitrary m-time differentiable function $u : [0, L] \to \mathbb{C}$ with the periodic boundary condition, and define its Fourier transform $\{u_k\}_{k\in\mathbb{Z}}$ by*

$$u(x) := \sum_{k\in\mathbb{Z}} u_k e^{2\pi ikx/L}. \tag{B.8}$$

Then, there exists a real number $C'_q > 0$ such that if

$$\sum_{k=1}^{\infty}\left(\frac{k}{L}\right)^{2q}(|u_k|^2+|u_{-k}|^2) \le C'_q M^2 \tag{B.9}$$

holds for $M > 0$, the function u is (q, M)-regular.

Proof. Let us define a continuous function $f(t)$ over $t \ge 0$ by

$$f(t) = \frac{|\sin\pi t|}{t^\sigma}. \tag{B.10}$$

We note that the right-hand side converges in the limit $t \to +0$ since $0 < \sigma \le 1$. Moreover, for $t \ge \frac{1}{2}$ we have $f(t) \le t^{-\sigma} \le 2^\sigma = f(\frac{1}{2})$. Therefore, the function

$f(t)$ has a maximum $a = \max_{t\geq 0} f(t)$ at somewhere in $0 \leq t \leq \frac{1}{2}$. Using this maximum, we may write

$$\sin^2 \frac{\pi k\varepsilon}{L} = \left(\frac{\pi k\varepsilon}{L}\right)^{2\sigma} \left[f\left(\frac{\pi k|\varepsilon|}{L}\right)\right]^2 \leq a^2 \left(\frac{\pi k|\varepsilon|}{L}\right)^{2\sigma}. \tag{B.11}$$

Therefore, if u satisfies the assumption (B.9), we have

$$\begin{aligned}\sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2)\left(\frac{k}{L}\right)^{2m} \sin^2 \frac{\pi k\varepsilon}{L} &\leq \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2)\left(\frac{k}{L}\right)^{2m} a^2 \left(\frac{\pi k|\varepsilon|}{L}\right)^{2\sigma} \\ &= \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2)\left(\frac{k}{L}\right)^{2q} a^2 (\pi|\varepsilon|)^{2\sigma} \\ &\leq a^2 (\pi|\varepsilon|)^{2\sigma} C'_q M^2. \end{aligned} \tag{B.12}$$

Therefore, if we set C'_q to be

$$C'_q = [(2\pi)^{2m} \pi^{2\sigma} a^2]^{-1} = (2^m \pi^q a)^{-2}, \tag{B.13}$$

the function $u(t)$ becomes (q, M)-regular by virtue of Lem. B.1. □

Theorem B.3 (Theorem 5.4 in the main text) *Let us set $0 < q \leq 1$. We consider an arbitrary function $u : [0, L] \to \mathbb{C}$ with the periodic boundary condition, and define its Fourier transform $\{u_k\}_{k\in\mathbb{Z}}$ by*

$$u(x) := \sum_{k\in\mathbb{Z}} u_k e^{2\pi i k x/L}. \tag{B.14}$$

Then, there exists a real number $C_q > 0$ such that if u is (q, M)-regular, the high-wavenumber components of $\{u_k\}$ can be bounded as

$$\sum_{k=K}^{\infty} (|u_k|^2 + |u_{-k}|^2) \leq C_q M^2 \left(\frac{L}{K}\right)^{2q} \tag{B.15}$$

for an arbitrary integer $K > 0$.

Proof. By Lem. B.1, a (q, M)-regular function $u(x)$ satisfies

$$\sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2) \sin^2 \frac{\pi k\varepsilon}{L} \leq M^2 |\varepsilon|^{2q}, \tag{B.16}$$

where we note that $0 < q \leq 1$ leads to $m = 0$ and $\sigma = q$. We restrict ε to a positive number and divide by ε^{2q+1} to obtain

$$\sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2) \frac{\sin^2 \frac{\pi k\varepsilon}{L}}{\varepsilon^{2q+1}} \leq \frac{M^2}{\varepsilon}. \tag{B.17}$$

Now, we fix some positive numbers $\beta > \alpha > 0$ and integrate Eq. (B.17) over $\frac{\alpha L}{K} \le \varepsilon \le \frac{\beta L}{K}$ to obtain

$$\int_{\frac{\alpha L}{K}}^{\frac{\beta L}{K}} d\varepsilon \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2) \frac{\sin^2 \frac{\pi k \varepsilon}{L}}{\varepsilon^{2q+1}} \le \int_{\frac{\alpha L}{K}}^{\frac{\beta L}{K}} d\varepsilon \frac{M^2}{\varepsilon} = M^2 \log \frac{\beta}{\alpha}. \tag{B.18}$$

We rewrite the leftmost-hand side of Eq. (B.18) by replacing ε with $t = \frac{L}{K}\varepsilon$:

$$\begin{aligned}
\int_{\frac{\alpha L}{K}}^{\frac{\beta L}{K}} d\varepsilon \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2) \frac{\sin^2 \frac{\pi k \varepsilon}{L}}{\varepsilon^{2q+1}} &= \int_{\alpha}^{\beta} \left(\frac{K}{L}\right)^{2q} dt \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2) \frac{\sin^2 \frac{\pi k t}{K}}{t^{2q+1}} \\
&= \left(\frac{K}{L}\right)^{2q} \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2) \int_{\alpha}^{\beta} dt \frac{\sin^2 \frac{\pi k t}{K}}{t^{2q+1}} \\
&= \left(\frac{K}{L}\right)^{2q} F\left(\frac{k}{K}\right) \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2),
\end{aligned} \tag{B.19}$$

where we have defined

$$F(s) = \int_{\alpha}^{\beta} dt \frac{\sin^2(\pi s t)}{t^{2q+1}}. \tag{B.20}$$

The function $F(s)$ is continuous and positive for $s \ge 1$. Furthermore, we also find that $F(s)$ does not vanish in the limit of $s \to \infty$ by virtue of the Riemann–Lebesgue lemma:

$$\lim_{s\to\infty} F(s) = \lim_{s\to\infty} \int_{\alpha}^{\beta} dt \frac{1 - \cos(2\pi s t)}{2t^{2q+1}} = \int_{\alpha}^{\beta} \frac{dt}{2t^{2q+1}} > 0. \tag{B.21}$$

Therefore, the infimum $m = \inf_{s\ge 1} F(s)$ must be a positive number.

Combining Eqs. (B.18) and (B.19), we obtain

$$\begin{aligned}
\left(\frac{K}{L}\right)^{2q} \sum_{k=K}^{\infty} m(|u_k|^2 + |u_{-k}|^2) &\le \left(\frac{K}{L}\right)^{2q} F\left(\frac{k}{K}\right) \sum_{k=1}^{\infty} (|u_k|^2 + |u_{-k}|^2) \\
&\le M^2 \log \frac{\beta}{\alpha}.
\end{aligned} \tag{B.22}$$

By setting $C_q = \frac{1}{m} \log \frac{\beta}{\alpha}$, the desired inequality (B.15) is obtained. □

Appendix C

Proof of Prop. 4.4

In this appendix, we prove Prop. 4.4, which bounds from below the infidelity of the postselected state in the Hamiltonian estimation.

Let us recall the basic notation regarding this problem. $H_{\boldsymbol{\theta}} = \sum_{j=1}^{m} \theta_j X_j$ is a Hamiltonian in a d-dimensional Hilbert space $\mathcal{H}_S = \mathbb{C}^d$ that linearly depends on the vector $\boldsymbol{\theta} \in \mathbb{R}^m$. The initial state $|\mathrm{M}\rangle = |\mathrm{M}\rangle_{\mathcal{H}_S}$ is set to be the *maximally entangled state* (MES) in $\mathcal{H}_S \otimes \mathcal{H}_A$, where the ancilla $\mathcal{H}_A$ is also d-dimensional. For an arbitrary operator O on $\mathcal{H}_S$, the MES satisfies

$$\langle \mathrm{M}|O|\mathrm{M}\rangle = \frac{1}{d} \operatorname{tr} O \tag{C.1}$$

and particularly,

$$\|O|\mathrm{M}\rangle\| = \frac{1}{d^{1/2}} \|O\|_{\mathrm{HS}} = \frac{1}{d^{1/2}} \big(\operatorname{tr}[O^\dagger O]\big)^{1/2}. \tag{C.2}$$

The probe state $|q_{\boldsymbol{\theta}}\rangle = e^{-i\tau H_{\boldsymbol{\theta}}}|\mathrm{M}\rangle$ is obtained by the Hamiltonian evolution on the MES over time τ. Them, we consider the projection operator P onto the $(m+1)$-dimensional subspace spanned by

$$|\mathrm{M}\rangle, X_1|\mathrm{M}\rangle, \ldots, X_m|\mathrm{M}\rangle, \tag{C.3}$$

and perform postselection onto this subspace:

$$|\tilde{q}_{\boldsymbol{\theta}}\rangle = \frac{P|q_{\boldsymbol{\theta}}\rangle}{\|P|q_{\boldsymbol{\theta}}\rangle\|}. \tag{C.4}$$

Then, the proposition in question is stated as follows:

Proposition C.1 (Proposition 4.4 in the main text) *Suppose that three constants $0 < c_1 < 1$, $c_2 \geq 1$, and $c_3 \geq 1$ are given. Then, there exists a constant $\alpha > 0$ such that following statement holds: Let $0 < b \leq c_3$, and $\{H_{\boldsymbol{\theta}}\}$ be a Hamiltonian model in d-dimensional Hilbert space such that*

$$\frac{1}{d} \operatorname{tr} H_{\boldsymbol{\theta}}^2 = b^2 \|\boldsymbol{\theta}\|^2, \qquad \frac{1}{d} \operatorname{tr} H_{\boldsymbol{\theta}}^4 \leq c_2^2 b^2 \|\boldsymbol{\theta}\|^4. \tag{C.5}$$

Then, the postselected state $|\tilde{q}_{\boldsymbol{\theta}}\rangle$ satisfies

$$1 - \big|\langle \tilde{q}_{\boldsymbol{\theta}^*} | \tilde{q}_{\boldsymbol{\theta}} \rangle\big|^2 \geq c_1 b^2 \tau^2 \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \tag{C.6}$$

for arbitrary $\|\boldsymbol{\theta}^*\|, \|\boldsymbol{\theta}\| \leq \alpha/\tau$.

Proof. From the definition in Eq. (C.4), we obtain

$$\begin{aligned}1-\left|\langle\tilde{q}_{\boldsymbol{\theta}^*}|\tilde{q}_{\boldsymbol{\theta}}\rangle\right|^2 &= \frac{\langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}^*}|P|q_{\boldsymbol{\theta}^*}\rangle - |\langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}^*}\rangle|^2}{\langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}^*}|P|q_{\boldsymbol{\theta}^*}\rangle}\\ &\geq \langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle\langle q_{\boldsymbol{\theta}^*}|P|q_{\boldsymbol{\theta}^*}\rangle - |\langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}^*}\rangle|^2\\ &= \langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle\langle \Delta q|P|\Delta q\rangle - |\langle q_{\boldsymbol{\theta}}|P|\Delta q\rangle|^2, \end{aligned} \tag{C.7}$$

where $|\Delta q\rangle = |q_{\boldsymbol{\theta}^*}\rangle - |q_{\boldsymbol{\theta}}\rangle$. The primary goal is to show the approximations

$$|q_{\boldsymbol{\theta}}\rangle \approx |q'_{\boldsymbol{\theta}}\rangle := (I - i\tau H_{\boldsymbol{\theta}})|\mathrm{M}\rangle \tag{C.8}$$

$$|\Delta q\rangle \approx |\Delta q'\rangle := (-i\tau\Delta H)|\mathrm{M}\rangle \qquad (\Delta H = H_{\boldsymbol{\theta}^*} - H_{\boldsymbol{\theta}}) \tag{C.9}$$

in a quantitative manner. To be more specific, we bound the errors

$$\epsilon_1 = \big\||q_{\boldsymbol{\theta}}\rangle - |q'_{\boldsymbol{\theta}}\rangle\big\|, \qquad \epsilon_2 = \big\||\Delta q\rangle - |\Delta q'\rangle\big\|. \tag{C.10}$$

In fact, with these approximations, the rightmost-hand side of Eq. (C.7) becomes

$$\begin{aligned}&\langle q'_{\boldsymbol{\theta}}|P|q'_{\boldsymbol{\theta}}\rangle\langle\Delta q'|P|\Delta q'\rangle - |\langle q'_{\boldsymbol{\theta}}|P|\Delta q'\rangle|^2,\\ &= \frac{1}{d}\|I - i\tau H_{\boldsymbol{\theta}}\|_{\mathrm{HS}}^2\cdot\frac{1}{d}\|-i\tau\Delta H\|_{\mathrm{HS}}^2 - \left|\frac{1}{d}\operatorname{tr}(I - i\tau H_{\boldsymbol{\theta}})^\dagger(-i\tau\Delta H)\right|^2\\ &= (1+\tau^2b^2\|\boldsymbol{\theta}\|^2)(\tau^2b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2) - \left(\tau^2b^2\|\boldsymbol{\theta}\|\cdot\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|\right)^2\\ &= \tau^2b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2, \end{aligned} \tag{C.11}$$

where we have used $\|\boldsymbol{\theta}\| \leq \alpha/\tau$.

First, let us show Eq. (C.8). Since $|e^{-\alpha} - 1 + i\alpha|^2 \leq frac\alpha^2 2$, we have

$$\begin{aligned}\epsilon_1 = \big\||q_{\boldsymbol{\theta}}\rangle - |q'_{\boldsymbol{\theta}}\rangle\big\| &= \big\||q_{\boldsymbol{\theta}}\rangle - (I - i\tau H_{\boldsymbol{\theta}})|M\rangle\big\|\\ &= \big\|(e^{-i\tau H_{\boldsymbol{\theta}}} - I + i\tau H_{\boldsymbol{\theta}})|M\rangle\big\|\\ &= \frac{1}{d^{1/2}}\operatorname{tr}\big[|e^{-i\tau H_{\boldsymbol{\theta}}} - I + i\tau H_{\boldsymbol{\theta}}|^2\big]\\ &= \frac{1}{d^{1/2}}\big(\operatorname{tr}\big[|e^{-i\tau H_{\boldsymbol{\theta}}} - I + i\tau H_{\boldsymbol{\theta}}|^2\big]\big)^{1/2}\\ &\leq \frac{1}{d^{1/2}}\left(\frac{\tau^4}{4}\operatorname{tr}[H_{\boldsymbol{\theta}}^4]\right)^{1/2}\\ &\leq \frac{1}{d^{1/2}}\left(\frac{\tau^4c_2^2b^2}{4}\|\boldsymbol{\theta}\|^4\right)^{1/2} \leq \frac{c_2b\alpha^2}{2}. \end{aligned} \tag{C.12}$$

Next, we treat the approximation in Eq. (C.9). First, the vector $|\Delta q\rangle$ can be rewritten as

$$\begin{aligned}|\Delta q\rangle &= (e^{-i\tau H_{\boldsymbol{\theta}^*}} - e^{-i\tau H_{\boldsymbol{\theta}}})|\mathrm{M}\rangle\\ &= -i\int_0^\tau dt e^{-itH_{\boldsymbol{\theta}^*}}(H_{\boldsymbol{\theta}^*} - H_{\boldsymbol{\theta}})e^{-i(\tau-t)H_{\boldsymbol{\theta}}}|\mathrm{M}\rangle \end{aligned} \tag{C.13}$$

and hence

$$|\Delta q\rangle - |\Delta q'\rangle = -i\int_0^\tau dt[e^{-itH_{\boldsymbol{\theta}^*}}\Delta H e^{-i(\tau-t)H_{\boldsymbol{\theta}}} - \Delta H]|\mathrm{M}\rangle. \tag{C.14}$$

Here, the Hilbert-Schmidt norm can be evaluated by the Hölder inequality as

$$\begin{aligned}
&\|e^{-itH_{\boldsymbol{\theta}^*}}\Delta H e^{-i(\tau-t)H_{\boldsymbol{\theta}}} - \Delta H\|_{\mathrm{HS}} \\
&\le \|(e^{-itH_{\boldsymbol{\theta}^*}} - I)\Delta H e^{-i(\tau-t)H_{\boldsymbol{\theta}}}\|_{\mathrm{HS}} + \|e^{-itH_{\boldsymbol{\theta}^*}}\Delta H(e^{-i(\tau-t)H_{\boldsymbol{\theta}}} - I)\|_{\mathrm{HS}} \\
&\le \left(\mathrm{tr}[|e^{-itH_{\boldsymbol{\theta}^*}} - I|^4]\right)^{1/4}\left(\mathrm{tr}[(\Delta H)^4]\right)^{1/4} + \left(\mathrm{tr}[(\Delta H)^4]\right)^{1/4}\left(\mathrm{tr}[|e^{-i(\tau-t)H_{\boldsymbol{\theta}}} - I|^4]\right)^{1/4} \\
&\le \left(\mathrm{tr}[t^4 H_{\boldsymbol{\theta}^*}^4]\right)^{1/4}\left(\mathrm{tr}[(\Delta H)^4]\right)^{1/4} + \left(\mathrm{tr}[(\Delta H)^4]\right)^{1/4}\left(\mathrm{tr}[(\tau-t)^4 H_{\boldsymbol{\theta}}^4]\right)^{1/4} \\
&\le d^{1/2} t c_2 b\|\boldsymbol{\theta}^*\|\cdot\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| + d^{1/2}(\tau-t)c_2 b\|\boldsymbol{\theta}\|\cdot\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \\
&\le d^{1/2} c_2 b\alpha\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|.
\end{aligned} \tag{C.15}$$

Therefore,

$$\begin{aligned}
\epsilon_2 = \||\Delta q\rangle - |\Delta q'\rangle\| &= \frac{1}{d^{1/2}}\left\|\int_0^\tau dt[e^{-itH_{\boldsymbol{\theta}^*}}\Delta H e^{-i(\tau-t)H_{\boldsymbol{\theta}}} - \Delta H]\right\|_{\mathrm{HS}} \\
&\le \frac{\tau}{d^{1/2}}\cdot d^{1/2} c_2 b\alpha\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| = c_2 b\alpha\tau\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|.
\end{aligned} \tag{C.16}$$

Now, we evaluate the approximation error of Eq. (C.11) in terms of e ϵ_1 and ϵ_2:

$$\begin{aligned}
&\left|\langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle\langle\Delta q|P|\Delta q\rangle - \langle q'_{\boldsymbol{\theta}}|P|q'_{\boldsymbol{\theta}}\rangle\langle\Delta q'|P|\Delta q'\rangle\right| \\
&\le \langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle|\langle\Delta q'|P|\Delta q'\rangle - \langle\Delta q|P|\Delta q\rangle| + |\langle q'_{\boldsymbol{\theta}}|P|q'_{\boldsymbol{\theta}}\rangle - \langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle|\langle\Delta q'|P|\Delta q'\rangle \\
&\le \epsilon_2(2\||\Delta q'\rangle\| + \epsilon_2) + \epsilon_1(2+\epsilon_1)\||\Delta q'\rangle\|^2 \\
&\le c_2\alpha\left(2 + c_2\alpha + b\alpha + \frac{c_2 b^2\alpha^3}{4}\right)\tau^2 b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2;
\end{aligned} \tag{C.17}$$

$$\begin{aligned}
&\left||\langle q_{\boldsymbol{\theta}}|P|\Delta q\rangle|^2 - |\langle q'_{\boldsymbol{\theta}}|P|\Delta q'\rangle|^2\right| \\
&\le \left||\langle q_{\boldsymbol{\theta}}|P|\Delta q\rangle|^2 - |\langle q_{\boldsymbol{\theta}}|P|\Delta q'\rangle|^2\right| + \left||\langle q_{\boldsymbol{\theta}}|P|\Delta q'\rangle|^2 - |\langle q'_{\boldsymbol{\theta}}|P|\Delta q'\rangle|^2\right| \\
&\le \epsilon_2(2\||\Delta q'\rangle\| + \epsilon_2) + \epsilon_1(2+\epsilon_1)\||\Delta q'\rangle\|^2 \\
&\le c_2\alpha\left(2 + c_2\alpha + b\alpha + \frac{c_2 b^2\alpha^3}{4}\right)\tau^2 b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2.
\end{aligned} \tag{C.18}$$

Combining Eqs. (C.7), (C.11), (C.17), and (C.18), we finally obtain

$$\begin{aligned}
1 - |\langle\tilde{q}_{\boldsymbol{\theta}^*}|\tilde{q}_{\boldsymbol{\theta}}\rangle|^2 &\ge \langle q_{\boldsymbol{\theta}}|P|q_{\boldsymbol{\theta}}\rangle\langle\Delta q|P|\Delta q\rangle - |\langle q_{\boldsymbol{\theta}}|P|\Delta q\rangle|^2 \\
&\ge \tau^2 b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 - 2c_2\alpha\left(c_2\alpha + 2 + b\alpha + \frac{c_2 b^2\alpha^3}{4}\right)\tau^2 b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \\
&\ge \left[1 - 2c_2\alpha\left(2 + c_2\alpha + c_3\alpha + \frac{c_2 c_3^2\alpha^3}{4}\right)\right]\tau^2 b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2 \\
&\ge c_1\tau^2 b^2\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2
\end{aligned} \tag{C.19}$$

by taking sufficiently small $\alpha > 0$. □

Appendix D

Bounds on x-error for multi-photon probe states

D.1 Rigorous upper bound for general Gaussian states

We recall the assumption in Sec. 6.3.1. The function $\varphi(x)$ is a (q, M)-regular function with $q = 1 + \sigma$, and hence

$$f(x) := \hbar\varphi'(x) \tag{D.1}$$

becomes a $(\sigma, \hbar M)$-regular function. Moreover, the initial state has an N-variate Gaussian wavefunction whose position covariance C_x and momentum covariance C_p satisfies the uncertainty relation $C_p = (\hbar/2)C_x^{-1}$.

For $N = 1$, the x-error E_x is defined in Eq. (6.57) by

$$\begin{aligned} E_x^2 = R(x_0) &:= \int [f(x) - \langle p^* \rangle]^2 dx \rho_s(x_0 - x) \\ &= \int [f(x+\epsilon) - \langle p^* \rangle]^2 d\epsilon \rho_s(\epsilon) \end{aligned} \tag{D.2}$$

where $\langle p^* \rangle = \int f(x+\epsilon) d\epsilon \rho_s(\epsilon)$. In this case, we can bound the mean-square x-error from above:

$$\overline{E_x^2} = \int_0^L \frac{dx_0}{L} R(x_0) \leq C^2 \hbar^2 M^2 s^{2\sigma}, \qquad C^2 = \pi^{-1/2}\Gamma\left(\sigma + \frac{1}{2}\right). \tag{D.3}$$

owing to the $(\sigma, \hbar M)$-regularity of f (see Eq. (6.60)).

In the multi-photon case, the x-error is defined from $E_x^2 = N^{-2}\mathbf{1}^{\mathrm{T}} V_x \mathbf{1}$, where the matrix V_x is defined in Eq. (6.45):

$$[V_x]_{jk} = \int [f(x_j) - \langle p^* \rangle][f(x_k) - \langle p^* \rangle] dx \rho(\boldsymbol{x} - \boldsymbol{x}_0; C_x). \tag{D.4}$$

If we denote by Σ the 2×2 covariance matrix with respect to x_j and x_k:

$$\Sigma := \begin{pmatrix} [C_x]_{jj} & [C_x]_{jk} \\ [C_x]_{kj} & [C_x]_{kk} \end{pmatrix}, \tag{D.5}$$

the matrix entries $[V_x]_{jk}$ can be rewritten as

$$[V_x]_{jk} = \iint [f(x_0+\epsilon_j) - \langle p^* \rangle][f(x_0+\epsilon_k) - \langle p^* \rangle] d\epsilon_j d\epsilon_k \rho(\epsilon_j, \epsilon_k; \Sigma). \tag{D.6}$$

We note that $[C_x]_{jj} = [C_x]_{kk} = s^2$ is required since we measure the wavelet of lengthscale s.

Lemma D.1 *The magnitude of each entry in the covariance matrix V_x can be bounded from above by*

$$|[V_x]_{jk}| \leq [R(x_0)\tilde{R}(x_0)]^{1/2}. \tag{D.7}$$

Here, $\tilde{R}(x_0)$ is defined by the following equations:

$$\tilde{R}(x_0) = \int [\tilde{f}(x+\epsilon) - \langle p^* \rangle]^2 d\epsilon \rho_{s_1}(\epsilon), \tag{D.8}$$

$$\tilde{f}(x) = \int dx' f(x') \rho_{s_2}(x - x'), \tag{D.9}$$

$$s_1 = \frac{|[C_x]_{jk}|}{s}, \qquad s_2 = \left(s^2 - s_1^2\right)^{1/2}. \tag{D.10}$$

First, we show the following lemma:

Proof. We begin with Eq. (D.6). The pair of variables (ϵ_j, ϵ_k) are subject to the Gaussian distribution with covariance Σ. By defining $\epsilon_k = \frac{s_1}{s}\epsilon_j + \epsilon'_k$, ϵ_j and ϵ'_k becomes independent Gaussian variables with respective variance s^2 and s_2^2. Hence we may write

$$\begin{aligned}
[V_x]_{jk} &= \iint [f(x_0+\epsilon_j) - \langle p^* \rangle][f(x_0 + \tfrac{s_1}{s}\epsilon_j + \epsilon'_k) - \langle p^* \rangle] d\epsilon_j d\epsilon'_k \rho_s(\epsilon_j)\rho_{s_2}(\epsilon'_k) \\
&= \int [f(x_0+\epsilon_j) - \langle p^* \rangle][\tilde{f}(x_0 + \tfrac{s_1}{s}\epsilon_j) - \langle p^* \rangle] d\epsilon_j \rho_s(\epsilon_j).
\end{aligned} \tag{D.11}$$

Therefore, from the Schwartz inequality, we have

$$\begin{aligned}
|[V_x]_{jk}|^2 &\leq \int [f(x_0+\epsilon_j) - \langle p^* \rangle]^2 d\epsilon_j \rho_s(\epsilon_j) \int [\tilde{f}(x_0 + \tfrac{s_1}{s}\epsilon_j) - \langle p^* \rangle] d\epsilon_j \rho_s(\epsilon_j) \\
&= \int [f(x_0+\epsilon) - \langle p^* \rangle]^2 d\epsilon \rho_s(\epsilon) \int \left[\tilde{f}(x_0+\epsilon') - \langle p^* \rangle\right]^2 d\epsilon' \rho_{s_1}(\epsilon') \\
&= R(x_0)\tilde{R}(x_0).
\end{aligned}$$

□

In this lemma, the magnitude of covariance is bounded by $R(x_0)$ and its analogue $\tilde{R}(x_0)$, with f and s replaced by $\tilde{f}$ and s_1 defined in Eqs. (D.9) and (D.10). We also note that $\langle p^* \rangle$ is invariant with this replacement:

$$\langle p^* \rangle = \int f(x+\epsilon) d\epsilon \rho_s(\epsilon) = \int \tilde{f}(x+\epsilon') d\epsilon' \rho_{s_1}(\epsilon'). \tag{D.12}$$

With these facts in mind, the average covariance $\overline{V_x}$ can be evaluated.

Proposition D.2 *Let* $\overline{V_x} = \int_0^L \frac{dx_0}{L} V_x$ *be the average of the covariance matrix* V_x *over* $[0, L]$*. Then, we have*

$$|\overline{V_x}| \leq C^2 \hbar^2 M^2 |[C_x]_{jk}|^\sigma. \tag{D.13}$$

Proof. First, we show that $\tilde{f}$ is also $(\sigma, \hbar M)$-regular:

$$\begin{aligned}\int [\tilde{f}(x+\epsilon) - \tilde{f}(x)]^2 dx &\leq \int \left| \int [\tilde{f}(x+\epsilon+\epsilon') - \tilde{f}(x+\epsilon')] \rho_{s_2}(\epsilon') d\epsilon' \right|^2 dx \\ &\leq \iint |\tilde{f}(x+\epsilon+\epsilon') - \tilde{f}(x+\epsilon')|^2 \rho_{s_2}(\epsilon') d\epsilon' dx \\ &\leq \iint |\tilde{f}(x'+\epsilon) - \tilde{f}(x')|^2 \rho_{s_2}(\epsilon') d\epsilon' dx \leq M^2 \sigma^{2\sigma}. \end{aligned} \tag{D.14}$$

Therefore, the average of $\tilde{R}(x_0)$ can be bounded from above form the same argument as (D.3), when we obtain

$$\int_0^L \frac{dx_0}{L} \tilde{R}(x_0) \leq C^2 \hbar^2 M^2 s_1^{2\sigma}. \tag{D.15}$$

Finally, Lemma D.1 yields

$$\begin{aligned}|\overline{V_x}| &\leq \int \frac{dx_0}{L} [R(x_0) \tilde{R}(x_0)]^{1/2} \\ &\leq \left[\int \frac{dx_0}{L} R(x_0) \int \frac{dx_0}{L} \tilde{R}(x_0) \right]^{1/2} \\ &\leq C\hbar M s^\sigma \cdot C\hbar M s_1^\sigma = C^2 \hbar^2 M^2 |[C_x]_{jk}|^\sigma, \end{aligned} \tag{D.16}$$

where we have employed Eqs. (D.3) and (D.15). □

Now, recalling that $E_x^2 = N^{-2} \mathbf{1}^{\mathrm{T}} V_x \mathbf{1}$, we derive the following inequality:

$$\overline{E_x^2} = N^{-2} \mathbf{1}^{\mathrm{T}} \overline{V_x} \mathbf{1} \leq \frac{C^2 \hbar^2 M^2}{N^2} \sum_{j,k=1}^{N} |[C_x]_{jk}|^\sigma. \tag{D.17}$$

In this way, we have a rigorous upper bound on the mean-square x-error for the multi-photon Gaussian states. Furthermore, this upper bound optimal in the leading term of s. In fact, the inequality in Eq. (D.17) can be asymptotically saturated in the limit $s \to 0$ by setting $\varphi(x)$ to be a piecewise quadratic function.

D.2 Restriction to specific Gaussian states

In the main text, the estimation error is optimized in terms of two variables $\mathbf{1}^{\mathrm{T}} C_p \mathbf{1}$ and $\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}$ under the following constraints:

$$(\mathbf{1}^{\mathrm{T}} C_p \mathbf{1}) \cdot (\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}) \leq \frac{\hbar^2 N^2}{4}, \tag{D.18}$$

$$(\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}) \leq N^2 s^2. \tag{D.19}$$

Furthermore, since we are interested in the benefit of quantum correlation, we may exclude the case in which the momentum is positively correlated and the position is negatively correlated, since such states result in larger error than separable probe states. This implies $[C_x]_{jk} \geq 0$ for every j, k, and particularly,

$$Ns^2 \leq \mathbf{1}^{\mathrm{T}} C_x \mathbf{1} \leq N^2 s^2. \tag{D.20}$$

Therefore, we would like to evaluate the upper bound in Eq. (D.17) in terms of $\mathbf{1}^{\mathrm{T}} C_x \mathbf{1} = \sum_{j,k} [C_x]_{jk}$. In fact, by using the fact that $0 \leq [C_x]_{jk} \leq s^2$ for every j, k, a lower bound and an upper bound can be found:

$$s^{2(\sigma-1)} (\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}) \leq \sum_{j,k=1}^{N} [C_x]_{jk}^{\sigma} \leq N^{2(1-\sigma)} (\mathbf{1}^{\mathrm{T}} C_x \mathbf{1})^{\sigma}. \tag{D.21}$$

In fact, either inequality can be saturated up to the leading term in N. Intuitively speaking, the lower bound is saturated (i.e. the best cases) when the probe state consists of small separate groups of strongly correlated photons, and the upper bound is saturated (i.e. the worst cases) when the entire photons are weakly correlated.

When the lower bound in Eq. (D.21) is saturated, the approximate inequality in Eq. (6.64) becomes the exact inequality:

$$\overline{E_x^2} \leq \hbar^2 C^2 M^2 \frac{\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}}{N^2 s^{2-2\sigma}}. \tag{D.22}$$

On the other hand, when the upper bound in Eq. (D.21) is saturated, we only obtain a looser upper bound on the mean-square x-error:

$$\overline{E_x^2} \leq \hbar^2 C^2 M^2 \frac{(\mathbf{1}^{\mathrm{T}} C_x \mathbf{1})^{\sigma}}{N^{2\sigma}}. \tag{D.23}$$

This implies that the exponent in N is different between the best case and the worst case. Since we are interested in the least estimation error possible, we would like to restrict the position covariance C_x such that the lower bound is saturated.

Now, we require that the covariance matrix be written in a diagonal block from, with each block $J_n(\varepsilon)$ being

$$J_n(\varepsilon) = \begin{pmatrix} 1 & 1-\frac{n}{n-1}\varepsilon & \cdots & 1-\frac{n}{n-1}\varepsilon \\ 1-\frac{n}{n-1}\varepsilon & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1-\frac{n}{n-1}\varepsilon \\ 1-\frac{n}{n-1}\varepsilon & \cdots & 1-\frac{n}{n-1}\varepsilon & 1 \end{pmatrix} \qquad (n \times n \text{ matrix}). \tag{D.24}$$

In particular, we assume that C_x consists of l blocks of $J_{n-1}(\varepsilon)$ and $k-l$ blocks of $J_n(\varepsilon)$:

$$C_x = s^2 \begin{pmatrix} J_{n-1}(\varepsilon) & & & & & \\ & \ddots & & & 0 & \\ & & J_{n-1}(\varepsilon) & & & \\ & & & J_n(\varepsilon) & & \\ & 0 & & & \ddots & \\ & & & & & J_n(\epsilon) \end{pmatrix}, \tag{D.25}$$

Here, the integers n, k, l must satisfy $N = nk - l$. In fact, the lower bound in Eq. (D.21) can be saturated in the limit $\varepsilon \to 0$.

However, this type of restriction of the covariance matrix causes a problem: We may not treat $\mathbf{1}^{\mathrm{T}} C_p \mathbf{1}$ and $\mathbf{1}^{\mathrm{T}} C_x \mathbf{1}$ as free variables satisfying Eqs. (D.18) and (D.20), as we did in the main text. To be more specific, let η be a number between 1 and N. We expect that there exists a covariance matrix C_x such that

$$\mathbf{1}^{\mathrm{T}} C_x \mathbf{1} = N\eta s^2, \qquad \mathbf{1}^{\mathrm{T}} C_p \mathbf{1} = \frac{\hbar^2 N}{4\eta s^2} \tag{D.26}$$

hold, but this is not always possible for arbitrary η. This problem can be solved by compromising with an additional factor of 2:

Theorem D.3 *Let $1 \leq \eta \leq N$. There exists a matrix C_x such that*

$$\overline{E_x^2} \leq \frac{C^2\hbar^2 M^2}{N^2} \sum_{j,k=1}^{N} \left|[C_x]_{jk}\right|^{\sigma} \leq 2 \cdot \frac{C^2\hbar^2 M^2 \eta s^{2\sigma}}{N}, \tag{D.27}$$

$$E_p^2 = \frac{\mathbf{1}^{\mathrm{T}} C_p \mathbf{1}}{N^2} \leq 2 \cdot \frac{\hbar^2}{4N\eta s^2}. \tag{D.28}$$

This implies that the approximated inequalities in Eq. (6.67) *is broken by at most a factor of* 2 *by choosing an appropriate Gaussian state.*

Proof. Let k be an integer such that $k \leq N/\eta < k+1$, and n, l be integers such that $N = nk - l$ and $0 \leq l < k$. Such n can be determined from $n - 1 < N/k \leq n$, which implies $n < \eta + 1$. We consider the covariance matrix C_x in Eq. (D.25) with such integers n, k, l.

The number of nonzero entries in C_x is at most

$$l^2(n-1)^2 + (k-l)^2 n^2 \leq n[(n-1)l + n(k-l)] = nN \leq (\eta+1)N. \tag{D.29}$$

Therefore, we obtain

$$\sum_{j,k=1}^{N} \left|[C_x]_{jk}\right|^{\sigma} \leq (\eta+1)Ns^{2\sigma} \leq 2\eta N s^{2\sigma}, \tag{D.30}$$

which yields Eq. (D.27).

Next, we consider the momentum covariance:

$$C_x = \frac{\hbar^2}{4s^2}\begin{pmatrix} J_{n-1}^{-1}(\varepsilon) & & & & & \\ & \ddots & & & 0 & \\ & & J_{n-1}^{-1}(\varepsilon) & & & \\ & & & J_n^{-1}(\varepsilon) & & \\ & 0 & & & \ddots & \\ & & & & & J_n^{-1}(\epsilon) \end{pmatrix}. \tag{D.31}$$

It can be straightforwardly checked that the matrix $J_n^{-1}(\varepsilon)$ satisfies $\mathbf{1}^{\mathrm{T}} J_n^{-1}(\varepsilon)\mathbf{1} = \frac{1}{1-\varepsilon}$, and therefore

$$\mathbf{1}^{\mathrm{T}} C_p \mathbf{1} = \frac{\hbar^2}{4s^2} \cdot \frac{k}{1-\varepsilon}. \tag{D.32}$$

When $\eta = N$, we have $k = 1$ and $\frac{k}{1-\varepsilon} = \frac{1}{1-\varepsilon} \leq 2$ for $\varepsilon \leq \frac{1}{2}$. When $1 \leq \eta < N$, we have $k < N/\eta + 1$ and

$$\mathbf{1}^{\mathrm{T}} C_p \mathbf{1} = \frac{k}{1-\varepsilon} < \frac{1}{1-\varepsilon}\left(\frac{N}{\eta} + 1\right) \leq \frac{2N}{\eta} \tag{D.33}$$

for sufficiently small ε. In either case, the inequality (D.28) has been shown to hold. □

Bibliography

[1] D. Deutsch and R. Jozsa. "Rapid Solution of Problems by Quantum Computation." *Proc. R. Soc. A* **439**, 553–558 (1992).

[2] P. W. Shor. "Algorithms for quantum computation: Discrete logarithms and factoring." In *Proc. 35th Ann. Symp. on Fundamentals of Computer Science*. IEEE, 124–134 (1994).

[3] L. K. Grover. "A fast quantum mechanical algorithm for database search." In *Proc. 28th Ann. ACM Symp. on Theory of computing*. ACM, 212–219 (1996).

[4] P. W. Shor and J. Preskill. "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol." *Phys. Rev. Lett.* **85**, 441–444 (2000).

[5] N. Gisin *et al.* "Quantum cryptography." *Rev. Mod. Phys.* **74**, 145–195 (2002).

[6] M. Hillery, V. Bužek, and A. Berthiaume. "Quantum secret sharing." *Phys. Rev. A* **59**, 1829–1834 (1999).

[7] V. Giovannetti, S. Lloyd, and L. Maccone. "Quantum Metrology." *Phys. Rev. Lett.* **96**, 010401 (2006).

[8] V. Giovannetti, S. Lloyd, and L. Maccone. "Advances in quantum metrology." *Nat. Photonics* **5**, 222 (2011).

[9] P. Bouyer and M. A. Kasevich. "Heisenberg-limited spectroscopy with degenerate Bose-Einstein gases." *Phys. Rev. A* **56**, R1083–R1086 (1997).

[10] M. D. Vidrighin *et al.* "Joint estimation of phase and phase diffusion for quantum metrology." *Nat. Commun.* **5**, 3532 (2014).

[11] M. J. Holland and K. Burnett. "Interferometric detection of optical phase shifts at the Heisenberg limit." *Phys. Rev. Lett.* **71**, 1355–1358 (1993).

[12] M. de Burgh and S. D. Bartlett. "Quantum methods for clock synchronization: Beating the standard quantum limit without entanglement." *Phys. Rev. A* **72**, 042301 (2005).

[13] E. M. Kessler *et al.* "Heisenberg-Limited Atom Clocks Based on Entangled Qubits." *Phys. Rev. Lett.* **112**, 190403 (2014).

[14] M. Fraas. "An Analysis of the Stationary Operation of Atomic Clocks." *Commun. in Math. Phys.* **348**, 363–393 (2016).

[15] C. M. Caves. "Quantum-mechanical noise in an interferometer." *Phys. Rev. D* **23**, 1693–1708 (1981).

[16] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kołodyński. "Quantum Limits in Optical Interferometry." *Progr. Opt.* **60**, 345–435 (2015).

[17] R. Schnabel *et al.* "Quantum metrology for gravitational wave astronomy." *Nat. Commun.* **1**, 121 (2010).

[18] D. W. Berry and H. M. Wiseman. "Adaptive phase measurements for narrowband squeezed beams." *Phys. Rev. A* **73**, 063824 (2006).

[19] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă. "The elusive Heisenberg limit in quantum-enhanced metrology." *Nat. Commun.* **3**, 1063 (2012).

[20] A. W. Chin, S. F. Huelga, and M. B. Plenio. "Quantum Metrology in Non-Markovian Environments." *Phys. Rev. Lett.* **109**, 233601 (2012).

[21] R. Chaves *et al.* "Noisy Metrology beyond the Standard Quantum Limit." *Phys. Rev. Lett.* **111**, 120401 (2013).

[22] S. Deffner and E. Lutz. "Quantum Speed Limit for Non-Markovian Dynamics." *Phys. Rev. Lett.* **111**, 010402 (2013).

[23] A. Smirne *et al.* "Ultimate Precision Limits for Noisy Frequency Estimation." *Phys. Rev. Lett.* **116**, 120801 (2016).

[24] M. Szczykulska, T. Baumgratz, and A. Datta. "Multi-parameter quantum metrology." *Adv. Phys. X* **1**, 621–639 (2016).

[25] C. Macchiavello. "Optimal estimation of multiple phases." *Phys. Rev. A* **67**, 062302 (2003).

[26] H. Imai and A. Fujiwara. "Geometry of optimal estimation scheme for $SU(D)$ channels." *J. Phys. A* **40**, 4391–4400 (2007).

[27] N. Kura. "Geometrical analysis on multi-parameter Hamiltonian problems and their time complexity." Master's thesis, University of Tokyo (2017).

[28] S.-i. Amari. *Differential-Geometrical Methods in Statistics*. Springer New York (1985).

[29] H. Cramér. *Mathematical methods of statistics*. Princeton University Press (1999).

[30] H. Yuan. "Sequential Feedback Scheme Outperforms the Parallel Scheme for Hamiltonian Parameter Estimation." *Phys. Rev. Lett.* **117**, 160801 (2016).

[31] N. Kura and M. Ueda. "Finite-error metrological bounds on multiparameter Hamiltonian estimation." *Phys. Rev. A* **97**, 012101 (2018).

[32] L. Galleani and P. Tavella. "Detection and identification of atomic clock anomalies." *Metrologia* **45**, S127–S133 (2008).

[33] The LIGO Scientific Collaboration. "A gravitational wave observatory operating beyond the quantum shot-noise limit." *Nat. Phys.* **7**, 962 (2011).

[34] The EHT Collaboration. "First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole." *Astrophys. J.* **875**, L4 (2019).

[35] The EHT Collaboration. "First M87 Event Horizon Telescope Results. III. Data Processing and Calibration." *Astrophys. J.* **875**, L3 (2019).

[36] M. D. Donsker. "Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems." *Ann. Math. Stat.* **23**, 277–281 (1952).

[37] G. Wahba. "Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation." *Ann. Stat.* **3**, 15–29 (1975).

[38] M. P. Wand and M. C. Jones. *Kernel smoothing*. Chapman and Hall/CRC (1994).

[39] D. W. Berry and H. M. Wiseman. "Adaptive quantum measurements of a continuously varying phase." *Phys. Rev. A* **65**, 043803 (2002).

[40] D. W. Berry, M. J. W. Hall, and H. M. Wiseman. "Stochastic Heisenberg Limit: Optimal Estimation of a Fluctuating Phase." *Phys. Rev. Lett.* **111**, 113601 (2013).

[41] D. W. Berry *et al.* "Quantum Bell-Ziv-Zakai Bounds and Heisenberg Limits for Waveform Estimation." *Phys. Rev. X* **5**, 031018 (2015).

[42] M. Tsang. "Subdiffraction incoherent optical imaging via spatial-mode demultiplexing." *New J. Phys.* **19**, 023054 (2017).

[43] H. T. Dinani and D. W. Berry. "Adaptive estimation of a time-varying phase with a power-law spectrum via continuous squeezed states." *Phys. Rev. A* **95**, 063821 (2017).

[44] J.-F. Cardoso. "Blind signal separation: statistical principles." *Proc. IEEE* **86**, 2009–2025 (1998).

[45] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media (2013).

[46] J. Canny. "A Computational Approach to Edge Detection." *IEEE Trans. Pattern Analysis and Machine Intelligence* **8**, 679–698 (1986).

[47] S. Geman and D. Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-6**, 721–741 (1984).

[48] A. Cochocki and R. Unbehauen. *Neural networks for optimization and signal processing*. John Wiley & Sons, Inc. (1993).

[49] S. Xie and Z. Tu. "Holistically-Nested Edge Detection." In *The IEEE International Conference on Computer Vision (ICCV)* (2015).

[50] S. Mallat and S. Zhong. "Signal characterization from multiscale edges." In *Proc.. 10th International Conference on Pattern Recognition.* **i**, 891–896 (1990).

[51] S. Zhong and S. Mallat. "Characterization of Signals from Multiscale Edges." *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**, 710–732 (1992).

[52] Y. Bromberg, O. Katz, and Y. Silberberg. "Ghost imaging with a single detector." *Phys. Rev. A* **79**, 053840 (2009).

[53] G. Brida, M. Genovese, and I. Ruo Berchera. "Experimental realization of sub-shot-noise quantum imaging." *Nat. Photonics* **4**, 227 (2010).

[54] S. R. Ghaleh *et al.* "Improved edge detection in computational ghost imaging by introducing orbital angular momentum." *Appl. Opt.* **57**, 9609–9614 (2018).

[55] S. Zacks. *Parametric statistical inference: basic theory and modern approaches*. Elsevier (2014).

[56] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press (2011).

[57] D. Petz. "Monotone metrics on matrix spaces." *Linear Algebra Appl.* **244**, 81–96 (1996).

[58] P. Gibilisco and T. Isola. "On the characterisation of paired monotone metrics." *Ann. Inst. Stat. Math.* **56**, 369–381 (2004).

[59] A. Fujiwara and H. Nagaoka. "Quantum Fisher metric and estimation for pure state models." *Phys. Letters A* **201**, 119–124 (1995).

[60] A. Hayashi, T. Hashimoto, and M. Horibe. "Reexamination of optimal quantum state estimation of pure states." *Phys. Rev. A* **72**, 032325 (2005).

[61] V. S. Varadarajan. *Lie groups, Lie algebras, and their representations*. Springer Science & Business Media (2013).

[62] M.-D. Choi. "Completely positive linear maps on complex matrices." *Linear Algebra Appl.* **10**, 285–290 (1975).

[63] A. Jamiołkowski. "Linear transformations which preserve trace and positive semidefiniteness of operators." *Rep. Math. Phys.* **3**, 275–278 (1972).

[64] B. E. Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*. Springer Science & Business Media (2013).

[65] P. C. Humphreys *et al.* "Quantum Enhanced Multiple Phase Estimation." *Phys. Rev. Lett.* **111**, 070403 (2013).

[66] W. F. Stinespring. "Positive functions on C*-algebras." *Proc. Am. Math. Soc.* **2**, 211–216 (1955).

[67] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Quantum Circuit Architecture." *Phys. Rev. Lett.* **101**, 60401 (2008).

[68] B. L. Higgins *et al.* "Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements." *New J. Phys.* **11**, 073023 (2009).

[69] W. S. Cleveland. "Robust Locally Weighted Regression and Smoothing Scatterplots." *J. Am. Stat. Assoc.* **74**, 829–836 (1979).

[70] A. Schaum. "Theory and Design of Local Interpolators." *CVGIP: Graphical Models and Image Processing* **55**, 464–481 (1993).

[71] E. H. W. Meijering, K. J. Zuiderveld, and M. A. Viergever. "Image reconstruction by convolution with symmetrical piecewise nth-order polynomial kernels." *IEEE Trans. Image Processing* **8**, 192–201 (1999).

[72] T. Blu, P. Thevenaz, and M. Unser. "Complete parameterization of piecewise-polynomial interpolation kernels." *IEEE Trans. Image Processing* **12**, 1297–1309 (2003).

[73] G. H. Hardy. "Weierstrass's non-differentiable function." *Trans. Am. Math. Soc.* **17**, 301–325 (1916).

[74] H. L. Van Trees. *Detection, estimation, and modulation theory.* John Wiley & Sons (2004).

[75] Y. K. Belayev. "CONTINUITY AND HOLDER'S CONDITIONS FOR SAMPLE FUNCTIONS OF STATIONARY GAUSSIAN PROCESSES." In *Proc. 4th Berkeley Symp. Math. Statist. and Prob.* 23–33 (1961).

[76] N. Kono. "On the modulus of continuity of sample functions of Gaussian processes." *J. Math. Kyoto Univ.* **10**, 493–536 (1970).

[77] P. Bocchieri and A. Loinger. "Quantum Recurrence Theorem." *Phys. Rev.* **107**, 337–338 (1957).

[78] A. Peres. "Recurrence Phenomena in Quantum Dynamics." *Phys. Rev. Lett.* **49**, 1118 (1982).

[79] H. Chernoff. "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations." *Ann. Math. Stat.* **23**, 493–507 (1952).

[80] W. Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables." *J. Am. Stat. Assoc.* **58**, 13–30 (1963).

[81] A. Fujiwara. "Strong consistency and asymptotic efficiency for adaptive quantum estimation problems." *J. Phys. A* **39**, 12489 (2006).

[82] E. Nitzan, T. Routtenberg, and J. Tabrikian. "A new class of Bayesian cyclic bounds for periodic parameter estimation." *IEEE Trans. Signal Processing* (2016).

[83] S. Prossdorf. "Convergence of Fourier series of Hölder continuous functions." *Mathematische Nachrichten* **69**, 7–14 (1975).

[84] C. J. Stone. "Optimal Rates of Convergence for Nonparametric Estimators." *Ann. Stat.* **8**, 1348–1360 (1980).

[85] B. L. Higgins *et al.* "Entanglement-free Heisenberg-limited phase estimation." *Nature* **450**, 393–396 (2007).

[86] D. V. Strekalov *et al.* "Observation of Two-Photon "Ghost" Interference and Diffraction." *Phys. Rev. Lett.* **74**, 3600–3603 (1995).

[87] R. S. Bennink, S. J. Bentley, and R. W. Boyd. ""Two-Photon" Coincidence Imaging with a Classical Source." *Phys. Rev. Lett.* **89**, 113601 (2002).

[88] R. L. Hudson. "When is the wigner quasi-probability density non-negative?" *Rep. Math. Phys.* **6**, 249–252 (1974).

[89] P. Dallaire, C. Besse, and B. Chaib-draa. "An approximate inference with Gaussian process to latent functions from uncertain data." *Neurocomputing* **74**, 1945–1955 (2011).

[90] M. I. Kolobov. "The spatial behavior of nonclassical light." *Rev. Mod. Phys.* **71**, 1539–1589 (1999).

[91] A. Christ *et al.* "Probing multimode squeezing with correlation functions." *New J. Phys.* **13**, 033027 (2011).

[92] D. F. Smirnov and A. S. Troshin. "New phenomena in quantum optics: photon antibunching, sub-Poisson photon statistics, and squeezed states." *Soviet Phys. Uspekhi* **30**, 851–874 (1987).

[93] P. M. Anisimov *et al.* "Quantum Metrology with Two-Mode Squeezed Vacuum: Parity Detection Beats the Heisenberg Limit." *Phys. Rev. Lett.* **104**, 103602 (2010).

[94] Z. Y. Ou, J.-K. Rhee, and L. J. Wang. "Photon bunching and multiphoton interference in parametric down-conversion." *Phys. Rev. A* **60**, 593–604 (1999).

[95] O. A. Ivanova *et al.* "Multiphoton correlations in parametric down-conversion and their measurement in the pulsed regime." *Quantum Electronics* **36**, 951–956 (2006).

[96] B. Blauensteiner *et al.* "Photon bunching in parametric down-conversion with continuous-wave excitation." *Phys. Rev. A* **79**, 63846 (2009).

[97] D. F. Walls. "Squeezed states of light." *Nature* **306**, 141 (1983).

[98] N. B. Grosse *et al.* "Measuring Photon Antibunching from Continuous Variable Sideband Squeezing." *Phys. Rev. Lett.* **98**, 153603 (2007).

[99] A. Prochazka *et al. Signal analysis and prediction.* Springer Science & Business Media (2013).

[100] J. Rice. "Bandwidth Choice for Nonparametric Regression." *Ann. Stat.* **12**, 1215–1230 (1984).

[101] F. Faa di Bruno. “Sullo sviluppo delle funzioni.” *Annali di scienze matematiche e fisiche* **6**, 479–480 (1855).

[102] L. Encinas and J. Masqué. “A short proof of the generalized Faà di bruno’s formula.” *Appl. Math. Lett.* **16**, 975–979 (2003).

[103] L. Carlitz. “Eulerian Numbers and Polynomials.” *Math. Magazine* **32**, 247–260 (1959).

[104] T. K. Petersen. *Eulerian numbers*. Ed. by T. K. Petersen. Springer New York (2015).