Asymptotic Quantum Statistical Estimation

(量子推定における漸近理論)

Koichi Yamagata

Osaka University

February 2013

Abstract

The present dissertation explores asymptotic quantum state estimation theory and its applications. The first half of the dissertation is devoted to investigating the ultimate limit of estimation precision in an asymptotic framework, assuming that any collective measurements are available. To this end, we extend the theory of weak local asymptotic normality, an essential ingredient in the classical asymptotic statistics, to a quantum regime. Meanwhile, it should be noticed that realizing collective measurements over a number of quantum systems is quite demanding, or even infeasible, in the current state of the art. In view of applications, therefore, it is also important to elaborate the estimation theory in which we make no use of quantum correlation, and the latter half of the dissertation is devoted to problems in this direction.

Let $\mathcal{H}$ be a finite dimensional Hilbert space that represents the physical system of interest. We say a pair of density operators ρ and σ on $\mathcal{H}$ are *mutually absolutely continuous*, $\rho \sim \sigma$ in symbols, if there exist a Hermitian operator $\mathcal{L}(\sigma|\rho)$ that satisfies

$$\sigma = \mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)} \rho \, \mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}.$$

We shall call such a Hermitian operator $\mathcal{L}(\sigma|\rho)$ a *quantum log-likelihood ratio*. The following theorem, one of the main results in the dissertation, generalizes the theory of local asymptotic normality (LAN) and Le Cam's third lemma in classical statistics [Theorem 2.9]:

Theorem. *Given a sequence $\mathcal{H}^{(n)}$ of finite dimensional Hilbert spaces, let*

$$\mathcal{S}^{(n)} = \left\{ \rho_\theta^{(n)} \; ; \; \theta \in \Theta \subset \mathbb{R}^d \right\}$$

be a quantum statistical model on $\mathcal{H}^{(n)}$, where $\rho_\theta^{(n)}$ is a parametric family of density operators and Θ is an open set. Let $X^{(n)} = (X_1^{(n)}, \ldots, X_r^{(n)})$ be a list of observables on $\mathcal{H}^{(n)}$. Fix a point $\theta_0 \in \Theta$. Assume $\mathcal{S}^{(n)}$ and $X^{(n)}$ satisfy the following conditions:

1. *for any $\theta \in \Theta$ and $n \in \mathbb{N}$, $\rho_\theta^{(n)}$ is mutually absolutely continuous to $\rho_{\theta_0}^{(n)}$,*
2. *there exist a list $\Delta^{(n)} = (\Delta_1^{(n)}, \ldots, \Delta_d^{(n)})$ of observables on each $\mathcal{H}^{(n)}$ that satisfies*

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{(n)} \right) \underset{q}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^* & J \end{pmatrix} \right),$$

   where Σ and J are Hermitian positive semidefinite matrices of size $r \times r$ and $d \times d$, respectively, with $\mathrm{Re}\, J > 0$, and τ is a complex matrix of size $r \times d$. (The arrow $\underset{q}{\rightsquigarrow}$ denotes a quantum extension of convergence in law, and $N(*,*)$ denotes a quantum Gaussian state.)
3. *quantum log-likelihood ratio $\mathcal{L}_h^{(n)} := \mathcal{L}\left(\rho_{\theta_0 + h/\sqrt{n}}^{(n)} \middle| \rho_{\theta_0}^{(n)} \right)$ is expanded in $h \in \mathbb{R}^d$ as*

$$\mathcal{L}_h^{(n)} = h^i \Delta_i^{(n)} - \frac{1}{2}(J_{ij} h^i h^j) I^{(n)} + o\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{(n)} \right).$$

   (The term $o(*,*)$ denotes an infinitesimal term defined in Section 2.2.)

It then follows that

$$\left(X^{(n)}, \rho_{\theta_0 + h/\sqrt{n}}^{(n)} \right) \underset{q}{\rightsquigarrow} N((\mathrm{Re}\, \tau)\, h, \, \Sigma)$$

for any $h \in \mathbb{R}^d$.

This theorem is successfully applied to the proof of asymptotic achievability of the Holevo bound for the local shift parameter $h \in \mathbb{R}^d$ [Theorem 2.12]:

Theorem. *Let* $\left\{\rho_\theta\,;\ \theta\in\Theta\subset\mathbb{R}^d\right\}$ *be a quantum statistical model on a finite dimensional Hilbert space* $\mathcal{H}$*, and fix a point* $\theta_0\in\Theta$*. Suppose that* $\rho_\theta\sim\rho_{\theta_0}$ *for all* $\theta\in\Theta$*, and that the quantum log-likelihood ratio* $\mathcal{L}_h:=\mathcal{L}\left(\rho_{\theta_0+h}|\rho_{\theta_0}\right)$ *is differentiable in* h *around* $h=0$ *and twice differentiable at* $h=0$*. For any countable dense subset* D *of* $\mathbb{R}^d$ *and any weight matrix* G *(*$d\times d$ *positive real matrix), there exist a sequence* $M^{(n)}$ *of estimators on the model* $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\,;\ h\in\mathbb{R}^d\right\}$ *that enjoys*

$$\lim_{n\to\infty}E_h^{(n)}[M^{(n)}]=h$$

and

$$\lim_{n\to\infty}\mathrm{Tr}\,GV_h^{(n)}[M^{(n)}]=C_{\theta_0}\left(\rho_\theta,G\right)$$

for every $h\in D$*. Here* $C_{\theta_0}\left(\rho_\theta,G\right)$ *is the Holevo bound at* θ_0*.*

This theorem clarifies the importance of the Holevo bound. However, the use of collective measurements, which is essential in achieving the bound, is beyond the reach of our current technology. In the latter half of the dissertation, therefore, we proceed to asymptotic quantum estimation schemes based on separable measurements. Among others, the efficiency of the quantum state tomography, a standard method widely used by experimental physicists, is scrutinized from the viewpoint of the quantum parameter estimation theory in which the trace of the weighted covariance matrix is to be minimized. The following theorem asserts that the quantum tomography is optimal if and only if a physically unnatural weight is adopted [Theorem 3.3]:

Theorem. *Let* $\mathcal{S}:=\{\tau_x\mid x=(x^1,x^2,x^3)\in\mathcal{X}\}$ *be the set of strictly positive density operators on* $\mathcal{H}=\mathbb{C}^2$ *parametrized by the Stokes parameters* $x\in\mathcal{X}:=\{x\in\mathbb{R}^3\mid(x^1)^2+(x^2)^2+(x^3)^2<1\}$ *as*

$$\tau_x:=\frac{1}{2}(I+x^1\sigma_1+x^2\sigma_2+x^3\sigma_3),$$

where $\sigma_1,\sigma_2,\sigma_3$ *are the Pauli matrices. Suppose we have an unknown quantum state* $\tau=\tau_x\in\mathcal{S}$*. Tomography is optimal if and only if the weight* H_x *is proportional to the following special one:*

$$H_x^{(T)}:=\begin{pmatrix}\frac{1}{1-(x^1)^2} & -\frac{(x^1)(x^2)}{(1-(x^1)^2)(1-(x^2)^2)} & -\frac{(x^3)(x^1)}{(1-(x^3)^2)(1-(x^1)^2)}\\ -\frac{(x^1)(x^2)}{(1-(x^1)^2)(1-(x^2)^2)} & \frac{1}{1-(x^2)^2} & -\frac{(x^2)(x^3)}{(1-(x^2)^2)(1-(x^3)^2)}\\ -\frac{(x^3)(x^1)}{(1-(x^3)^2)(1-(x^1)^2)} & -\frac{(x^2)(x^3)}{(1-(x^2)^2)(1-(x^3)^2)} & \frac{1}{1-(x^3)^2}\end{pmatrix}.$$

We also report the first experimental demonstration of an adaptive quantum state estimation (AQSE). The angle of linear polarization of single photons, or the phase parameter between the right and the left circularly polarization, is estimated using AQSE, and the strong consistency and asymptotic efficiency are experimentally verified.

List of Publications Related to the Dissertation

- K. Yamagata, "Efficiency of quantum state tomography for qubits," *International Journal of Quantum Information*, **9**, 1167 (2011).
- R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, "Experimental demonstration of adaptive quantum state estimation," *Physical Review Letters*, **109**, 130404 (2012).

Acknowledgments

The author thanks Prof. A. Fujiwara for precious discussions, comments, and careful much effort for me. The author thanks Prof. R. D. Gill for accepting me to study beside him at Leiden University, and providing me some valuable problems. The author also thanks Prof. H. Nagaoka, Prof. S. Takeuchi, Prof. T. Ogawa, and Dr. R. Okamoto for valuable discussions.

Contents

Chapter 1

Introduction

Quantum estimation theory was pioneered by Helstrom in late 1960s [20, 21]. He advocated an optical communication theory based on quantum physics and mathematical statistics, and studied a parameter estimation problem of optical signals. He derived a quantum counterpart of the logarithmic derivative called the symmetric logarithmic derivative (SLD) and a quantum extension of the Cramér-Rao inequality called the SLD Cramér-Rao inequality. In 1970s, Holevo, Yuen and Lax studied several theoretically important models [50, 24]. Especially, Yuen and Lax solved the simultaneous estimation problem of the complex amplitudes of coherent signals under Gaussian thermal noise. In that work, they introduced the right logarithmic derivative (RLD) and the RLD Cramér-Rao inequality to solve the two-dimensional parameter estimation problem for the first time. Today, their result are practically used as a quantum heterodyne measurement. In 1990s, Fujiwara and Matsumoto studied the estimation theory of pure state models intensively and revealed its relation with Berry-Uhlmann's geometrical phase [6, 7, 8]. After that, Matsumoto proved that the Holevo bound can be achievable for any pure state model [35] (see Section 2.C for a simple proof). Furthermore, qubit state estimation problem without invoking collective measurement was studied by Nagaoka (two-dimensional case [37]) and later by Hayashi (three-dimensional case [17]). Gill and Massar also treated the same problems independently from an entirely different point of view [13] (see Section 3.A for a simplified argument). In 2000s, some results about asymptotic theories of quantum state estimations appeared. Fujiwara proved the strong consistency and asymptotic efficiency of an adaptive quantum state estimation [9]. Hayashi and Matsumoto [19] showed the asymptotic achievability of the Holevo bound for a quantum statistical model on a Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$. Following their work, Guţă and Kahn [15, 31] developed a theory of (strong) quantum local asymptotic normality for a restricted class of models.

The purpose of the present dissertation is to explore a new asymptotic quantum state estimation theory and its applications. Let $\mathcal{S} = \left\{\rho_\theta \in \mathcal{S}(\mathcal{H})\,;\ \theta \in \Theta \subset \mathbb{R}^d\right\}$ be a quantum statistical model comprizing smoothly parametrized quantum state ρ_θ, where $S(\mathcal{H})$ is the set of quantum states (density operators) on a Hilbert space $\mathcal{H}$. Our purpose is to estimate the unknown parameter θ as efficient as possible. An estimator $\hat{M}$ for the parameter θ of this model, given by a positive-operator valued measure (POVM) on Θ, is called unbiased if

$$E_\theta[\hat{M}] = \theta \tag{1.1}$$

for all $\theta \in \Theta$, where $E_\theta[\cdot]$ denotes the expectation with respect to ρ_θ. An estimator $\hat{M}$ is called locally unbiased [24] at $\theta_0 \in \Theta$ if the condition (1.1) is satisfied around θ_0 up to the first order of the Taylor expansion. A locally unbiased estimator $\hat{M}$ at θ_0 satisfies the following inequality:

$$V_{\theta_0}[\hat{M}] \geq {J_{\theta_0}^{(S)}}^{-1}, \tag{1.2}$$

where $V_{\theta_0}[\cdot]$ denotes the covariance matrix with respect to ρ_{θ_0}, and $J_{\theta_0}^{(S)}$ is the quantum Fisher information matrix at θ_0 given by $J_{\theta_0}^{(S)} := \left[\mathrm{Re}\,\mathrm{Tr}\,\rho_{\theta_0} L_i L_j\right]_{1\leq i,j\leq d}$, where L_i is a ith SLD defined

by the self-adjoint operator satisfying the equation

$$\frac{\partial}{\partial\theta^i}\rho_\theta\,\Big|_{\theta=\theta_0} = \frac{1}{2}\left(L_i\rho_{\theta_0} + \rho_{\theta_0}L_i\right).$$

The optimal estimator achieving the SLD Cramér-Rao lower bound $J_{\theta_0}^{(S)^{-1}}$ always exits when θ is one-dimensional, while it is not achievable in general because the optimal measurements for each coordinate θ^i become incompatible. Put differently, the inequality (1.2) cannot be saturated in general because of the non-commutativity of the SLDs. To avoid this difficulty, we often adopt an alternative strategy to seek the estimator which minimizes $\mathrm{Tr}\, GV_{\theta_0}[\hat{M}]$, where G is a given $d\times d$ real positive definite matrix called a *weight* [24, 21]. The inequality

$$\mathrm{Tr}\, GV_{\theta_0}[\hat{M}] \geq C_{\theta_0}\left(\rho_\theta, G\right) \geq \mathrm{Tr}\, GJ_{\theta_0}^{(S)^{-1}} \tag{1.3}$$

is more informative than (1.2), where the quantity $C_{\theta_0}\left(\rho_\theta, G\right)$ is the Holevo bound [24] at θ_0 defined by

$$\begin{aligned} C_{\theta_0}\left(\rho_\theta, G\right) := \min_{V,B}\{\mathrm{Tr}\, GV\,;\ & V \text{ is a real matrix such that } V \geq Z(B),\ Z_{ij}(B) = \mathrm{Tr}\,\rho_{\theta_0}B_jB_i, \\ & B_1,\ldots,B_d \text{ are Hermitian operators on } \mathcal{H} \text{ such that } \mathrm{Re}\,\mathrm{Tr}\,\rho_{\theta_0}L_iB_j = \delta_{ij}\}. \end{aligned} \tag{1.4}$$

For any $n\in\mathbb{N}$, the Holevo bound for the nth i.i.d. extention model $\mathcal{S}^{(n)} := \left\{\rho_\theta^{\otimes n}\;\middle|\;\theta\in\Theta\subset\mathbb{R}^d\right\}$ is $\frac{1}{n}C_{\theta_0}\left(\rho_\theta, G\right)$, and

$$n\mathrm{Tr}\, GV_{\theta_0}^{(n)}[\hat{M}^{(n)}] \geq C_{\theta_0}\left(\rho_\theta, G\right), \tag{1.5}$$

where $V_{\theta_0}^{(n)}[\cdot]$ denotes the covariance matrix respect to $\rho_{\theta_0}^{\otimes n}$, and $\hat{M}^{(n)}$ is a collective estimator of $\rho_\theta^{\otimes n}$ which is locally unbiased. It is expected that the lower bound in (1.5) is achievable asymptotically because the sequence of models $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n};\ h\in\mathbb{R}^d\right\}$ with shrinking parameter h "converges" to a quantum Gaussian shift model in some sence. This property is called quantum local asymptotic normality (QLAN). Earlier research about QLAN is given by Guţă and Kahn [15, 31]. They proved that $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n};\ h\in\mathbb{R}^d\right\}$ and a quantum Gaussian shift model can be translated by quantum channels to each other asymptotically. Although their result is powerful, their QLAN has several drawbacks. It can be applicable only when a parametrization θ of $\mathcal{S}(\mathcal{H})$ takes a special form. Furthermore, it does not work if the reference state ρ_{θ_0} has a multiplicity of eigenvalues. Here we aim at developing QLAN theory applicable to any quantum statistical model satisfying a mild smoothness condition. Our approach is based on a new quantum extension of the log-likelihood ratio.

The optimal estimators appeared in QLAN theory are necessarily collective ones. It is, however, difficult to implement collective measurements over a number of constituent systems. We therefore confine our attention to separable estimators in the latter half of the dissertation. We prove that the quantum state tomography, one of the standard technique widely used by experimental physicists, is in general much less efficient than the optimal estimator. Note that the optimal estimator depends on the true value θ_0 of the parameter. In such a case, we necessarily invoke an adaptive estimation scheme [9]. We demonstrate that such an adaptive estimation scheme can be realized by a state-of-the-art technique in quantum optics.

The dissertation is organized as follows. In Chapter 2, we develop a theory of QLAN based on a new quantum log-likelihood ratio, and prove that the Holevo bound is asymptotically achievable. In Chapter 3, the efficiency of tomography is studied in depth, to conclude that the tomography is optimal if and only if a physically unnatural weight is adopted. We also give some numerical simulations to compare the asymptotic performance of the tomography and the optimal adaptive estimation schemes. In Chapter 4, experimental demonstration of an adaptive quantum state estimation (AQSE) is reported. The angle of linear polarization of single photons, or the phase parameter between the right and the left circularly polarization, is estimated using AQSE, and the strong consistency and asymptotic efficiency are experimentally verified.

Chapter 2

Quantum Local Asymptotic Normality Based on a New Quantum Likelihood Ratio

Abstract

We develop a theory of local asymptotic normality in a quantum regime based on a novel quantum analogue of the log-likelihood ratio. This formulation is applicable to any quantum statistical model satisfying a mild smoothness condition. As an application, we prove the asymptotic achievability of the Holevo bound for the local shift parameter.

2.1 Motivation

Given a (classical) statistical model $\mathcal{S} = \{p_\theta\,;\,\theta \in \Theta\}$ on a probability space $(\Omega, \mathcal{F}, \mu)$ indexed by a parameter θ that ranges over an open subset Θ of $\mathbb{R}^d$, let us introduce a local parameter $h := \sqrt{n}(\theta - \theta_0)$ around a fixed $\theta_0 \in \Theta$. If the parametrization $\theta \mapsto p_\theta$ is sufficiently smooth, it is known that the statistical properties of the model $\left\{p_{\theta_0+h/\sqrt{n}}^{\otimes n}\,;\,h \in \mathbb{R}^d\right\}$ is similar to that of the Gaussian shift model $\left\{N(h, J_{\theta_0}^{-1})\,;\,h \in \mathbb{R}^d\right\}$ for large n, where $p_\theta^{\otimes n}$ is the nth i.i.d. extension of p_θ, and J_{θ_0} is the Fisher information matrix of the model p_θ at θ_0. This property is called the local asymptotic normality of the model $\mathcal{S}$ [47].

More generally, a sequence $\left\{p_\theta^{(n)}\,;\,\theta \in \Theta \subset \mathbb{R}^d\right\}$ of statistical models on $(\Omega^{(n)}, \mathcal{F}^{(n)}, \mu^{(n)})$ is called *locally asymptotically normal* (LAN) at $\theta_0 \in \Theta$ if there exist a $d \times d$ positive matrix J and random vectors $\Delta^{(n)} = (\Delta_1^{(n)}, \ldots, \Delta_d^{(n)})$ such that $\Delta^{(n)} \overset{0}{\rightsquigarrow} N(0, J)$ and

$$\log \frac{p_{\theta_0+h/\sqrt{n}}^{(n)}}{p_{\theta_0}^{(n)}} = h^i \Delta_i^{(n)} - \frac{1}{2} h^i h^j J_{ij} + o_{p_{\theta_0}}(1)$$

for all $h \in \mathbb{R}^d$. Here the arrow $\overset{h}{\rightsquigarrow}$ stands for the convergence in distribution under $p_{\theta_0+h/\sqrt{n}}^{(n)}$, the remainder term $o_{p_{\theta_0}}(1)$ converges in probability to zero under $p_{\theta_0}^{(n)}$, and Einstein's summation convention is used. The above expansion is similar in form to the log-likelihood ratio of the Gaussian shift model:

$$\log \frac{dN(h, J^{-1})}{dN(0, J^{-1})}(X^1, \ldots, X^d) = h^i (X^j J_{ij}) - \frac{1}{2} h^i h^j J_{ij}.$$

This is the underlying mechanism behind the statistical similarities between models $\left\{p_{\theta_0+h/\sqrt{n}}^{(n)}\,;\,h \in \mathbb{R}^d\right\}$ and $\left\{N(h, J^{-1})\,;\,h \in \mathbb{R}^d\right\}$.

In order to put the similarities to practical use, one needs some mathematical devices. In general, a statistical theory comprises two parts. One is to prove the existence of a statistic that possesses a certain desired property (direct part), and the other is to prove the non-existence of a statistic that exceeds that property (converse part). In the problem of asymptotic efficiency, for example, the converse part, the impossibility to do asymptotically better than the best which can be done in the limit situation, is ensured by the following proposition, which is usually referred to as "Le Cam's third lemma" [47].

Proposition 2.1. *Suppose* $\left\{p_\theta^{(n)}\,;\,\theta\in\Theta\subset\mathbb{R}^d\right\}$ *is LAN at* $\theta_0\in\Theta$*, with* $\Delta^{(n)}$ *and* J *being as above, and let* $X^{(n)}=(X_1^{(n)},\dots,X_r^{(n)})$ *be a sequence of random vectors. If the joint distribution of* $X^{(n)}$ *and* $\Delta^{(n)}$ *converges to a Gaussian distribution, in that*

$$\begin{pmatrix}X^{(n)}\\ \Delta^{(n)}\end{pmatrix}\overset{0}{\rightsquigarrow}N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\Sigma & \tau\\ {}^t\tau & J\end{pmatrix}\right),$$

then $X^{(n)}\overset{h}{\rightsquigarrow}N(\tau h,\Sigma)$ *for all* $h\in\mathbb{R}^d$.

Now, it appears from this lemma that it already tells us something about the direct problem. In fact, by putting $X^{(n)j}:=\sum_{k=1}^d\left[J^{-1}\right]^{jk}\Delta_k^{(n)}$, we have

$$\begin{pmatrix}X^{(n)}\\ \Delta^{(n)}\end{pmatrix}\overset{0}{\rightsquigarrow}N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}J^{-1} & I\\ I & J\end{pmatrix}\right),$$

so that $X^{(n)}\overset{h}{\rightsquigarrow}N(h,J^{-1})$ follows from Proposition 2.1. This proves the existence of an asymptotically efficient estimator for h. In the real world however, we do not know θ_0 (obviously!). Thus the existence of an asymptotically optimal estimator for h does not translate into the existence of an asymptotically optimal estimator of θ. In fact, the usual way that Le Cam's third lemma is used in the subsequent analysis is in order to prove the so-called representation theorem, [47, Theorem 7.10]. This theorem can be used to tell us in several precise mathematical senses that no estimator can asymptotically do better than what can be achieved in the limiting Gaussian model.

For instance, Van der Vaart's version of the representation theorem leads to the asymptotic minimax theorem, telling us that the worst behaviour of an estimator as θ varies in a shrinking (1 over root n) neighbourhood of θ_0 cannot improve on what we expect from the limiting problem. This theorem applies to *all* possible estimators, but only discusses their *worst* behaviour in a neighbourhood of θ. Another option is to use the representation theorem to derive the convolution theorem, which tells us that *regular* estimators (estimators whose asymptotic behaviour in a small neighbourhood of θ is more or less stable as the parameter varies) have a limiting distribution which in a very strong sense is more disperse than the optimal limiting distribution which we expect from the limiting statistical problem.

This chapter addresses a quantum extension of LAN (abbreviated as QLAN). As in the classical statistics, one of the important subjects of QLAN is to show the existence of an estimator (direct part) that enjoys certain desired properties. Some earlier works of asymptotic quantum parameter estimation theory revealed the asymptotic achievability of the Holevo bound, a quantum extension of the Cramér-Rao type bound (cf., Appendices 2.A, 2.B). Using a group representation theoretical method, Hayashi and Matsumoto [19] showed that the Holevo bound for the quantum statistical model $\mathcal{S}(\mathbb{C}^2)=\{\rho_\theta\,;\,\theta\in\Theta\}$ comprising the totality of density operators on the Hilbert space $\mathcal{H}\simeq\mathbb{C}^2$ is asymptotically achievable at a given single point $\theta_0\in\Theta$. Following their work, Guţă and Kahn [15, 31] developed a theory of strong QLAN, and proved that the Holevo bound is asymptotically uniformly achievable around a given $\theta_0\in\Theta$ for the quantum statistical model $\mathcal{S}(\mathbb{C}^D)=\{\rho_\theta\,;\,\theta\in\Theta\}$ comprising the totality of density operators on the finite dimensional Hilbert space $\mathcal{H}\simeq\mathbb{C}^D$. They proved that $\left\{\rho_{h/\sqrt{n}}^{\otimes n}\,;\,h\in\mathbb{R}^d\right\}$ and a quantum Gaussian shift model can be translated by quantum channels to each other asymptotically. Although their result is powerful, their QLAN has several drawbacks. First of all, their method works only for i.i.d. extension of the totality $\mathcal{S}(\mathcal{H})$ of the quantum states on the

Hilbert space $\mathcal{H}$, and is not applicable to generic submodels of $\mathcal{S}(\mathcal{H})$. Moreover, it makes use of a special parametrization θ of $\mathcal{S}(\mathcal{H})$, in which the change of eigenvalues and eigenvectors are treated as essential. Furthermore, it does not work if the reference state ρ_{θ_0} has a multiplicity of eigenvalues. Hayashi and Matsumoto's formulation [19] also suffers from the same problems.

The purpose of the present chapter is to develop a theory of (weak) QLAN based on a new quantum extension of the log-likelihood ratio. This formulation is applicable to any quantum statistical model satisfying a mild smoothness condition, and is free from artificial setups such as the use of a special coordinate system and/or non-degeneracy of eigenvalues of the reference state at which QLAN works. We also prove asymptotic achievability of the Holevo bound for the local shift parameter h that belong to a dense subset of $\mathbb{R}^d$.

This chapter is organized as follows. The main results are summarized in Section 2.2. We first introduce a novel type of quantum log-likelihood ratio, and define a quantum extension of local asymptotic normality in a quite analogous way to the classical LAN. We then explore some basic properties of QLAN, including a sufficient condition for an i.i.d. model to be QLAN, and a quantum extension of Le Cam's third lemma. Proofs of those results are provided in Section 2.3. Section 2.4 is devoted to application of QLAN, including the asymptotic achievability of the Holevo bound and asymptotic estimation theory for some typical qubit models. For the reader's convenience, a brief account of quantum estimation theory are presented in appendices A-D. Those prerequisites are used throughout this chapter.

It is also important to notice the limits of this work, which means that there are many open problems left to study in the future. In the classical case, the theory of LAN builds, of course, on the rich theory of convergence in distribution, as studied in probability theory. In the quantum case, there still does not exist a full parallel theory. Some of the most useful lemmas in the classical theory simply are not true when translated in the quantum domain. For instance, in the classical case, we know that if the sequence of random variables X_n converges in distribution to a random variable X, and at the same time the sequence Y_n converges in probability to a constant c, then this implies joint convergence in distribution of (X_n, Y_n) to the pair (X, c). The obvious analogue of this in the quantum domain is simply untrue. In fact, there is not even a general theory of convergence in distribution at all: there is only a theory of convergence in distribution towards quantum Gaussian limits. Unfortunately, even in this special case the natural analogue of the just mentioned result simply fails to be true.

Because of these obstructions we are not at present able to follow the standard route from Le Cam's third lemma to the representation theorem, and from there to asymptotic minimax or convolution theorems.

However we believe that this chapter presents some notable steps in this direction. Moreover, just as with Le Cam's third lemma, one is able to use the lemma to construct what can be conjectured to be asymptotically optimal measurement and estimation schemes. We make some more remarks on these possibilities later in this chapter.

2.2 Main results

Definition 2.2 (Quantum log-likelihood ratio)**.** *We say a pair of density operators ρ and σ on a finite dimensional Hilbert space $\mathcal{H}$ are mutually absolutely continuous, $\rho \sim \sigma$ in symbols, if there exist a Hermitian operator $\mathcal{L}$ that satisfies*

$$\sigma = \mathrm{e}^{\frac{1}{2}\mathcal{L}} \rho \, \mathrm{e}^{\frac{1}{2}\mathcal{L}}.$$

We shall call such a Hermitian operator $\mathcal{L}$ a quantum log-likelihood ratio. When the reference states ρ and σ need to be specified, $\mathcal{L}$ shall be denoted by $\mathcal{L}(\sigma|\rho)$, so that

$$\sigma = \mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)} \rho \, \mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}.$$

We use the convention that $\mathcal{L}(\rho|\rho) = 0$.

Example 2.3. *Faithful states are always mutually absolutely continuous. In fact, given $\rho > 0$*

and $\sigma > 0$, they are related as $\sigma = \mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}\rho\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}$, where

$$\mathcal{L}(\sigma|\rho) = 2\log\left(\sqrt{\rho^{-1}}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\sqrt{\rho^{-1}}\right) = 2\log\left(\sqrt{\sigma}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right)^{-1}\sqrt{\sigma}\right).$$

Note that $\operatorname{Tr}\rho\,\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}$ is identical to the fidelity between ρ and σ, and $\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}$ is nothing but the operator geometric mean $\sigma\#\rho^{-1}$, where $A\#B := A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2}$ for positive operators A, B.

Example 2.4. *Pure states $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\xi\rangle\langle\xi|$ are mutually absolutely continuous if and only if $\langle\xi|\psi\rangle \neq 0$.*

Proof. Suppose first that $\rho \sim \sigma$. Then

$$|\langle\xi|\psi\rangle|^2 = \operatorname{Tr}\rho\sigma = \operatorname{Tr}\,|\psi\rangle\langle\psi|\,\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}\,|\psi\rangle\langle\psi|\,\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)} = \left|\langle\psi|\,\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}\,|\psi\rangle\right|^2 > 0.$$

Suppose next that $\langle\xi|\psi\rangle \neq 0$. Then

$$R := I + \frac{1}{|\langle\xi|\psi\rangle|}|\xi\rangle\langle\xi| - |\psi\rangle\langle\psi|$$

is positive definite, and $\mathcal{L}(\sigma|\rho) := 2\log R$ satisfies

$$\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}\,|\psi\rangle = R\,|\psi\rangle = \frac{\langle\xi|\psi\rangle}{|\langle\xi|\psi\rangle|}\,|\xi\rangle,$$

showing that $\rho \sim \sigma$. Note that $\operatorname{Tr}\rho\,\mathrm{e}^{\frac{1}{2}\mathcal{L}(\sigma|\rho)}$ is the fidelity again. □

Given a $d \times d$ real skew-symmetric matrix S, let $\mathrm{CCR}(S)$ be the CCR algebra defined by

$$\mathrm{e}^{\sqrt{-1}X_i}\mathrm{e}^{\sqrt{-1}X_j} = \mathrm{e}^{\sqrt{-1}S_{ij}}\mathrm{e}^{\sqrt{-1}(X_i+X_j)} \quad (1 \le i, j \le d),$$

(see [34, 44, 28, 24]). We call $X = (X_1, \ldots, X_d)$ the basic canonical observables of $\mathrm{CCR}(S)$. A state ϕ on $\mathrm{CCR}(S)$ is characterized by the *characteristic function* $\mathcal{F}_\xi\{\phi\} := \phi(\mathrm{e}^{\sqrt{-1}\xi^i X_i})$, where $\xi = (\xi^i)_{i=1}^d \in \mathbb{R}^d$ and Einstein's summation convention is used. A state ϕ on $\mathrm{CCR}(S)$ is called a *quantum Gaussian state*, denoted by $\phi \sim N(h, J)$, if the characteristic function takes the form

$$\mathcal{F}_\xi\{\phi\} = \mathrm{e}^{\sqrt{-1}\xi^i h_i - \frac{1}{2}\xi^i\xi^j V_{ij}},$$

where $h = (h_i)_{i=1}^d \in \mathbb{R}^d$ and $V = (V_{ij})$ is a real symmetric matrix such that the Hermitian matrix $J := V + \sqrt{-1}S$ is positive semidefinite. When the canonical observables X need to be specified, we also use the notation $(X, \phi) \sim N(h, J)$.

We will discuss relationships between a quantum Gaussian state ϕ on a CCR and a state on another algebra. In such a case, we need to use the *quasi-characteristic function*

$$\phi\left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i}\right) = \exp\left(\sum_{t=1}^r\left(\sqrt{-1}\xi_t^i h_i - \frac{1}{2}\xi_t^i\xi_t^j J_{ji}\right) - \sum_{t=1}^r\sum_{s=t+1}^r \xi_t^i\xi_s^j J_{ji}\right), \tag{2.1}$$

of a quantum Gaussian state, where $(X, \phi) \sim N(h, J)$ and $\{\xi_t\}_{t=1}^r$ is a finite subset of $\mathbb{C}^d$ [28].

Given a sequence $\mathcal{H}^{(n)}$, $n \in \mathbb{N}$, of finite dimensional Hilbert spaces, let $X^{(n)} = (X_1^{(n)}, \ldots, X_d^{(n)})$ and $\rho^{(n)}$ be a list of observables and a density operator on each $\mathcal{H}^{(n)}$. We say the sequence $\left(X^{(n)}, \rho^{(n)}\right)$ *converges in law to a quantum Gaussian state* $N(h, J)$, denoted as $(X^{(n)}, \rho^{(n)}) \underset{q}{\rightsquigarrow} N(h, J)$, if

$$\lim_{n\to\infty} \operatorname{Tr}\rho^{(n)}\left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i^{(n)}}\right) = \phi\left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i}\right)$$

for any finite subset $\{\xi_t\}_{t=1}^r$ of $\mathbb{C}^d$, where $(X, \phi) \sim N(h, J)$. Here we do not intend to introduce the notion of "quantum convergence in law" in general. We use this notion only for quantum Gaussian states in the sense of convergence of quasi-characteristic function.

The following is a version of the quantum central limit theorem (see [28], for example).

Proposition 2.5 (Quantum central limit theorem)**.** *Let A_i $(1 \le i \le d)$ and ρ be observables and a state on a finite dimensional Hilbert space $\mathcal{H}$ such that $\operatorname{Tr}\rho A_i = 0$, and let*

$$X_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes A_i \otimes I^{\otimes(n-k)}.$$

Then $(X^{(n)}, \rho^{\otimes n}) \underset{q}{\rightsquigarrow} N(0, J)$, where J is the Hermitian matrix whose (i,j)th entry is given by $J_{ij} = \operatorname{Tr}\rho A_j A_i$.

For later convenience, we introduce the notion of an "infinitesimal" object relative to the convergence $(X^{(n)}, \rho^{(n)}) \underset{q}{\rightsquigarrow} N(0, J)$ as follows. Given a list $X^{(n)} = (X_1^{(n)}, \dots, X_d^{(n)})$ of observables and a state $\rho^{(n)}$ on each $\mathcal{H}^{(n)}$ that satisfy $(X^{(n)}, \rho^{(n)}) \underset{q}{\rightsquigarrow} N(0, J) \sim (X, \phi)$, we say a sequence $R^{(n)}$ of observables, each being defined on $\mathcal{H}^{(n)}$, is *infinitesimal relative to the convergence* $(X^{(n)}, \rho^{(n)}) \underset{q}{\rightsquigarrow} N(0, J)$ if it satisfies

$$\lim_{n\to\infty} \operatorname{Tr}\rho^{(n)} \left(\prod_{t=1}^{r} \mathrm{e}^{\sqrt{-1}\left(\xi_t^i X_i^{(n)} + \eta_t R^{(n)}\right)} \right) = \phi\left(\prod_{t=1}^{r} \mathrm{e}^{\sqrt{-1}\xi_t^i X_i} \right) \tag{2.2}$$

for any finite subset of $\{\xi_t\}_{t=1}^r$ of $\mathbb{C}^d$ and any finite subset $\{\eta_t\}_{t=1}^r$ of $\mathbb{C}$. This is equivalent to saying that

$$\left(\begin{pmatrix} X^{(n)} \\ R^{(n)} \end{pmatrix}, \rho^{(n)} \right) \underset{q}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} J & 0 \\ 0 & 0 \end{pmatrix} \right),$$

and is much stronger a requirement than

$$(R^{(n)}, \rho^{(n)}) \underset{q}{\rightsquigarrow} N(0, 0).$$

An infinitesimal object $R^{(n)}$ relative to $(X^{(n)}, \rho^{(n)}) \underset{q}{\rightsquigarrow} N(0, J)$ will be denoted as $o(X^{(n)}, \rho^{(n)})$.

The following is in essence a simple extension of Proposition 2.5, but will turn out to be useful in applications.

Lemma 2.6. *In addition to assumptions of Proposition 2.5, let $P(n)$, $n \in \mathbb{N}$, be a sequence of observables on $\mathcal{H}$, and let*

$$R^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes P(n) \otimes I^{\otimes(n-k)}.$$

If $\lim_{n\to\infty} P(n) = 0$ and $\lim_{n\to\infty} \sqrt{n} \operatorname{Tr}\rho P(n) = 0$, then $R^{(n)} = o(X^{(n)}, \rho^{\otimes n})$.

We are now ready to extend the notion of local asymptotic normality to a quantum regime.

Definition 2.7 (QLAN)**.** *Given a sequence $\mathcal{H}^{(n)}$ of finite dimensional Hilbert spaces, let $\mathcal{S}^{(n)} = \left\{ \rho_\theta^{(n)} \; ; \; \theta \in \Theta \subset \mathbb{R}^d \right\}$ be a quantum statistical model on $\mathcal{H}^{(n)}$, where $\rho_\theta^{(n)}$ is a parametric family of density operators and Θ is an open set. We say $\mathcal{S}^{(n)}$ is quantum locally asymptotically normal (QLAN) at $\theta_0 \in \Theta$ if the following conditions are satisfied:*

1. for any $\theta \in \Theta$ and $n \in \mathbb{N}$, $\rho_\theta^{(n)}$ is mutually absolutely continuous to $\rho_{\theta_0}^{(n)}$,
2. there exist a list $\Delta^{(n)} = (\Delta_1^{(n)}, \dots, \Delta_d^{(n)})$ of observables on each $\mathcal{H}^{(n)}$ that satisfies

$$\left(\Delta^{(n)}, \rho_{\theta_0}^{(n)} \right) \underset{q}{\rightsquigarrow} N(0, J),$$

where J is a $d \times d$ Hermitian positive semidefinite matrix with $\operatorname{Re} J > 0$,

3. quantum log-likelihood ratio $\mathcal{L}_h^{(n)} := \mathcal{L}\left(\rho_{\theta_0+h/\sqrt{n}}^{(n)} \middle| \rho_{\theta_0}^{(n)}\right)$ is expanded in $h \in \mathbb{R}^d$ as

$$\mathcal{L}_h^{(n)} = h^i \Delta_i^{(n)} - \frac{1}{2}(J_{ij} h^i h^j) I^{(n)} + o(\Delta^{(n)}, \rho_{\theta_0}^{(n)}), \tag{2.3}$$

where $I^{(n)}$ is the identity operator on $\mathcal{H}^{(n)}$.

It is also possible to extend Le Cam's third lemma (Proposition 2.1) to a quantum regime. To this end, however, we need a device to handle the infinitesimal residual term in (2.3) in a more elaborate way.

Definition 2.8. *Let $\mathcal{S}^{(n)} = \left\{\rho_\theta^{(n)};\ \theta \in \Theta \subset \mathbb{R}^d\right\}$ be as in Definition 2.7, and let $X^{(n)} = (X_1^{(n)}, \ldots, X_r^{(n)})$ be a list of observables on $\mathcal{H}^{(n)}$. We say the pair $(\mathcal{S}^{(n)}, X^{(n)})$ is* jointly QLAN *at $\theta_0 \in \Theta$ if the following conditions are satisfied:*

1. for any $\theta \in \Theta$ and $n \in \mathbb{N}$, $\rho_\theta^{(n)}$ is mutually absolutely continuous to $\rho_{\theta_0}^{(n)}$,
2. there exist a list $\Delta^{(n)} = (\Delta_1^{(n)}, \ldots, \Delta_d^{(n)})$ of observables on each $\mathcal{H}^{(n)}$ that satisfies

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{(n)}\right) \underset{q}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^* & J \end{pmatrix}\right), \tag{2.4}$$

where Σ and J are Hermitian positive semidefinite matrices of size $r \times r$ and $d \times d$, respectively, with $\operatorname{Re} J > 0$, and τ is a complex matrix of size $r \times d$.

3. quantum log-likelihood ratio $\mathcal{L}_h^{(n)} := \mathcal{L}\left(\rho_{\theta_0+h/\sqrt{n}}^{(n)} \middle| \rho_{\theta_0}^{(n)}\right)$ is expanded in $h \in \mathbb{R}^d$ as

$$\mathcal{L}_h^{(n)} = h^i \Delta_i^{(n)} - \frac{1}{2}(J_{ij} h^i h^j) I^{(n)} + o\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{(n)}\right). \tag{2.5}$$

With Definition 2.8, we can state a quantum extension of Le Cam's third lemma as follows.

Theorem 2.9. *Let $\mathcal{S}^{(n)}$ and $X^{(n)}$ be as in Definition 2.8. If $(\rho_\theta^{(n)}, X^{(n)})$ is jointly QLAN at $\theta_0 \in \Theta$, then*

$$\left(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{(n)}\right) \underset{q}{\rightsquigarrow} N((\operatorname{Re}\tau)\, h,\ \Sigma)$$

for any $h \in \mathbb{R}^d$.

In applications, we often handle i.i.d. extensions. In classical statistics, a sequence of i.i.d. extensions of a model is LAN if the log-likelihood ratio is twice differentiable [47]. Quite analogously, we can prove that a sequence of i.i.d. extensions of a quantum statistical model is QLAN if the quantum log-likelihood ratio is twice differentiable.

Theorem 2.10. *Let $\left\{\rho_\theta;\ \theta \in \Theta \subset \mathbb{R}^d\right\}$ be a quantum statistical model on a finite dimensional Hilbert space $\mathcal{H}$ satisfying $\rho_\theta \sim \rho_{\theta_0}$ for all $\theta \in \Theta$, where $\theta_0 \in \Theta$ is an arbitrarily fixed point. If $\mathcal{L}_h := \mathcal{L}(\rho_{\theta_0+h} | \rho_{\theta_0})$ is differentiable around $h = 0$ and twice differentiable at $h = 0$, then $\left\{\rho_\theta^{\otimes n};\ \theta \in \Theta \subset \mathbb{R}^d\right\}$ is QLAN at θ_0: that is, $\rho_\theta^{\otimes n} \sim \rho_{\theta_0}^{\otimes n}$, and*

$$\Delta_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes L_i \otimes I^{\otimes(n-k)}$$

and $J_{ij} := \operatorname{Tr} \rho_{\theta_0} L_j L_i$, with L_i being the ith symmetric logarithmic derivative at $\theta_0 \in \Theta$, satisfy conditions (ii) (iii) in Definition 2.7.

By combining Theorem 2.10 with Theorem 2.9, we obtain the following.

Corollary 2.11. *Let $\{\rho_\theta\,;\ \theta\in\Theta\subset\mathbb{R}^d\}$ be a quantum statistical model on $\mathcal{H}$ satisfying $\rho_\theta\sim\rho_{\theta_0}$ for all $\theta\in\Theta$, where $\theta_0\in\Theta$ is an arbitrarily fixed point. Further, let $\{B_i\}_{1\le i\le r}$ be observables on $\mathcal{H}$ satisfying $\operatorname{Tr}\rho_{\theta_0}B_i=0$ for $i=1,\dots,r$. If $\mathcal{L}_h:=\mathcal{L}\left(\rho_{\theta_0+h}|\rho_{\theta_0}\right)$ is differentiable around $h=0$ and twice differentiable at $h=0$, then the pair $\left(\left\{\rho_\theta^{\otimes n}\right\},\,X^{(n)}\right)$ of i.i.d. extension model $\left\{\rho_\theta^{\otimes n}\right\}$ and the list $X^{(n)}=\{X_i^{(n)}\}_{1\le i\le r}$ of observables defined by*

$$X_i^{(n)}:=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}I^{\otimes(k-1)}\otimes B_i\otimes I^{\otimes(n-k)}$$

is jointly QLAN at θ_0, and

$$\left(X^{(n)},\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\right)\underset{q}{\rightsquigarrow}N((\operatorname{Re}\tau)\,h,\Sigma)$$

for any $h\in\mathbb{R}^d$, where Σ is the $r\times r$ positive semidefinite matrix defined by $\Sigma_{ij}=\operatorname{Tr}\rho_{\theta_0}B_jB_i$ and τ is the $r\times d$ matrix defined by $\tau_{ij}=\operatorname{Tr}\rho_{\theta_0}L_jB_i$ with L_i being the ith SLD at θ_0.

As in the classical case, Corollary 2.11 prompts us to expect that any estimator for a quantum Gaussian shift model $\left\{N((\operatorname{Re}\tau)\,h,\Sigma)\,;\ h\in\mathbb{R}^d\right\}$ could be realized asymptotically on $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\,;\ h\in\mathbb{R}^d\right\}$. This program will be partly demonstrated in Section 2.4 in the form of achievability of the Holevo bound.

2.3 Proofs of main theorems

2.3.1 Proof of Lemma 2.6

We shall prove (2.2) for $\{\xi_t\}_{t=1}^r\subset\mathbb{C}^d$ and $\{\eta_t\}_{t=1}^r\subset\mathbb{C}$.

$$\begin{aligned}
&\operatorname{Tr}\rho^{\otimes n}\left(\prod_{t=1}^{r}\mathrm{e}^{\sqrt{-1}\left(\xi_t^iX_i^{(n)}+\eta_tR^{(n)}\right)}\right)\\
&\quad=\operatorname{Tr}\rho^{\otimes n}\left[\prod_{t=1}^{r}\exp\left\{\frac{\sqrt{-1}}{\sqrt{n}}\sum_{k=1}^{n}I^{\otimes(k-1)}\otimes\left(\xi_t^iA_i+\eta_tP(n)\right)\otimes I^{\otimes(n-k)}\right\}\right]\\
&\quad=\operatorname{Tr}\rho^{\otimes n}\left[\prod_{t=1}^{r}\left\{\exp\left(\frac{\sqrt{-1}}{\sqrt{n}}\left(\xi_t^iA_i+\eta_tP(n)\right)\right)\right\}^{\otimes n}\right]\\
&\quad=\operatorname{Tr}\rho^{\otimes n}\left[\left\{\prod_{t=1}^{r}\exp\left(\frac{\sqrt{-1}}{\sqrt{n}}\left(\xi_t^iA_i+\eta_tP(n)\right)\right)\right\}^{\otimes n}\right]\\
&\quad=\left[\operatorname{Tr}\rho\left\{\prod_{t=1}^{r}\exp\left(\frac{\sqrt{-1}}{\sqrt{n}}\left(\xi_t^iA_i+\eta_tP(n)\right)\right)\right\}\right]^n\\
&\quad=\left[\operatorname{Tr}\rho\left\{\sum_{k_1,\dots,k_r\in\mathbb{Z}_+}\left(\frac{\sqrt{-1}}{\sqrt{n}}\right)^{k_1+\cdots+k_r}\prod_{t=1}^{r}\frac{1}{k_t!}\left(\xi_t^iA_i+\eta_tP(n)\right)^{k_t}\right\}\right]^n,
\end{aligned}$$

where $\mathbb{Z}_+=\{0,1,2,\dots\}$. The terms corresponding to $k_1+\cdots+k_r=1$ in the summand are

$$\operatorname{Tr}\rho\left(\frac{\sqrt{-1}}{\sqrt{n}}\sum_{t=1}^{r}\left(\xi_t^iA_i+\eta_tP(n)\right)\right)=\left(\sum_{t=1}^{r}\eta_t\right)\frac{\sqrt{-1}}{\sqrt{n}}\operatorname{Tr}\rho P(n)=o\left(\frac{1}{n}\right)$$

because $\mathrm{Tr}\,\rho A_i = 0$ and $\mathrm{Tr}\,\rho P(n) = o(\frac{1}{\sqrt{n}})$. The terms corresponding to $k_1 + \cdots + k_r = 2$ are

$$\begin{aligned}
&-\frac{1}{n}\mathrm{Tr}\,\rho\left\{\sum_{k_1+\cdots+k_r=2}\left(\prod_{t=1}^{r}\frac{1}{k_t!}\left(\xi_t^i A_i+\eta_t P(n)\right)^{k_t}\right)\right\}\\
&\quad=-\frac{1}{2n}\sum_{t=1}^{r}\mathrm{Tr}\,\rho\left(\xi_t^i A_i+\eta_t P(n)\right)^2-\frac{1}{n}\sum_{t=1}^{r}\sum_{s=t+1}^{r}\mathrm{Tr}\,\rho\left(\xi_t^i A_i+\eta_t P(n)\right)\left(\xi_s^j A_j+\eta_s P(n)\right)\\
&\quad=-\frac{1}{2n}\sum_{t=1}^{r}\xi_t^i\xi_t^j\mathrm{Tr}\,\rho A_iA_j-\frac{1}{n}\sum_{t=1}^{r}\sum_{s=t+1}^{r}\xi_t^i\xi_s^j\mathrm{Tr}\,\rho A_iA_j+o\left(\frac{1}{n}\right)\\
&\quad=-\frac{1}{2n}\sum_{t=1}^{r}\xi_t^i\xi_t^j J_{ji}-\frac{1}{n}\sum_{t=1}^{r}\sum_{s=t+1}^{r}\xi_t^i\xi_s^j J_{ji}+o\left(\frac{1}{n}\right).
\end{aligned}$$

In the third line, we used the fact that $P(n) = o(1)$. Let us denote the terms corresponding to $k_1 + \cdots + k_r \geq 3$ by

$$r_n := \mathrm{Tr}\,\rho\left\{\sum_{k_1+\cdots+k_r\geq 3}\left(\frac{\sqrt{-1}}{\sqrt{n}}\right)^{(k_1+\cdots+k_r)}\prod_{t=1}^{r}\frac{1}{k_t!}\left(\xi_t^i A_i+\eta_t P(n)\right)^{k_t}\right\}.$$

Then

$$\begin{aligned}
|r_n| &\leq \sum_{k_1+\cdots+k_r\geq 3}\left\|\left(\frac{1}{\sqrt{n}}\right)^{(k_1+\cdots+k_r)}\prod_{t=1}^{r}\frac{1}{k_t!}\left(\xi_t^i A_i+\eta_t P(n)\right)^{k_t}\right\|\\
&\leq \frac{1}{n\sqrt{n}}\sum_{k_1+\cdots+k_r\geq 3}\left\|\prod_{t=1}^{r}\frac{1}{k_t!}\left(\xi_t^i A_i+\eta_t P(n)\right)^{k_t}\right\|\\
&\leq \frac{1}{n\sqrt{n}}\sum_{k_1+\cdots+k_r\geq 3}\prod_{t=1}^{r}\frac{1}{k_t!}\left\|\xi_t^i A_i+\eta_t P(n)\right\|^{k_t}\\
&\leq \frac{1}{n\sqrt{n}}\sum_{k_1,\ldots,k_r\in\mathbb{Z}_+}\prod_{t=1}^{r}\frac{1}{k_t!}\left\|\xi_t^i A_i+\eta_t P(n)\right\|^{k_t}\\
&= \frac{1}{n\sqrt{n}}\prod_{t=1}^{r}\exp\left\|\xi_t^i A_i+\eta_t P(n)\right\|\\
&\leq \frac{1}{n\sqrt{n}}\exp\left(\sum_{t=1}^{r}\left(\left\|\xi_t^i A_i\right\|+\left\|\eta_t P(n)\right\|\right)\right).
\end{aligned}$$

Since $\lim_{n\to\infty} P(n) = 0$, the operators $P(n)$ are uniformly bounded. As a consequence, $\lim_{n\to\infty} n\,|r_n| = 0$, so that $r_n = o\left(\frac{1}{n}\right)$. Thus we conclude that

$$\begin{aligned}
\lim_{n\to\infty}\mathrm{Tr}\,\rho^{\otimes n}\left(\prod_{t=1}^{r}\mathrm{e}^{\sqrt{-1}\left(\xi_t^i X_i^{(n)}+\eta_t R^{(n)}\right)}\right) &= \lim_{n\to\infty}\left(1-\frac{1}{2n}\sum_{t=1}^{r}\xi_t^i\xi_t^j J_{ji}-\frac{1}{n}\sum_{t=1}^{r}\sum_{s=t+1}^{r}\xi_t^i\xi_s^j J_{ji}+o\left(\frac{1}{n}\right)\right)^n\\
&= \exp\left(-\frac{1}{2}\sum_{t=1}^{r}\xi_t^i\xi_t^j J_{ji}-\sum_{t=1}^{r}\sum_{s=t+1}^{r}\xi_t^i\xi_s^j J_{ji}\right)\\
&= \phi\left(\prod_{t=1}^{r}\mathrm{e}^{\sqrt{-1}\xi_t^i X_i}\right).
\end{aligned}$$

The last equation is due to (2.1) with $h = 0$.

2.3.2 Proof of Theorem 2.9

Let $X_1, \dots, X_r, \Delta_1, \dots, \Delta_d$ be the basic canonical observables of CCR $\left(\mathrm{Im}\begin{pmatrix}\Sigma & \tau\\ \tau^* & J\end{pmatrix}\right)$, and $\tilde{\phi}$ the quantum Gaussian state $N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Sigma & \tau\\ \tau^* & J\end{pmatrix}\right)$ on that CCR. Assumption (2.5) guarantees that the quantities

$$R_h^{(n)} := \mathcal{L}_h^{(n)} - \left\{ h^i \Delta_i^{(n)} - \frac{1}{2} J_{ij} h^i h^j I^{(n)} \right\}$$

enjoy $R_h^{(n)} = o\left(\begin{pmatrix}X^{(n)}\\ \Delta^{(n)}\end{pmatrix}, \rho_{\theta_0}^{(n)}\right)$ for each $h \in \mathbb{R}^d$. Consequently, for a finite subset $\{\xi_t\}_{t=1}^r$ of $\mathbb{C}^d$,

$$\begin{aligned}
&\mathrm{Tr}\, \rho_{\theta_0+h/\sqrt{n}}^{(n)} \left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i^{(n)}} \right) \\
&= \mathrm{Tr}\, \left(\mathrm{e}^{\frac{1}{2}\mathcal{L}_h^{(n)}} \rho_{\theta_0}^{(n)} \mathrm{e}^{\frac{1}{2}\mathcal{L}_h^{(n)}} \right) \left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i^{(n)}} \right) \\
&= \mathrm{e}^{-\frac{1}{2}h^i h^j J_{ij}} \mathrm{Tr}\, \rho_{\theta_0}^{(n)} \mathrm{e}^{\frac{1}{2}\left(h^i \Delta_i^{(n)} + R_h^{(n)}\right)} \left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i^{(n)}} \right) \mathrm{e}^{\frac{1}{2}\left(h^i \Delta_i^{(n)} + R_h^{(n)}\right)} \\
&= \mathrm{e}^{-\frac{1}{2}h^i h^j J_{ij}} \mathrm{Tr}\, \rho_{\theta_0}^{(n)} \left(\mathrm{e}^{-\sqrt{-1}\left(\frac{\sqrt{-1}}{2} h^i \Delta_i^{(n)} + \frac{\sqrt{-1}}{2} R_h^{(n)}\right)} \right) \left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i^{(n)}} \right) \left(\mathrm{e}^{-\sqrt{-1}\left(\frac{\sqrt{-1}}{2} h^i \Delta_i^{(n)} + \frac{\sqrt{-1}}{2} R_h^{(n)}\right)} \right).
\end{aligned}$$

Since $R_h^{(n)}$ is infinitesimal relative to the convergence (2.4), we see from (2.2) that

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{Tr}\, \rho_{\theta_0+h/\sqrt{n}}^{(n)} \left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i^{(n)}} \right) \\
&= \mathrm{e}^{-\frac{1}{2}h^i h^j J_{ij}} \tilde{\phi} \left(\mathrm{e}^{-\sqrt{-1}\frac{\sqrt{-1}}{2} h^i \Delta_i} \left(\prod_{t=1}^r \mathrm{e}^{\sqrt{-1}\xi_t^i X_i} \right) \mathrm{e}^{-\sqrt{-1}\frac{\sqrt{-1}}{2} h^i \Delta_i} \right) \\
&= \mathrm{e}^{-\frac{1}{2}h^i h^j J_{ij}} \exp\left(-\frac{1}{2} \sum_{t=0}^{r+1} \tilde{\xi}_t^i \tilde{\xi}_t^j \tilde{\Sigma}_{ji} - \sum_{t=0}^{r+1} \sum_{s=t+1}^{r+1} \tilde{\xi}_t^i \tilde{\xi}_s^j \tilde{\Sigma}_{ji} \right) \\
&= \mathrm{e}^{-\frac{1}{2}h^i h^j J_{ij}} \exp\left(-\frac{1}{2} \left\{ -\frac{1}{4} h^j h^j J_{ji} + \sum_{t=1}^r \xi_t^i \xi_t^j \Sigma_{ji} - \frac{1}{4} h^j h^j J_{ji} \right\} \right) \\
&\qquad \times \exp\left(\frac{\sqrt{-1}}{2} \sum_{t=1}^r h^i \xi_t^j \tau_{ji} + \frac{\sqrt{-1}}{2} \sum_{t=1}^r \xi_t^i h^j \tau_{ji} + \frac{1}{4} h^i h^j J_{ji} - \sum_{t=1}^r \sum_{s=t+1}^r \xi_t^i \xi_s^j \Sigma_{ji} \right) \\
&= \exp\left(\sum_{t=1}^r \left(\sqrt{-1} \xi_t^i h^j \left(\mathrm{Re}\, \tau\right)_{ij} - \frac{1}{2} \xi_t^i \xi_t^j \Sigma_{ji} \right) - \sum_{t=1}^r \sum_{s=t+1}^r \xi_t^i \xi_s^j \Sigma_{ji} \right),
\end{aligned}$$

where $\tilde{\Sigma} := \begin{pmatrix}\Sigma & \tau\\ \tau^* & J\end{pmatrix}$ and $(\tilde{\xi}_0, \tilde{\xi}_1, \dots, \tilde{\xi}_r, \tilde{\xi}_{r+1}) := (-\frac{\sqrt{-1}}{2} h, \xi_1, \dots, \xi_r, -\frac{\sqrt{-1}}{2} h)$, and (2.1) was used at the second equation. This is the quasi-characteristic function of $N((\mathrm{Re}\, \tau)\, h, \Sigma)$.

2.3.3 Proof of Theorem 2.10

Since

$$\begin{aligned}
\rho_\theta^{\otimes n} &= \left(\mathrm{e}^{\frac{1}{2}\mathcal{L}\left(\rho_\theta | \rho_{\theta_0}\right)} \rho_{\theta_0} \mathrm{e}^{\frac{1}{2}\mathcal{L}\left(\rho_\theta | \rho_{\theta_0}\right)} \right)^{\otimes n} \\
&= \left(\mathrm{e}^{\frac{1}{2}\sum_{k=1}^n I^{\otimes(k-1)} \otimes \mathcal{L}\left(\rho_\theta | \rho_{\theta_0}\right) \otimes I^{\otimes(n-k)}} \right) \rho_{\theta_0}^{\otimes n} \left(\mathrm{e}^{\frac{1}{2}\sum_{k=1}^n I^{\otimes(k-1)} \otimes \mathcal{L}\left(\rho_\theta | \rho_{\theta_0}\right) \otimes I^{\otimes(n-k)}} \right),
\end{aligned}$$

we see that

$$\mathcal{L}\left(\rho_\theta^{\otimes n}\,\middle|\,\rho_{\theta_0}^{\otimes n}\right) = \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes \mathcal{L}\left(\rho_\theta|\rho_{\theta_0}\right) \otimes I^{\otimes(n-k)}. \tag{2.6}$$

This proves $\rho_\theta^{\otimes n} \sim \rho_{\theta_0}^{\otimes n}$ for all $\theta \in \Theta$ and $n \in \mathbb{N}$.

Before proceeding to the proof of (ii) and (iii) in Definition 2.8, we give some preliminary consideration. Let the quantum log-likelihood ratio $\mathcal{L}_h := \mathcal{L}\left(\rho_{\theta_0+h}|\rho_{\theta_0}\right)$ be expanded into

$$\mathcal{L}_h = h^i A_i + B_{ij} h^i h^j + o(h^2),$$

where $A_i\,(1 \le i \le d)$ and $B_{ij}\,(1 \le i, j \le d)$ are Hermitian operators on $\mathcal{H}$. Observe that A_i is the SLD in the ith direction. In fact,

$$\begin{aligned}
\rho_{\theta_0+h} &= \exp\left[\frac{1}{2}\left(h^i A_i + o(h)\right)\right] \rho_{\theta_0} \exp\left[\frac{1}{2}\left(h^i A_i + o(h)\right)\right] \\
&= \rho_{\theta_0} + \frac{1}{2} h^i \left(\rho_{\theta_0} A_i + A_i \rho_{\theta_0}\right) + o(h),
\end{aligned}$$

so that

$$\partial_i \rho_{\theta_0} = \frac{1}{2}\left(\rho_{\theta_0} A_i + A_i \rho_{\theta_0}\right).$$

This observation also shows that $\mathrm{Tr}\,\rho_{\theta_0} A_i = 0$ for all i. On the other hand,

$$\begin{aligned}
\mathrm{Tr}\,\rho_{\theta_0+h} &= \mathrm{Tr}\,\rho_{\theta_0} \exp\left(h^i A_i + B_{ij} h^i h^j + o(h^2)\right) \\
&= \mathrm{Tr}\,\rho_{\theta_0} \left(I + \left(h^i A_i + B_{ij} h^i h^j\right) + \frac{1}{2}\left(h^i A_i + B_{ij} h^i h^j\right)^2 + o(h^2)\right) \\
&= 1 + h^i \left(\mathrm{Tr}\,\rho_{\theta_0} A_i\right) + h^i h^j \mathrm{Tr}\,\rho_{\theta_0} \left(B_{ij} + \frac{1}{2} A_i A_j\right) + o(h^2) \\
&= 1 + h^i h^j \mathrm{Tr}\,\rho_{\theta_0} \left(B_{ij} + \frac{1}{2} A_i A_j\right) + o(h^2).
\end{aligned}$$

Since $\mathrm{Tr}\,\rho_{\theta_0+h} = 1$ for all h, the above equation leads to

$$\mathrm{Tr}\,\rho_{\theta_0} \left(B_{ij} + \frac{1}{2} A_i A_j\right) = 0. \tag{2.7}$$

Now we prove (ii). Let $J_{ij} := \mathrm{Tr}\,\rho_{\theta_0} A_j A_i$, and let

$$\Delta_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes A_i \otimes I^{\otimes(n-k)}.$$

It then follows from the quantum central limit theorem (Proposition 2.5) that $\left(\Delta^{(n)}, \rho_{\theta_0}^{\otimes n}\right) \underset{q}{\rightsquigarrow} N(0, J)$.

Finally, we prove (iii). It follows from (2.6) that

$$\mathcal{L}_h^{(n)} = \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes \mathcal{L}_{h/\sqrt{n}} \otimes I^{\otimes(n-k)}.$$

Let us show that

$$R_h^{(n)} := \mathcal{L}_h^{(n)} - \left(h^i \Delta_i^{(n)} - \frac{1}{2}\left(J_{ij} h^i h^j\right) I^{\otimes n}\right)$$

is infinitesimal relative to the convergence $\left(\Delta^{(n)}, \rho_{\theta_0}^{\otimes n}\right) \underset{q}{\rightsquigarrow} N(0, J)$. It is rewritten as

$$\begin{aligned} R_h^{(n)} &= \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes \left[\mathcal{L}_{h/\sqrt{n}} - \frac{1}{\sqrt{n}} h^i A_i + \frac{1}{2n}\left(J_{ij} h^i h^j\right) I\right] \otimes I^{\otimes(n-k)} \\ &= \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes \left[\frac{1}{\sqrt{n}} h^i A_i + \frac{1}{n} B_{ij} h^i h^j + o(\frac{1}{n}) - \frac{1}{\sqrt{n}} h^i A_i + \frac{1}{2n}\left(J_{ij} h^i h^j\right) I\right] \otimes I^{\otimes(n-k)} \\ &= \sum_{k=1}^{n} I^{\otimes(k-1)} \otimes \left[\frac{1}{n} B_{ij} h^i h^j + \frac{1}{2n}\left(J_{ij} h^i h^j\right) I + o(\frac{1}{n})\right] \otimes I^{\otimes(n-k)} \\ &= \sum_{k=1}^{n} I^{\otimes k-1} \otimes \frac{1}{\sqrt{n}} P(n) \otimes I^{\otimes(n-k)}, \end{aligned}$$

where

$$P(n) := \sqrt{n}\left(\frac{1}{n}\left(B_{ij} + \frac{1}{2} J_{ij} I\right) h^i h^j + o(\frac{1}{n})\right).$$

Note that $\lim_{n\to\infty} P(n) = 0$, and that

$$\begin{aligned} \lim_{n\to\infty} \sqrt{n} \operatorname{Tr} \rho_{\theta_0} P(n) &= \operatorname{Tr} \rho_{\theta_0}\left(B_{ij} + \frac{1}{2} J_{ij} I\right) h^i h^j \\ &= \operatorname{Tr} \rho_{\theta_0}\left(B_{ij} + \frac{1}{2} J_{ji} I\right) h^i h^j \\ &= \operatorname{Tr} \rho_{\theta_0}\left(B_{ij} + \frac{1}{2} A_i A_j\right) h^i h^j \\ &= 0 \end{aligned}$$

because of (2.7). It then follows from Lemma 2.6 that $R_h^{(n)} = o(\Delta^{(n)}, \rho_{\theta_0}^{\otimes n})$ for each $h \in \mathbb{R}^d$. This completes the proof.

2.3.4 Proof of Corollary 2.11

That $\rho_\theta^{\otimes n} \sim \rho_{\theta_0}^{\otimes n}$ was proven in the proof of Theorem 2.10. Let $\Delta_1^{(n)}, \ldots, \Delta_d^{(n)}$ be as in the proof of Theorem 2.10. It then follows from the quantum central limit theorem that

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{\otimes n}\right) \underset{q}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^* & J \end{pmatrix}\right). \tag{2.8}$$

Further, because of Lemma 2.6, the sequence $R_h^{(n)}$ of observables given in the proof of Theorem 2.10 is also infinitesimal relative to the convergence (2.8). Now that $(\rho_\theta^{\otimes n}, X^{(n)})$ are jointly QLAN at θ_0, the property $\left(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\right) \underset{q}{\rightsquigarrow} N((\operatorname{Re}\tau)h, \Sigma)$ is an immediate consequence of Theorem 2.9. This completes the proof.

2.4 Applications to quantum statistics

Quantum Le Cam's third lemma (Corollary 2.11) implies convergence of $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\,;\, h \in \mathbb{R}^d\right\}$ to a quantum Gaussian shift model $\left\{N((\operatorname{Re}\tau)\, h,\, \Sigma)\,;\, h \in \mathbb{R}^d\right\}$. This fact prompts us to expect that, for sufficiently large n, the estimation problem for the parameter h of $\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}$ could be reduced to that for the shift parameter h of a quantum Gaussian shift model $N((\operatorname{Re}\tau)\, h,\, \Sigma)$. The latter problem has been well-established to date (see Appendix 2.B). In particular, the best strategy for estimating the shift parameter h is the one that achieves the Holevo bound $C_h\left(N((\operatorname{Re}\tau)\, h, \Sigma),\, G\right)$, (see Theorem 2.25). Moreover, it can be shown (see Corollary 2.24) that the Holevo bound $C_h\left(N((\operatorname{Re}\tau)\, h, \Sigma),\, G\right)$ is identical to the Holevo bound $C_{\theta_0}\left(\rho_\theta,\, G\right)$ for the

model ρ_θ at θ_0 This observation strongly suggests the existence of a sequence $M^{(n)}$ of estimators for the parameter h of $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\right\}_n$ that asymptotically achieves the Holevo bound $C_{\theta_0}(\rho_\theta, G)$.

This section is devoted to materialize this program: we prove that there exist a sequence $M^{(n)}$ of estimators on $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\right\}_n$ that is asymptotically unbiased and achieves the Holevo bound $C_{\theta_0}(\rho_\theta, G)$ for all h that belong to a dense subset of $\mathbb{R}^d$. Since this result requires only twice differentiability of the quantum log-likelihood ratio of the model ρ_θ, it will be useful in a wide range of statistical estimation problems.

2.4.1 Achievability of the Holevo bound

Theorem 2.12. *Let $\{\rho_\theta\,;\,\theta\in\Theta\subset\mathbb{R}^d\}$ be a quantum statistical model on a finite dimensional Hilbert space $\mathcal{H}$, and fix a point $\theta_0\in\Theta$. Suppose that $\rho_\theta\sim\rho_{\theta_0}$ for all $\theta\in\Theta$, and that the quantum log-likelihood ratio $\mathcal{L}_h := \mathcal{L}(\rho_{\theta_0+h}|\rho_{\theta_0})$ is differentiable in h around $h=0$ and twice differentiable at $h=0$. For any countable dense subset D of $\mathbb{R}^d$ and any weight matrix G, there exist a sequence $M^{(n)}$ of estimators on the model $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\,;\,h\in\mathbb{R}^d\right\}$ that enjoys*

$$\lim_{n\to\infty} E_h^{(n)}[M^{(n)}] = h$$

and

$$\lim_{n\to\infty} \operatorname{Tr} G V_h^{(n)}[M^{(n)}] = C_{\theta_0}(\rho_\theta, G)$$

for every $h\in D$. Here $C_{\theta_0}(\rho_\theta, G)$ is the Holevo bound at θ_0.

Proof. Let $\mathcal{D} := \mathcal{D}_{\rho_{\theta_0}}$ be the commutation operator with respect to the state ρ_{θ_0} (see Appendix 2.A), and let $\mathcal{T}$ be the minimal $\mathcal{D}$ invariant extension of the SLD tangent space $\operatorname{span}_{\mathbb{R}}\{L_i\}_{i=1}^d$ of the model $\{\rho_\theta\}$ at $\theta=\theta_0$, i.e., the smallest $\mathcal{D}$ invariant real linear subspace of Hermitian operators on $\mathcal{H}$ containing all the SLDs $\{L_i\}_{i=1}^d$ of ρ_θ at θ_0. The minimality ensures that $\operatorname{Tr}\rho_{\theta_0}A=0$ for all $A\in\mathcal{T}$ because $\mathcal{T}'=\{A\in\mathcal{T}\,;\,\operatorname{Tr}\rho_{\theta_0}A=0\}$ is also $\mathcal{D}$ invariant.

Let $\{D_j\}_{j=1}^r$ be a basis of $\mathcal{T}$, thus $d\le r$. Let Σ be an $r\times r$ matrix whose (i,j)th entry is given by $\Sigma_{ij}=\operatorname{Tr}\rho_{\theta_0}D_jD_i$, and let τ be an $r\times d$ matrix whose (i,j)th entry is given by $\tau_{ij}=\operatorname{Tr}\rho_{\theta_0}L_jD_i$. According to Theorem 2.19 in Appendix 2.A, the Holevo bound for a weight $G>0$ can be expressed as

$$\begin{aligned} C_{\theta_0}(\rho_\theta, G) &= \min_F\{\operatorname{Tr} GZ + \operatorname{Tr}\left|\sqrt{G}\operatorname{Im} Z\sqrt{G}\right|\,;\, Z = {}^tF\Sigma F, \\ &\qquad F \text{ is an } r\times d \text{ real matrix satisfying } {}^tF\operatorname{Re}(\tau)=I\}. \end{aligned} \tag{2.9}$$

Letting

$$X_i^{(n)} := \frac{1}{\sqrt{n}}\sum_{k=1}^n I^{\otimes(k-1)}\otimes D_i\otimes I^{\otimes(n-k)} \quad (1\le i\le r),$$

Corollary 2.11 asserts that $\left(\left\{\rho_\theta^{\otimes n}\right\}, X^{(n)}\right)$ is jointly QLAN at θ_0, and that

$$\left(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{\otimes n}\right) \underset{q}{\rightsquigarrow} N((\operatorname{Re}\tau)h, \Sigma). \tag{2.10}$$

Let F be the matrix that attains the minimum in (2.9), and let $Z := {}^tF\Sigma F$, $\tilde{V} := \operatorname{Re} Z$, $\tilde{S} := \operatorname{Im} Z$, $\hat{V} = \sqrt{G^{-1}}\left|\sqrt{G}\operatorname{Im} Z\sqrt{G}\right|\sqrt{G^{-1}}$, and $\hat{Z} = \hat{V} - \sqrt{-1}\tilde{S}$. It then follows from Corollary 2.24 and Theorem 2.25 in Appendix 2.B that

$$C_{\theta_0}(\rho_\theta, G) = \operatorname{Tr} G\left(\tilde{V}+\hat{V}\right).$$

Further, Lemma 2.13 below assures that there exist a finite dimensional Hilbert space $\hat{\mathcal{H}}$ and a state σ and observables B_i $(1\le i\le d)$ on $\hat{\mathcal{H}}$ such that $\operatorname{Tr}\sigma B_i=0$ and $\operatorname{Tr}\sigma B_jB_i=\hat{Z}_{ij}$. Let

$$\overline{X}_i^{(n)} := \tilde{X}_i^{(n)}\otimes\hat{I}^{\otimes n} + I^{\otimes n}\otimes Y_i^{(n)} \quad (1\le i\le d),$$

where $\tilde{X}^{(n)} := F_i^k X_k^{(n)}$ $(1 \le i \le d)$,

$$Y_i^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \hat{I}^{\otimes(k-1)} \otimes B_i \otimes \hat{I}^{\otimes(n-k)} \quad (1 \le i \le d),$$

and $\hat{I}$ is the identity on $\hat{\mathcal{H}}$. A crucial observation is that $\left(\overline{X}^{(n)}, \overline{\rho}_h^{(n)}\right)$, where $\overline{\rho}_h^{(n)} := \rho_{\theta_0+h/\sqrt{n}}^{\otimes n} \otimes \sigma^{\otimes n}$, converges to a classical Gaussian state:

$$\left(\overline{X}^{(n)}, \overline{\rho}_h^{(n)}\right) \underset{q}{\rightsquigarrow} N(h, \tilde{V} + \hat{V}), \tag{2.11}$$

for all $h \in \mathbb{R}^d$. In fact,

$$\begin{aligned}
\lim_{n\to\infty} \operatorname{Tr} \overline{\rho}_h^{(n)} \left(\prod_{t=1}^{s} \mathrm{e}^{\sqrt{-1}\xi_t^i \overline{X}_i^{(n)}} \right) &= \lim_{n\to\infty} \operatorname{Tr} \overline{\rho}_h^{(n)} \left\{ \left(\prod_{t=1}^{s} \mathrm{e}^{\sqrt{-1}\xi_t^i \tilde{X}_i^{(n)}} \right) \otimes \left(\prod_{t=1}^{s} \mathrm{e}^{\sqrt{-1}\xi_t^i Y_i^{(n)}} \right) \right\} \\
&= \lim_{n\to\infty} \left[\operatorname{Tr} \rho_{\theta_0+h/\sqrt{n}}^{\otimes n} \left(\prod_{t=1}^{s} \mathrm{e}^{\sqrt{-1}\xi_t^i \tilde{X}_i^{(n)}} \right) \right] \left[\operatorname{Tr} \sigma^{\otimes n} \left(\prod_{t=1}^{s} \mathrm{e}^{\sqrt{-1}\xi_t^i Y_i^{(n)}} \right) \right] \\
&= \phi_h \left(\prod_{t=1}^{s} \mathrm{e}^{\sqrt{-1}\xi_t^i \tilde{X}_i} \right) \psi \left(\prod_{t=1}^{s} \mathrm{e}^{\sqrt{-1}\xi_t^i Y_i} \right),
\end{aligned} \tag{2.12}$$

where $\tilde{X}_i := F_i^k X_k$ $(1 \le i \le d)$ are canonical observables with $X_1, \ldots, X_r$ being the basic canonical observables of CCR $(\operatorname{Im}\Sigma)$ and $(X, \phi_h) \sim N((\operatorname{Re}\tau)h, \Sigma)$, and $Y_1, \ldots, Y_d$ are the basic canonical observables of CCR $\left(\operatorname{Im}\hat{Z}\right)$ with $(Y, \psi) \sim N(0, \hat{Z})$. In the last line in (2.12), we used (2.10) as well as the quantum central limit theorem for $Y^{(n)}$. By using the explicit form (2.1) of the quasi-characteristic function for the quantum Gaussian state, (2.12) is rewritten as

$$\begin{aligned}
&\exp\left(\sum_{t=1}^{r} \left(\sqrt{-1}\xi_t^i h_i - \frac{1}{2}\xi_t^i \xi_t^j Z_{ji} \right) - \sum_{t=1}^{r} \sum_{s=t+1}^{r} \xi_t^i \xi_s^j Z_{ji} \right) \exp\left(-\frac{1}{2} \sum_{t=1}^{r} \xi_t^i \xi_t^j \hat{Z}_{ji} - \sum_{t=1}^{r} \sum_{s=t+1}^{r} \xi_t^i \xi_s^j \hat{Z}_{ji} \right) \\
&\qquad = \exp\left(\sum_{t=1}^{r} \left(\sqrt{-1}\xi_t^i h_i - \frac{1}{2}\xi_t^i \xi_t^j (\tilde{V} + \hat{V})_{ji} \right) - \sum_{t=1}^{r} \sum_{s=t+1}^{r} \xi_t^i \xi_s^j (\tilde{V} + \hat{V})_{ji} \right).
\end{aligned}$$

This proves (2.11).

Now according to Lemma 2.14 below, there exist a quintuple sequence

$$M^{(n,m,\ell,q,p)} = \left\{ M_\omega^{(n,m,\ell,q,p)} \,;\, \omega \in \Omega^{(n,m,l,p,q)} \right\}$$

of POVMs on $\left(\mathcal{H} \otimes \hat{\mathcal{H}}\right)^{\otimes n}$, taking values in a certain finite subset $\Omega^{(n,m,l,p,q)}$ of $\mathbb{R}^d$, that enjoys the properties

$$\lim_{p\to\infty} \lim_{q\to\infty} \lim_{\ell\to\infty} \lim_{m\to\infty} \lim_{n\to\infty} \overline{E}_h^{(n)}[M^{(n,m,\ell,q,p)}] = h,$$

and

$$\lim_{p\to\infty} \lim_{q\to\infty} \lim_{\ell\to\infty} \lim_{m\to\infty} \lim_{n\to\infty} \overline{V}_h^{(n)}[M^{(n,m,\ell,q,p)}] = \tilde{V} + \hat{V},$$

for all $h \in \mathbb{R}^d$, where $\overline{E}_h^{(n)}[\,\cdot\,]$ and $\overline{V}_h^{(n)}[\,\cdot\,]$ denote the expectation and the covariance with respect to $\overline{\rho}_h^{(n)}$. It then follows from Lemma 2.15 below that for any countable dense subset D of $\mathbb{R}^d$ and any $h \in D$, there exist a subsequence $\{(n, m(n), \ell(n), q(n), p(n)\}_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \overline{E}_h^{(n)}[M^{(n,m(n),\ell(n),q(n),p(n))}] = h,$$

and

$$\lim_{n\to\infty} \overline{V}_h^{(n)}[M^{(n,m(n),\ell(n),q(n),p(n))}] = \tilde{V} + \hat{V}.$$

This implies that the POVM $M^{(n)}$ on $\mathcal{H}^{\otimes n}$ that is uniquely defined by the requirement

$$\mathrm{Tr}\,\rho^{(n)} M_\omega^{(n)} = \mathrm{Tr}\,\left(\rho^{(n)} \otimes \sigma^{\otimes n}\right) M_\omega^{(n,m(n),\ell(n),q(n),p(n))}$$

for all density operator $\rho^{(n)}$ on $\mathcal{H}^{\otimes n}$ and $\omega \in \Omega^{(n,m(n),l(n),p(n),q(n))}$ enjoys

$$\lim_{n\to\infty} E_h^{(n)}[M^{(n)}] = h,$$

$$\lim_{n\to\infty} V_h^{(n)}[M^{(n)}] = \tilde{V} + \hat{V}.$$

for all $h \in D$. Recalling that $\mathrm{Tr}\, G(\tilde{V}+\hat{V}) = C_{\theta_0}(\rho_\theta, G)$, the proof is complete. □

Lemma 2.13. *Given a $d \times d$ positive semidefinite Hermitian matrix J, there exist a finite dimensional Hilbert space $\mathcal{H}$ and a pure state ρ and observables A_i $(1 \le i \le d)$ on $\mathcal{H}$ such that $\mathrm{Tr}\,\rho A_i = 0$ and $\mathrm{Tr}\,\rho A_j A_i = J_{ij}$.*

Proof. Let $\mathcal{H} = \mathbb{C}^{d+1}$, and let $\{|i\rangle\}_{i=0}^d$ be a CONS of $\mathcal{H}$. We set $|\psi\rangle := |0\rangle$ and $|\ell_i\rangle := \sum_{k=1}^d \left[\sqrt{J}\right]_{ik} |k\rangle$ for $i = 1, \ldots, d$. Then $\rho := |\psi\rangle\langle\psi|$ and $A_i := |\ell_i\rangle\langle\psi| + |\psi\rangle\langle\ell_i|$ satisfy $\mathrm{Tr}\,\rho A_i = 0$ and $\mathrm{Tr}\,\rho A_j A_i = J_{ij}$. □

Lemma 2.14. *Given a sequence $\mathcal{H}^{(n)}$ of finite dimensional Hilbert spaces, let $X^{(n)} = \left(X_1^{(n)}, \ldots, X_d^{(n)}\right)$ be a list of observables on $\mathcal{H}^{(n)}$, and let $\left\{\rho_h^{(n)}\right\}_h$ be a family of density operators on $\mathcal{H}^{(n)}$ parametrized by $h \in \mathbb{R}^d$. If there is a real $d \times d$ positive definite matrix V such that*

$$\left(X^{(n)}, \rho_h^{(n)}\right) \underset{q}{\rightsquigarrow} N(h, V) \tag{2.13}$$

holds for all $h \in \mathbb{R}^d$, then there exist a quintuple sequence $\left\{M^{(n,m,\ell,q,p)}; (n,m,\ell,q,p) \in \mathbb{N}^5\right\}$ of POVMs on $\mathcal{H}^{(n)}$ that enjoy the properties

$$\lim_{p\to\infty}\lim_{q\to\infty}\lim_{\ell\to\infty}\lim_{m\to\infty}\lim_{n\to\infty} E_h^{(n)}[M^{(n,m,\ell,q,p)}] = h,$$

and

$$\lim_{p\to\infty}\lim_{q\to\infty}\lim_{\ell\to\infty}\lim_{m\to\infty}\lim_{n\to\infty} V_h^{(n)}[M^{(n,m,\ell,q,p)}] = V.$$

Proof. Let

$$\Omega^{(m,\ell)} := \left\{ \frac{\ell}{m}\overrightarrow{k} + \frac{\ell}{2m}(1, \ldots, 1);\ \overrightarrow{k} \in \mathbb{Z}^d \right\} \cap [-l, l]^d$$

be a finite subset of $\mathbb{R}^d$, comprising $(2m)^d$ lattice points in the hypercube $[-l,l]^d$, and let $\Omega^{(m,\ell,p)} := \Omega^{(m,\ell)} \cap [-p,p]^d$ and $\Omega_0^{(m,\ell,p)} := \Omega^{(m,\ell,p)} \cup \{0\}$. We introduce a Gaussian density function $f_\omega^{(q)}(x)$ on $\mathbb{R}^d$ centered at $\omega = (\omega_1, \ldots, \omega_d) \in \mathbb{R}^d$ by

$$f_\omega^{(q)}(x) := \left\{\prod_{i=1}^d g_{\omega_{d+1-i}}^{(q)}(x_{d+1-i})\right\}\left\{\prod_{i=1}^d g_{\omega_i}^{(q)}(x_i)\right\},$$

where $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and

$$g_s^{(q)}(t) := \left(\frac{q}{2\pi}\right)^{\frac{1}{4}} \exp\left(-\frac{q}{4}(t-s)^2\right), \qquad (s,t \in \mathbb{R}).$$

By using this function, we define a POVM $M^{(n,m,l,q,p)} = \left\{M_\omega^{(n,m,l,q,p)};\ \omega \in \Omega_0^{(m,\ell,p)}\right\}$ on $\mathcal{H}^{(n)}$ that takes values in the finite subset $\Omega_0^{(m,\ell,p)}$ by

$$M_\omega^{(n,m,\ell,q,p)} := R^{(m,\ell,q)}(X^{(n)})\left[f_\omega^{(q)}(X^{(n)}) + \frac{I^{(n)}}{(2m)^d}\right] R^{(m,\ell,q)}(X^{(n)})$$

for $\omega \in \Omega^{(m,\ell,p)}$, and

$$M_0^{(n,m,\ell,q,p)} := \sum_{\omega \in \Omega^{(m,\ell)} \setminus \Omega^{(m,\ell,p)}} \left\{ R^{(m,\ell,q)}(X^{(n)}) \left[\left(f_\omega^{(q)}(X^{(n)}) + \frac{I^{(n)}}{(2m)^d} \right) \right] R^{(m,\ell,q)}(X^{(n)}) \right\}.$$

Here

$$R^{(m,\ell,q)}(x) := g\left(\sum_{\omega \in \Omega^{(m,\ell)}} f_\omega^{(q)}(x) \right)$$

is the normalization with

$$g(t) := \frac{1}{\sqrt{t+1}}.$$

Intuitively speaking, the difference set $\Omega^{(m,\ell)} \setminus \Omega^{(m,\ell,p)}$ works as a "buffer" zone that gives the default outcome $\omega = 0$. This device is meaningful only when $p < \ell$.

We shall prove that

$$\lim_{p\to\infty} \lim_{q\to\infty} \lim_{\ell\to\infty} \lim_{m\to\infty} \lim_{n\to\infty} \sum_{\omega \in \Omega_0^{(m,\ell,p)}} P(\omega) \mathrm{Tr}\, \rho_h^{(n)} M_\omega^{(n,m,\ell,q,p)} = \int_{\mathbb{R}^d} P(\omega) p_h(\omega) \mathrm{d}\omega, \tag{2.14}$$

where $P(\omega)$ is an arbitrary polynomial of ω such that $P(0) = 0$ and $p_h(\omega)$ is a probability density function of the classical normal distribution $N(h, V)$. Once (2.14) has been proved, we can verify

$$\lim_{p\to\infty} \lim_{q\to\infty} \lim_{\ell\to\infty} \lim_{m\to\infty} \lim_{n\to\infty} E_h^{(n)}[M^{(n,m,\ell,q,p)}] = h$$

and

$$\lim_{p\to\infty} \lim_{q\to\infty} \lim_{\ell\to\infty} \lim_{m\to\infty} \lim_{n\to\infty} V_h^{(n)}[M^{(n,m,\ell,q,p)}] = V$$

just by letting $P(\omega) = \omega_i$ or $P(\omega) = \omega_i \omega_j$ $(1 \le i, j \le d)$ in (2.14).

The first limit $n \to \infty$ in (2.14) yields

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{\omega \in \Omega_0^{(m,\ell,p)}} P(\omega) \mathrm{Tr}\, \rho_h^{(n)} M_\omega^{(n,m,\ell,q,p)} \\
&= \lim_{n\to\infty} \sum_{\omega \in \Omega^{(m,\ell,p)}} P(\omega) \mathrm{Tr}\, \rho_h^{(n)} M_\omega^{(n,m,\ell,q,p)} \\
&= \lim_{n\to\infty} \sum_{\omega \in \Omega^{(m,\ell,p)}} P(\omega) \mathrm{Tr}\, \rho_h^{(n)} R^{(m,\ell,q)}(X^{(n)}) \left[f_\omega^{(q)}(X^{(n)}) + \frac{I^{(n)}}{(2m)^d} \right] R^{(m,\ell,q)}(X^{(n)}) \\
&= \sum_{\omega \in \Omega^{(m,\ell,p)}} P(\omega) E_h \left[R^{(m,\ell,q)}(X)^2 \left(f_\omega^{(q)}(X) + \frac{I}{(2m)^d} \right) \right] \\
&= \int_{\mathbb{R}^d} \frac{\sum_{\omega \in \Omega^{(m,\ell,p)}} P(\omega) \left(f_\omega^{(q)}(x) + \frac{1}{(2m)^d} \right)}{\sum_{\omega \in \Omega^{(m,\ell)}} \left(f_\omega^{(q)}(x) + \frac{1}{(2m)^d} \right)} p_h(x) \mathrm{d}x.
\end{aligned} \tag{2.15}$$

In the fourth line, we used the assumption (2.13) and Corollary 2.29 in Appendix 2.D, as well as the fact that functions $g_s^{(q)}(t)$ on $\mathbb{R}$ and $g(t)$ on $t \ge 0$ are both bounded and continuous. Further, $X = (X_1, \ldots, X_d)$ is a classical random vector that follow the normal distribution $N(h, V)$, and $E_h[\,\cdot\,]$ denotes the expectation with respect to $N(h, V)$. As for the second limit $m \to \infty$, due to

$$\left| \frac{\sum_{\omega \in \Omega^{(m,\ell,p)}} P(\omega) \left(f_\omega^{(q)}(x) + \frac{1}{(2m)^d} \right)}{\sum_{\omega \in \Omega^{(m,\ell)}} \left(f_\omega^{(q)}(x) + \frac{1}{(2m)^d} \right)} \right| \le \max_{\omega \in [-p,p]^d} |P(\omega)| < \infty,$$

the bounded convergence theorem yields

$$\begin{aligned}\lim_{m\to\infty}(2.15) &= \int_{\mathbb{R}^d}\lim_{m\to\infty}\frac{\left(\frac{\ell}{m}\right)^d\sum_{\omega\in\Omega^{(m,\ell,p)}}P(\omega)\left(f_\omega^{(q)}(x)+\frac{1}{(2m)^d}\right)}{\left(\frac{\ell}{m}\right)^d\sum_{\omega\in\Omega^{(m,\ell)}}\left(f_\omega^{(q)}(x)+\frac{1}{(2m)^d}\right)}p_h(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^d}\frac{\int_{\omega\in[-p,p]^d}P(\omega)p^{(q)}(\omega,x)\mathrm{d}\omega}{\int_{\omega\in[-\ell,\ell]^d}p^{(q)}(\omega,x)\mathrm{d}\omega}p_h(x)\mathrm{d}x, \qquad (2.16)\end{aligned}$$

where $p^{(q)}(\omega,x)=\left(\frac{q}{2\pi}\right)^{\frac{d}{2}}\exp\left(-\frac{q}{2}\sum_{i=1}^d(x_i-\omega_i)^2\right)$, and Darboux's theorem for the Riemann integral was used in the second line. Finally, the dominated convergence theorem and Fubini's theorem yield

$$\begin{aligned}\lim_{p\to\infty}\lim_{q\to\infty}\lim_{\ell\to\infty}(2.16) &= \lim_{p\to\infty}\lim_{q\to\infty}\int_{\mathbb{R}^d}\frac{\int_{\omega\in[-p,p]^d}P(\omega)p^{(q)}(\omega,x)\mathrm{d}\omega}{\int_{\mathbb{R}^d}p^{(q)}(\omega,x)\mathrm{d}\omega}p_h(x)\mathrm{d}x\\ &= \lim_{p\to\infty}\lim_{q\to\infty}\int_{\mathbb{R}^d}\left(\int_{\omega\in[-p,p]^d}P(\omega)p^{(q)}(\omega,x)\mathrm{d}\omega\right)p_h(x)\mathrm{d}x\\ &= \lim_{p\to\infty}\lim_{q\to\infty}\int_{\omega\in[-p,p]^d}\left(\int_{\mathbb{R}^d}p^{(q)}(\omega,x)p_h(x)\mathrm{d}x\right)P(\omega)\mathrm{d}\omega\\ &= \lim_{p\to\infty}\lim_{q\to\infty}\int_{\omega\in[-p,p]^d}p_h^{(q)}(\omega)P(\omega)\mathrm{d}\omega\\ &= \lim_{p\to\infty}\int_{\omega\in[-p,p]^d}p_h(\omega)P(\omega)\mathrm{d}\omega\\ &= \int_{\mathbb{R}^d}p_h(\omega)P(\omega)\mathrm{d}\omega, \qquad (2.17)\end{aligned}$$

where $p_h^{(q)}(\omega)$ is the density function of $N(h,V+\frac{1}{q}I)$. This completes the proof. □

Lemma 2.15. *For each $i\in\mathbb{N}$, let $\left\{a^i_{n_1n_2\cdots n_rn};(n_1,n_2,\ldots,n_r,n)\in\mathbb{N}^{(r+1)}\right\}$ be an $(r+1)$-tuple sequence on a normed space V. If, for each $i\in\mathbb{N}$, there exists an $\alpha^i\in V$ such that*

$$\lim_{n_1\to\infty}\lim_{n_2\to\infty}\cdots\lim_{n_r\to\infty}\lim_{n\to\infty}a^i_{n_1n_2\cdots n_rn}=\alpha^i,$$

then there exist a subsequence $\{(n_1(n),n_2(n),\ldots,n_r(n),n)\}_{n\in\mathbb{N}}$ that satisfies

$$\lim_{n\to\infty}a^i_{n_1(n)n_2(n)\cdots n_r(n)n}=\alpha^i$$

for all $i\in\mathbb{N}$.

Proof. We first prove the case when $r=1$. Let $a^i_{n_1}:=\lim_{n\to\infty}a^i_{n_1n}$. We construct a subsequence $\{(n_1(k),n(k))\}_{k\in\mathbb{N}}$ in a recursive manner as follows. We set $n_1(1)=n(1)=1$. For $k\geq 2$, it follows from $\lim_{n_1\to\infty}a^i_{n_1}=\alpha^i$ that there exist an $N_1(k)\in\mathbb{N}$ such that $n_1\geq N_1(k)$ implies

$$\max_{1\leq i\leq k}\left|a^i_{n_1}-\alpha^i\right|<\frac{1}{k}.$$

Thus the number $n_1(k):=\max\{N_1(k),n_1(k-1)+1\}$ satisfies

$$\max_{1\leq i\leq k}\left|a^i_{n_1(k)}-\alpha^i\right|<\frac{1}{k}. \qquad (2.18)$$

For this $n_1(k)$, it follows from $\lim_{n\to\infty}a^i_{n_1(k)n}=a^i_{n_1(k)}$ that there exist an $N(k)\in\mathbb{N}$ such that $n\geq N(k)$ implies

$$\max_{1\leq i\leq k}\left|a^i_{n_1(k)n}-a^i_{n_1(k)}\right|<\frac{1}{k}. \qquad (2.19)$$

Thus we set $n(k) := \max\{N(k), n(k-1)+1\}$.

Now let $k(n) := \max\{k;\, n(k) \le n\}$, which is non-decreasing in n and $\lim_{n\to\infty} k(n) = \infty$. We show that the subsequence $\{n_1(k(n)), n);\, n \in \mathbb{N}\}$ enjoys the required property: for all $i \in \mathbb{N}$,

$$\lim_{n\to\infty} a^i_{n_1(k(n))n} = \alpha^i.$$

Given $i \in \mathbb{N}$ and $\varepsilon > 0$ arbitrarily, there exist an $N \in \mathbb{N}$ such that $n \ge N$ implies $k(n) \ge \max\left\{i, \left\lceil \frac{2}{\varepsilon} \right\rceil\right\}$. Then for all $n \ge N$, we have

$$\begin{aligned}
\left|a^i_{n_1(k(n))n} - \alpha^i\right| &\le \left|a^i_{n_1(k(n))n} - a^i_{n_1(k(n))}\right| + \left|a^i_{n_1(k(n))} - \alpha^i\right| \\
&\le \max_{1\le j\le k(n)} \left|a^j_{n_1(k(n))n} - a^j_{n_1(k(n))}\right| + \max_{1\le j\le k(n)} \left|a^j_{n_1(k(n))} - \alpha^j\right| \\
&< \frac{2}{k(n)} \le \varepsilon.
\end{aligned}$$

In the third inequality, we used (2.18) and (2.19), as well as its premise $n \ge n(k(n)) \ge N(k(n))$.

The proof for a generic r is similar. □

2.4.2 Application to qubit state estimation

In order to demonstrate the power of our method, we explore qubit state estimation problems.

Example 2.16 (3-dimensional faithful state model)**.**

The first example is an ordinary one, comprising the totality of faithful quit states:

$$\mathcal{S}(\mathbb{C}^2) = \left\{ \rho_\theta = \frac{1}{2}\left(I + \theta^1\sigma_1 + \theta^2\sigma_2 + \theta^3\sigma_3\right)\ ;\ \theta = (\theta^i)_{1\le i\le 3} \in \Theta \right\}$$

where σ_i $(i = 1, 2, 3)$ are the standard Pauli matrices and Θ is the open unit ball in $\mathbb{R}^3$. Due to the rotational symmetry, we take the reference point to be $\theta_0 = (0, 0, r)$, with $0 \le r < 1$. By a direct calculation, we see that the SLDs of the model ρ_θ at $\theta = \theta_0$ are $(L_1, L_2, L_3) = \left(\sigma_1, \sigma_2, (rI + \sigma_3)^{-1}\right)$, and the SLD Fisher information matrix $J^{(S)}$ at θ_0 is given by the real part of the matrix

$$J := \left[\operatorname{Tr} \rho_{\theta_0} L_j L_i\right]_{ij} = \begin{pmatrix} 1 & -r\sqrt{-1} & 0 \\ r\sqrt{-1} & 1 & 0 \\ 0 & 0 & 1/(1-r^2) \end{pmatrix}.$$

Given a 3×3 real positive definite matrix G, the minimal value of the weighted covariances at $\theta = \theta_0$ is given by

$$\min_{\hat{M}} \operatorname{Tr} G V_{\theta_0}[\hat{M}] = C^{(1)}_{\theta_0}(\rho_\theta, G),$$

where the minimum is taken over all estimators $\hat{M}$ that are locally unbiased at θ_0, and

$$C^{(1)}_{\theta_0}(\rho_\theta, G) = \left(\operatorname{Tr}\sqrt{\sqrt{G} J^{(S)^{-1}} \sqrt{G}}\right)^2$$

is the Hayashi-Gill-Massar bound [17, 13] (see also [49]). On the other hand, the SLD tangent space is obviously $\mathcal{D}$ invariant, and the Holevo bound is given by

$$C_{\theta_0}(\rho_\theta, G) := \operatorname{Tr} G J^{(R)^{-1}} + \operatorname{Tr}\left|\sqrt{G}\operatorname{Im} J^{(R)^{-1}}\sqrt{G}\right|,$$

where

$$J^{(R)^{-1}} := (\operatorname{Re} J)^{-1} J (\operatorname{Re} J)^{-1} = \begin{pmatrix} 1 & -r\sqrt{-1} & 0 \\ r\sqrt{-1} & 1 & 0 \\ 0 & 0 & 1-r^2 \end{pmatrix}$$

is the inverse RLD Fisher information matrix (See Corollary 2.20 in Appendix 2.A).

It can be shown that the Hayashi-Gill-Massar bound is greater than the Holevo bound:

$$C^{(1)}_{\theta_0}(\rho_\theta, G) > C_{\theta_0}(\rho_\theta, G).$$

Let us check this fact for the special case when $G = J^{(S)}$. A direct computation shows that

$$C^{(1)}_{\theta_0}\left(\rho_\theta, J^{(S)}\right) = 9,$$

and

$$C_{\theta_0}\left(\rho_\theta, J^{(S)}\right) = 3 + 2r.$$

The left panel of Figure 2.1 shows the behavior of $C_{\theta_0}\left(\rho_\theta, J^{(S)}\right)$ (solid) and $C^{(1)}_{\theta_0}\left(\rho_\theta, J^{(S)}\right)$ (dashed) as functions of r. We see that the Holevo bound $C_{\theta_0}\left(\rho_\theta, J^{(S)}\right)$ is much smaller than $C^{(1)}_{\theta_0}\left(\rho_\theta, J^{(S)}\right)$.

Does this fact imply that the Holevo bound is of no use? The answer is contrary, as Theorem 2.12 asserts. We will demonstrate the asymptotic achievability of the Holevo bound. Let

$$\Delta^{(n)}_i := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes k-1} \otimes L_i \otimes I^{\otimes n-k}$$

and let $X^{(n)}_i := \Delta^{(n)}_i$ for $i = 1, 2, 3$. It follows from the quantum central limit theorem that

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho^{\otimes n}_{\theta_0} \right) \underset{q}{\rightsquigarrow} N\left(0, \begin{pmatrix} J & J \\ J & J \end{pmatrix}\right).$$

Since

$$\mathcal{L}(\theta) := \mathcal{L}(\rho_\theta | \rho_{\theta_0}) = 2 \log \left(\sqrt{\rho^{-1}_{\theta_0}} \sqrt{\sqrt{\rho_{\theta_0}} \rho_\theta \sqrt{\rho_{\theta_0}}} \sqrt{\rho^{-1}_{\theta_0}} \right)$$

is obviously of class C^∞ in θ, Corollary 2.11 shows that $\left(\left\{\rho^{\otimes n}_\theta\right\}, X^{(n)}\right)$ is jointly QLAN at θ_0, and that

$$\left(X^{(n)}, \rho^{\otimes n}_{\theta_0 + h/\sqrt{n}}\right) \underset{q}{\rightsquigarrow} N((\mathrm{Re}\, J)h, J)$$

for all $h \in \mathbb{R}^3$. This implies that a sequence of models $\left\{\rho^{\otimes n}_{\theta_0+h/\sqrt{n}}\,;\, h \in \mathbb{R}^d\right\}$ converges to a quantum Gaussian shift model $\left\{N((\mathrm{Re}\, J)h,\, J)\,;\, h \in \mathbb{R}^3\right\}$. Note that the imaginary part

$$S = \begin{pmatrix} 0 & -r\sqrt{-1} & 0 \\ r\sqrt{-1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

of the matrix J determines the CCR (S), as well as the corresponding basic canonical observables $X = (X^1, X^2, X^3)$. When $r \neq 0$, the above S has the following physical interpretation: X^1 and X^2 form a canonical pair of quantum Gaussian observables, while X^3 is a classical Gaussian random variable. In this way, the matrix J automatically tells us the structure of the limiting quantum Gaussian shift model.

Now, the best strategy for estimating the shift parameter h of the quantum Gaussian shift model $\left\{N((\mathrm{Re}\, J)h,\, J)\,;\, h \in \mathbb{R}^d\right\}$ is the one that achieves the Holevo bound $C_h\left(N((\mathrm{Re}\, J)\, h, J),\, G\right)$, (cf., Theorem 2.25 in Appendix 2.B). Moreover, this Holevo bound $C_h\left(N((\mathrm{Re}\, J)\, h, J),\, G\right)$ is identical to the Holevo bound $C_{\theta_0}(\rho_\theta,\, G)$ for the model ρ_θ at θ_0, (cf., Corollary 2.24. Recall that the matrix J is evaluated at θ_0 of the model ρ_θ). Theorem 2.12 combines these facts, and concludes that there exist a sequence $M^{(n)}$ of estimators on the model $\left\{\rho^{\otimes n}_{\theta_0+h/\sqrt{n}}\,;\, h \in \mathbb{R}^3\right\}$ that is asymptotically unbiased and achieves the common values of the Holevo bound:

$$\lim_{n\to\infty} \mathrm{Tr}\, G V^{(n)}_h[M^{(n)}] = C_h\left(N((\mathrm{Re}\, J)h, J), G\right) = C_{\theta_0}(\rho_\theta,\, G)$$

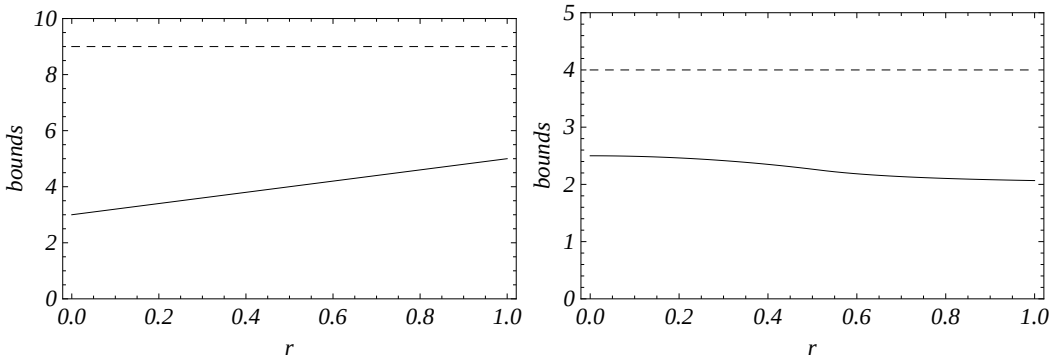

Figure 2.1: The left panel displays the Holevo bound $C_{(0,0,r)}\left(\rho_\theta, J^{(S)}\right)$ (solid) and the Hayashi-Gill-Massar bound $C^{(1)}_{(0,0,r)}\left(\rho_\theta, J^{(S)}\right)$ (dashed) for the 3-D model $\rho_\theta = \frac{1}{2}\left(I + \theta^1\sigma_1 + \theta^2\sigma_2 + \theta^3\sigma_3\right)$ as functions of $r = \|\theta\|$. The right panel displays the Holevo bound $C_{(0,r)}\left(\rho_\theta, J^{(S)}\right)$ (solid) and the Nagaoka bound $C^{(1)}_{(0,r)}\left(\rho_\theta, J^{(S)}\right)$ (dashed) for the 2-D model $\rho_\theta = \frac{1}{2}\left(I + \theta^1\sigma_1 + \theta^2\sigma_2 + \frac{1}{4}\sqrt{1-\|\theta\|^2}\,\sigma_3\right)$.

for all h that belong to a countable dense subset of $\mathbb{R}^3$.

It should be emphasized that the matrix J becomes the identity at the origin $\theta_0 = (0,0,0)$. This means that the limiting Gaussian shift model $\left\{N(h,J)\,;\, h \in \mathbb{R}^3\right\}$ is "classical." Since such a degenerate case cannot be treated in [15, 31, 19], our method has a clear advantage in applications.

Example 2.17 (Pure state model)**.**

The second example is to demonstrate that our formulation allows us to treat pure state models. Let us consider the model $\mathcal{S} = \{|\psi(\theta)\rangle\langle\psi(\theta)|\,;\, \theta = (\theta^i)_{1\le i\le 2} \in \Theta\}$ defined by

$$\psi(\theta) := \frac{1}{\sqrt{\cosh\|\theta\|}}\,\mathrm{e}^{\frac{1}{2}\left(\theta^1\sigma_1+\theta^2\sigma_2\right)}\begin{pmatrix}1\\0\end{pmatrix},$$

where Θ is an open subset of $\mathbb{R}^2$ containing the origin, and $\|\cdot\|$ denotes the Euclid norm. By a direct computation, the SLDs at $\theta_0 = (0,0)$ are $(L_1, L_2) = (\sigma_1, \sigma_2)$, and the SLD Fisher information matrix $J^{(S)}$ is the real part of the matrix

$$J = \left[\mathrm{Tr}\,\rho_{\theta_0}L_jL_i\right]_{ij} = \begin{pmatrix}1 & -\sqrt{-1}\\ \sqrt{-1} & 1\end{pmatrix},$$

that is, $J^{(S)} = I$. Since the SLD tangent space is $\mathcal{D}$ invariant [7], the Holevo bound for a weight $G > 0$ is represented as

$$C_{\theta_0}\left(\rho_\theta, G\right) := \mathrm{Tr}\,GJ^{(R)^{-1}} + \mathrm{Tr}\left|\sqrt{G}\,\mathrm{Im}\,J^{(R)^{-1}}\sqrt{G}\right|$$

where

$$J^{(R)^{-1}} := (\mathrm{Re}\,J)^{-1}J(\mathrm{Re}\,J)^{-1} = \begin{pmatrix}1 & -\sqrt{-1}\\ \sqrt{-1} & 1\end{pmatrix}$$

is the inverse RLD Fisher information matrix (see Corollary 2.20 in Appendix 2.A).

Let us demonstrate that our QLAN is applicable also to pure state models. Let

$$\Delta_i^{(n)} := \frac{1}{\sqrt{n}}\sum_{k=1}^{n} I^{\otimes k-1}\otimes L_i\otimes I^{\otimes n-k}$$

and let $X_i^{(n)} := \Delta_i^{(n)}$ for $i = 1, 2$. It follows from the quantum central limit theorem that

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{\otimes n} \right) \underset{q}{\rightsquigarrow} N\left(0, \begin{pmatrix} J & J \\ J & J \end{pmatrix}\right).$$

Since

$$\mathcal{L}(\theta) := \mathcal{L}\left(\rho_\theta | \rho_{\theta_0}\right) = \theta^1\sigma_1 + \theta^2\sigma_2 - \log\cosh\|\theta\|$$

is of class C^∞ with respect to θ, it follows from Corollary 2.11 that $\left(\left\{\rho_\theta^{\otimes n}\right\}, X^{(n)}\right)$ is jointly QLAN at θ_0, and that

$$(X^{(n)}, \rho_{\theta_0+h/\sqrt{n}}^{\otimes n}) \rightsquigarrow N((\mathrm{Re}\, J)h, J) = N(h, J^{(R)^{-1}})$$

for all $h \in \mathbb{R}^2$. Theorem 2.12 further asserts that there exist a sequence $M^{(n)}$ of estimators on the model $\left\{\rho_{\theta_0+h/\sqrt{n}}^{\otimes n};\ h \in \mathbb{R}^2\right\}$ that is asymptotically unbiased and achieves the Holevo bound:

$$\lim_{n\to\infty} \mathrm{Tr}\, G V_h^{(n)}[M^{(n)}] = C_h\left(N(h, J^{(R)^{-1}}),\, G\right) = C_{(0,0)}\left(\rho_\theta,\, G\right)$$

for all h that belong to a dense subset of $\mathbb{R}^3$. In fact, the sequence $M^{(n)}$ can be taken to be a separable one, making no use of quantum correlations [35]. (See also Appendix 2.C for a simple proof.) Note that the matrix $J^{(R)^{-1}}$ is degenerate, and the derived quantum Gaussian shift model $\left\{N(h, J^{(R)^{-1}})\right\}_h$ is a canonical coherent model [7].

Example 2.18 (2-dimensional faithful state model)**.**

The third example treats the case when the SLD tangent space is not $\mathcal{D}$ invariant. Let us consider the model

$$\mathcal{S} = \left\{ \rho_\theta = \frac{1}{2}\left(I + \theta^1\sigma_1 + \theta^2\sigma_2 + z_0\sqrt{1 - \|\theta\|^2}\,\sigma_3\right)\ ;\ \theta = (\theta^i)_{1\le i\le 2} \in \Theta \right\},$$

where $0 \le z_0 < 1$, and Θ is the open unit disk. Due to the rotational symmetry around z-axis, we take the reference point to be $\theta_0 = (0, r)$, with $0 \le r < 1$. By a direct calculation, we see that the SLDs at θ_0 are $(L_1,\, L_2) = \left(\sigma_1,\, \frac{1}{1-r^2}(\sigma_2 - rI)\right)$. It is important to notice that the SLD tangent space $\mathrm{span}\,\{L_i\}_{i=1}^2$ is not $\mathcal{D}$ invariant unless $r = 0$. In fact

$$\mathcal{D}\sigma_1 = z(r)\sigma_2 - r\sigma_3, \qquad \mathcal{D}\sigma_2 = -z(r)\sigma_1,$$

where $z(r) := E[\sigma_3] = z_0\sqrt{1-r^2}$. The minimal $\mathcal{D}$ invariant extension $\mathcal{T}$ of the SLD tangent space has a basis $(D_1,\, D_2,\, D_3) := (L_1,\, L_2,\, \sigma_3 - z(r)I)$. The matrices Σ, J, and τ appeared in Definition 2.8 and Corollary 2.11 are calculated as

$$\Sigma := [\mathrm{Tr}\,\rho_{\theta_0} D_j D_i]_{ij} = \begin{pmatrix} 1 & -\sqrt{-1}\dfrac{z_0^2}{z(r)} & r\sqrt{-1} - z(r) \\ \sqrt{-1}\dfrac{z_0^2}{z(r)} & \dfrac{z_0^2}{z(r)^2} & -\left(\dfrac{r}{z(r)} + \sqrt{-1}\right) z_0^2 \\ -r\sqrt{-1} - z(r) & -\left(\dfrac{r}{z(r)} - \sqrt{-1}\right) z_0^2 & 1 \end{pmatrix},$$

$$J := [\mathrm{Tr}\,\rho_{\theta_0} L_j L_i]_{ij} = \begin{pmatrix} 1 & -\sqrt{-1}\dfrac{z_0^2}{z(r)} \\ \sqrt{-1}\dfrac{z_0^2}{z(r)} & \dfrac{z_0^2}{z(r)^2} \end{pmatrix},$$

$$\tau := [\mathrm{Tr}\,\rho_{\theta_0} L_j \sigma_i]_{ij} = \begin{pmatrix} 1 & -\sqrt{-1}\dfrac{z_0^2}{z(r)} \\ \sqrt{-1}\dfrac{z_0^2}{z(r)} & \dfrac{z_0^2}{z(r)^2} \\ -r\sqrt{-1} - z(r) & -\left(\dfrac{r}{z(r)} - \sqrt{-1}\right) z_0^2 \end{pmatrix}.$$

Given a 2×2 real positive definite matrix G, the minimal value of the weighted covariances at $\theta = \theta_0$ is given by

$$\min_{\hat{M}} \operatorname{Tr} G V_{\theta_0}[\hat{M}] = C^{(1)}_{\theta_0}(\rho_\theta, G),$$

where the minimum is taken over all estimators $\hat{M}$ that are locally unbiased at θ_0, and

$$C^{(1)}_{\theta_0}(\rho_\theta, G) = \left(\operatorname{Tr} \sqrt{\sqrt{G} J^{(S)^{-1}} \sqrt{G}} \right)^2$$

is the Nagaoka bound [37] (see also [49]).

It can be shown that the Nagaoka bound is greater than the Holevo bound:

$$C^{(1)}_{\theta_0}(\rho_\theta, G) > C_{\theta_0}(\rho_\theta, G).$$

Let us check this fact for the special case when $G = J^{(S)}$. A direct computation shows that

$$C^{(1)}_{\theta_0}\left(\rho_\theta, J^{(S)}\right) = 4,$$

and

$$C_{\theta_0}\left(\rho_\theta, J^{(S)}\right) = \begin{cases} 2(1+z_0) - r^2(1-z_0^2), & \text{if } 0 \le r \le \sqrt{\dfrac{z_0}{1-z_0^2}} \\ 2 + \dfrac{z_0^2}{r^2(1-z_0^2)}, & \text{if } \sqrt{\dfrac{z_0}{1-z_0^2}} < r. \end{cases}$$

The right panel of Figure 2.1 shows the behavior of $C_{\theta_0}\left(\rho_\theta, J^{(S)}\right)$ (solid) and $C^{(1)}_{\theta_0}\left(\rho_\theta, J^{(S)}\right)$ with $z_0 = \frac{1}{4}$ (dashed) as functions of r. We see that Holevo bound $C_{\theta_0}\left(\rho_\theta, J^{(S)}\right)$ is much smaller than $C^{(1)}_{(0,r)}\left(\rho_\theta, J^{(S)}\right)$.

As in Example 2.16, we demonstrate that the Holevo bound is asymptotically achievable. Let

$$\Delta^{(n)}_i := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes k-1} \otimes L_i \otimes I^{\otimes n-k}, \qquad (i = 1, 2),$$

and let

$$X^{(n)}_j := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} I^{\otimes k-1} \otimes D_j \otimes I^{\otimes n-k}, \qquad (j = 1, 2, 3).$$

It then follows from the quantum central limit theorem that

$$\left(\begin{pmatrix} X^{(n)} \\ \Delta^{(n)} \end{pmatrix}, \rho_{\theta_0}^{\otimes n} \right) \underset{q}{\rightsquigarrow} N\left(0, \begin{pmatrix} \Sigma & \tau \\ \tau^* & J \end{pmatrix}\right).$$

Therefore, Corollary 2.11 shows that $\left(\left\{\rho_\theta^{\otimes n}\right\}, X^{(n)}\right)$ is jointly QLAN at θ_0, and that

$$(X^{(n)}, \rho^{\otimes n}_{\theta_0 + h/\sqrt{n}}) \underset{q}{\rightsquigarrow} N((\operatorname{Re}\tau)h, \Sigma)$$

for all $h \in \mathbb{R}^2$.

It should be noted that the off-diagonal block τ of the "quantum covariance" matrix is not a square matrix. This means that the derived quantum Gaussian shift model $\left\{N((\operatorname{Re}\tau)h, \Sigma);\ h \in \mathbb{R}^2\right\}$ forms a submanifold of the total quantum Gaussian shift model derived in Example 2.16, corresponding to a 2-dimensional linear subspace in the shift parameter space. Nevertheless, Theorem 2.12 asserts that there exist a sequence $M^{(n)}$ of estimators on the model $\left\{\rho^{\otimes n}_{\theta_0 + h/\sqrt{n}};\ h \in \mathbb{R}^3\right\}$ that is asymptotically unbiased and achieves the Holevo bound:

$$\lim_{n\to\infty} \operatorname{Tr} G V^{(n)}_h[M^{(n)}] = C_h\left(N((\operatorname{Re}\tau)h, \Sigma), G\right) = C_{\theta_0}(\rho_\theta, G)$$

for all h that belong to a dense subset of $\mathbb{R}^3$.

2.4.3 Translating estimation of h to estimation of θ

As we have seen in the previous subsections, our theory enables us to construct asymptotically optimal estimators of h in the local models indexed by the parameter $\theta_0 + h/\sqrt{n}$. In practice of course, θ_0 is unknown and hence estimation of h, with θ_0 known, is irrelevant. The actual sequence of measurements which we have constructed depends in all interesting cases on θ_0.

However, the results immediately inspire two-step (or adaptive) procedures, in which we first measure a small proportion of the quantum systems, in number n_1 say, using some standard measurement scheme, for instance separate particle quantum tomography. From these measurement outcomes we construct an initial estimate of θ, let us call it $\widetilde{\theta}$. We can now use our theory to compute the asymptotically optimal measurement scheme which corresponds to the situation $\theta_0 = \widetilde{\theta}$. We proceed to implement this measurement on the remaining quantum systems collectively, estimating h in the model $\theta = \widetilde{\theta} + h/\sqrt{n_2}$ where n_2 is the number of systems still available for the second stage.

What can we say about such a procedure? If $n_1/n \to \alpha > 0$ as $n \to \infty$ then we can expect that the initial estimate $\widetilde{\theta}$ is root n consistent. In smooth models, one would expect that in this case the final estimate $\widehat{\theta} = \widetilde{\theta} + \widehat{h}/\sqrt{n_2}$ would be asymptotically optimal *up to a factor* $1-\alpha$: its limiting variance will be a factor $(1-\alpha)^{-1}$ too large.

If however $n_1 \to \infty$ but $n_1/n \to \alpha = 0$ then one would expect this procedure to break down, unless the rate of growth of n_1 is very carefully chosen (and fast enough). On the other hand, instead of a direct two-step procedure, with the final estimate computed as $\widetilde{\theta} + \widehat{h}/\sqrt{n_2}$, one could be more careful in how the data obtained from the second stage measurement is used. Given the second step measurement, which results in an observed value $\widehat{h}$, one could write down the likelihood for h based on the given measurement and the initially specified model, and compute instead of the just mentioned one-step iterate, the actual maximum likelihood estimator of θ based on the second stage data. Such procedures have earlier been studied by Gill and Massar [13] and others, and shown in special cases to perform very well.

However, in general, the computational problem of even calculating the likelihood given data, measurement, and model, is challenging, due to the huge size of the Hilbert space of n copies of a finite dimensional quantum system.

2.5 Concluding remarks

We have developed a new theory of local asymptotic normality in a quantum regime based on a quantum extension of the log-likelihood ratio. This formulation is applicable to any model satisfying a mild smoothness condition, and is free from artificial setups such as the use of a special coordinate system and/or non-degeneracy of eigenvalues of the reference state. We also have proved asymptotic achievability of the Holevo bound for the local shift parameter on a dense subset of the parameter space.

There are of course many open questions left. Among others, it is not clear whether every sequence of statistics on a QLAN model can be realized on the limiting quantum Gaussian shift model. In classical statistics, such a problem has been solved affirmatively as the representation theorem, which asserts that, given a weakly convergent sequence $T^{(n)}$ of statistics on $\left\{p^{(n)}_{\theta_0+h/\sqrt{n}};\ h \in \mathbb{R}^d\right\}$, there exist a limiting statistics T on $\left\{N(h, J^{-1});\ h \in \mathbb{R}^d\right\}$ such that $T^{(n)} \overset{h}{\rightsquigarrow} T$. Representation theorem is useful in proving, for example, the non-existence of an asymptotically superefficient estimator (the converse part, as stated in Introduction). Moreover, the so-called convolution theorem and local asymptotic minimax theorem, which are the standard tools in discussing asymptotic lower bounds for estimation in LAN models, immediately follows [47]. Extending the representation theorem, convolution theorem, and local asymptotic minimax theorem to a quantum regime is an intriguing open problem. However it surely is possible to make some progress in this direction, as for instance the results of Gill and Guţă [11]. In that paper, the van Trees inequality was used to derive some results in a "poor man's" version of QLAN theory; see also [12].

It also remains to be seen whether our asymptotically optimal statistical procedures for the

local model with local parameter h can be translated into useful statistical procedures for the real world case in which θ_0 is unknown.

Appendices

Appendix 2.A Commutation operator and the Holevo bound

In the study of quantum statistics, Holevo [24] introduced useful mathematical tools called the square summable operators and the commutation operators associated with quantum states. Let $\mathcal{H}$ be a separable Hilbert space and let ρ be a density operator. We define a real Hilbert space $\mathcal{L}_h^2(\rho)$ associated with ρ by the completion of the set $\mathcal{B}_h(\mathcal{H})$ of bounded Hermitian operators with respect to the pre-inner product $\langle X, Y\rangle_\rho := \mathrm{Re}\,\mathrm{Tr}\,\rho XY$. Letting $\rho = \sum_j s_j|\psi_j\rangle\langle\psi_j|$ be the spectral representation, an element $X \in \mathcal{L}_h^2(\rho)$ can be regarded as an equivalence class of those Hermitian operators, called the square summable operators, which satisfy $\sum_j s_j\|X\psi_j\|^2 < \infty$ (so that $\psi_j \in \mathrm{Dom}(X)$ if $s_j \neq 0$) under the identification $X_1 \sim X_2$ if $X_1\psi_j = X_2\psi_j$ for $s_j \neq 0$. The space $\mathcal{L}_h^2(\rho)$ thus provides a convenient tool to cope with unbounded observables. Note that when $\mathcal{H}$ is finite dimensional, the setup is considerably simplified to be $\mathcal{L}_h^2(\rho) = \mathcal{B}_h(\mathcal{H})/\ker\rho$.

Let $\mathcal{L}^2(\rho)$ be the complexification of $\mathcal{L}_h^2(\rho)$, which is also regarded as the completion of $\mathcal{B}(\mathcal{H})$ with respect to the pre-inner product

$$\langle X, Y\rangle_\rho := \frac{1}{2}\mathrm{Tr}\,\rho(YX^* + X^*Y).$$

Thus $\mathcal{L}^2(\rho)$ is a complex Hilbert space with this inner product. Let us further introduce two sesquilinear forms on $\mathcal{B}(\mathcal{H})$ by

$$(X, Y)_\rho := \mathrm{Tr}\,\rho YX^*, \qquad [X, Y]_\rho := \frac{1}{2\sqrt{-1}}\mathrm{Tr}\,\rho(YX^* - X^*Y).$$

and extend them to $\mathcal{L}^2(\rho)$ by continuity.

The *commutation operator* $\mathcal{D}_\rho : \mathcal{L}^2(\rho) \to \mathcal{L}^2(\rho)$ with respect to ρ is defined by

$$[X, Y]_\rho = \langle X, \mathcal{D}_\rho Y\rangle_\rho,$$

which is formally represented by the operator equation

$$\mathcal{D}_\rho(X)\rho + \rho\mathcal{D}_\rho(X) = \sqrt{-1}\,(X\rho - \rho X).$$

(To be precise, Holevo's original definition is different from the above one by a factor of 2.) The operator $\mathcal{D}_\rho$ is a $\mathbb{C}$-linear bounded skew-adjoint operator. Moreover, since the forms $[\,\cdot\,,\,\cdot\,]_\rho$ and $\langle\,\cdot\,,\,\cdot\,\rangle_\rho$ are real on the real subspace $\mathcal{L}_h^2(\rho)$, this subspace is invariant under the operation of $\mathcal{D}_\rho$. Thus $\mathcal{D}_\rho$ can be regarded as an $\mathbb{R}$-linear bounded skew-adjoint operator when restricted to $\mathcal{L}_h^2(\rho)$ as $\mathcal{D}_\rho : \mathcal{L}_h^2(\rho) \to \mathcal{L}_h^2(\rho)$. When no confusion is likely to arise, we drop the subscript ρ of $\mathcal{D}_\rho$ and simply denote it as $\mathcal{D}$.

Let $\mathcal{S} = \{\rho_\theta\,;\, \theta \in \Theta \in \mathbb{R}^d\}$ be a quantum statistical model satisfying the conditions: 1) the parametrization $\theta \mapsto \rho_\theta$ is smooth and nondegenerate so that the derivatives $\{\partial\rho_\theta/\partial\theta^i\}_{1\le i\le d}$ exist in trace class and form a linearly independent set at each point $\theta \in \Theta$, and 2) there exists a constant c such that

$$\left|\frac{\partial}{\partial\theta^i}\,\mathrm{Tr}\,\rho_\theta X\right|^2 \le c\,\langle X, X\rangle_{\rho_\theta}$$

for all $X \in \mathcal{B}(\mathcal{H})$ and i. The second condition assures that the linear functionals $X \mapsto (\partial/\partial\theta^i)\mathrm{Tr}\,\rho_\theta X$ can be extended to continuous linear functionals on $\mathcal{L}^2(\rho_\theta)$. Given a quantum statistical model satisfying the above conditions, the *symmetric logarithmic derivative* (SLD) $L_{\theta,i}$ in the ith direction is defined as the operator in $\mathcal{L}^2(\rho_{\rho_\theta})$ satisfying

$$\frac{\partial}{\partial\theta^i}\mathrm{Tr}\,\rho_\theta X = \langle L_{\theta,i}, X\rangle_{\rho_\theta}.$$

It is easily verified that $L_{\theta,i} \in \mathcal{L}_h^2(\rho_\theta)$; so the definition is formally written as

$$\frac{\partial \rho_\theta}{\partial \theta^i} = \frac{1}{2}(L_{\theta,i}\,\rho_\theta + \rho_\theta\, L_{\theta,i}). \tag{2.20}$$

When no confusion occurs, we simply denote $L_{\theta,i}$ as L_i. Since L_i is a faithful operator representation of the tangent vector $\partial/\partial\theta^i$, we shall call the $\mathbb{R}$-linear space $\mathrm{span}_{\mathbb{R}}\{L_i\}_{i=1}^d$ the *SLD tangent space* of the model ρ_θ at θ. Incidentally the $d \times d$ real symmetric matrix $J_\theta := [\mathrm{Re}\,\mathrm{Tr}\,\rho_\theta L_i L_j]_{1\le i,j\le d}$ is called the *SLD Fisher information matrix* of the model $\mathcal{S}$ at θ.

An estimator $\hat{M}$ for the parameter θ of the model $\mathcal{S}$ is called *unbiased* if

$$E_\theta[\hat{M}] = \theta \tag{2.21}$$

for all $\theta \in \Theta$, where $E_\theta[\,\cdot\,]$ denotes the expectation with respect to ρ_θ. An estimator $\hat{M}$ is called *locally unbiased* at $\theta_0 \in \Theta$ if the condition (2.21) is satisfied around $\theta = \theta_0$ up to the first order of the Taylor expansion. It is well known that an estimator $\hat{M}$ that is locally unbiased at θ_0 satisfies the quantum (SLD) Cramér-Rao inequality, $V_{\theta_0}[\hat{M}] \ge J_{\theta_0}^{-1}$, where $V_{\theta_0}[\,\cdot\,]$ denotes the covariance matrix with respect to ρ_{θ_0}. The lower bound $J_{\theta_0}^{-1}$ cannot be attained in general due to the non-commutativity of the SLDs. Because of this fact, we often switch the problem to minimizing the weighted sum of covariances, $\mathrm{Tr}\,GV_{\theta_0}[\hat{M}]$, given a $d \times d$ real positive definite matrix G. It is known that this quantity also has a variety of Cramér-Rao type lower bounds [24]:

$$\mathrm{Tr}\,GV_{\theta_0}[\hat{M}] \ge C_{\theta_0}\left(\rho_\theta, G\right).$$

Among others, we concentrate our attention to the *Holevo bound* [24]:

$$\begin{aligned} C_{\theta_0}\left(\rho_\theta, G\right) \;:=\; & \min_{V,B}\{\mathrm{Tr}\,GV\;;\; V \text{ is a real matrix such that } V \ge Z(B),\, Z_{ij}(B) = \mathrm{Tr}\,\rho_{\theta_0} B_j B_i, \\ & B_1, \ldots, B_d \text{ are Hermitian operators on } \mathcal{H} \text{ such that } \mathrm{Re}\,\mathrm{Tr}\,\rho_{\theta_0} L_i B_j = \delta_{ij}\}. \end{aligned} \tag{2.22}$$

The minimization problem over V is explicitly solved, to obtain

$$\begin{aligned} C_{\theta_0}\left(\rho_\theta, G\right) \;=\; & \min_{B}\{\mathrm{Tr}\,GZ(B) + \mathrm{Tr}\left|\sqrt{G}\,\mathrm{Im}\,Z(B)\sqrt{G}\right|\;;\; Z_{ij}(B) = \mathrm{Tr}\,\rho_{\theta_0} B_j B_i, \\ & B_1, \ldots, B_d \text{ are Hermitian operators on } \mathcal{H} \text{ such that } \mathrm{Re}\,\mathrm{Tr}\,\rho_{\theta_0} L_i B_j = \delta_{ij}\}. \end{aligned}$$

Our aim here is to derive a further concise expression for it in terms of the minimal $\mathcal{D}$ invariant extension of the SLD tangent space.

Theorem 2.19. *Given a quantum statistical model $\{\rho_\theta\,;\,\theta \in \Theta \subset \mathbb{R}^d\}$ on $\mathcal{H}$, let $\mathcal{T}$ be the minimal $\mathcal{D}$ invariant extension of the SLD tangent space $\mathrm{span}_{\mathbb{R}}\{L_i\}_{i=1}^d$ of the model at $\theta = \theta_0$, and let $\{D_j\}_{j=1}^r$ be a basis of $\mathcal{T}$. The Holevo bound defined by (2.22) is rewritten as*

$$\begin{aligned} C_{\theta_0}\left(\rho_\theta, G\right) \;=\; & \min_{F}\{\mathrm{Tr}\,GZ + \mathrm{Tr}\left|\sqrt{G}\,\mathrm{Im}\,Z\sqrt{G}\right|\;;\; Z = {}^tF\Sigma F, \\ & F \text{ is an } r \times d \text{ real matrix satisfying } {}^tF\,\mathrm{Re}\,(\tau) = I\}, \end{aligned} \tag{2.23}$$

where Σ and τ are $r \times r$ and $r \times d$ complex matrices whose (i,j)th entries are given by $\Sigma_{ij} = \mathrm{Tr}\,\rho_{\theta_0} D_j D_i$ and $\tau_{ij} = \mathrm{Tr}\,\rho_{\theta_0} L_j D_i$.

Proof. Let $\mathcal{T}^\perp$ be the orthogonal complement of $\mathcal{T}$ in $\mathcal{L}_h^2(\rho_{\theta_0})$ with respect to the inner product $\langle\cdot,\cdot\rangle_{\rho_{\theta_0}}$, and let $\mathcal{P} : \mathcal{L}_h^2(\rho_{\theta_0}) \to \mathcal{T}$ and $\mathcal{P}^\perp : \mathcal{L}_h^2(\rho_{\theta_0}) \to \mathcal{T}^\perp$ be the projections associated with the decomposition $\mathcal{L}_h^2(\rho_{\theta_0}) = \mathcal{T} \oplus \mathcal{T}^\perp$. Note that if $X \in \mathcal{T}^\perp$ and $Y \in \mathcal{T}$, then

$$(X, Y)_{\rho_{\theta_0}} = \langle X, Y\rangle_{\rho_{\theta_0}} + \sqrt{-1}\,\langle X, \mathcal{D}Y\rangle_{\rho_{\theta_0}} = 0.$$

We show that the operators $\{B_j\}_{j=1}^d$ that achieve the minimum in (2.22) can be taken from $\mathcal{T}$. Let $\{B_j\}_{j=1}^d \subset \mathcal{L}_h^2(\rho_{\theta_0})$ satisfies the local unbiasedness condition $\mathrm{Re}\,\mathrm{Tr}\,\rho_{\theta_0} L_i B_j = \delta_{ij}$, which is rewritten as

$$\langle L_i, B_j\rangle_{\rho_{\theta_0}} = \delta_{ij}.$$

Then $\{\mathcal{P}(B_j)\}_{j=1}^d$ also satisfies the local unbiasedness

$$\langle L_i, \mathcal{P}(B_j)\rangle_{\rho_{\theta_0}} = \langle L_i, B_j\rangle_{\rho_{\theta_0}} = \delta_{ij}.$$

Further,

$$\begin{aligned} Z_{ij}(B) &= (B_i, B_j)_{\rho_{\theta_0}} = \big(\mathcal{P}(B_i) + \mathcal{P}^\perp(B_i), \mathcal{P}(B_j) + \mathcal{P}^\perp(B_j)\big)_{\rho_{\theta_0}} \\ &= (\mathcal{P}(B_i), \mathcal{P}(B_j))_{\rho_{\theta_0}} + \big(\mathcal{P}^\perp(B_i), \mathcal{P}^\perp(B_j)\big)_{\rho_{\theta_0}} = Z_{ij}(\mathcal{P}(B)) + Z_{ij}(\mathcal{P}^\perp(B)). \end{aligned}$$

Since $Z(\,\cdot\,)$ is a Gram matrix and is positive semidefinite, this decomposition implies that $Z(B) \geq Z(\mathcal{P}(B))$. Thus the observables B that minimize (2.22) can be taken from $\mathcal{T}$.

Let $B_j \in \mathcal{T}$ be expanded as $B_j = F_j^k D_k$, where F is an $r \times d$ real matrix. Then the local unbiasedness condition is rewritten as

$$\langle L_i, B_j\rangle_{\rho_{\theta_0}} = F_j^k \langle L_i, D_k\rangle_{\rho_{\theta_0}} = \delta_{ij},$$

or in a matrix form,

$${}^t F\,(\mathrm{Re}\,\tau) = I.$$

Further, the Gram matrix $Z(B)$ is rewritten as

$$Z_{ij}(B) = (B_i, B_j)_{\rho_{\theta_0}} = F_i^k F_j^\ell\,(D_k, D_\ell)_{\rho_{\theta_0}},$$

or,

$$Z(B) = {}^t F \Sigma F.$$

This proves the claim. □

When the SLD tangent space itself is $\mathcal{D}$ invariant, the Holevo bound can be represented in terms of the RLD Fisher information matrix as follows.

Corollary 2.20. *Let $\{\rho_\theta\,;\,\theta \in \Theta \subset \mathbb{R}^d\}$ be a quantum statistical model, and let L_i $(1 \leq i \leq d)$ be the SLDs at θ_0. If the SLD tangent space $\mathrm{span}_{\mathbb{R}}\{L_i\}_{i=1}^d$ at θ_0 is $\mathcal{D}$ invariant, then*

$$C_{\theta_0}(\rho_\theta, G) = \mathrm{Tr}\, G (J^{(R)})^{-1} + \mathrm{Tr}\,\left|\sqrt{G}\,\mathrm{Im}\,(J^{(R)})^{-1}\sqrt{G}\right|,$$

where $(J^{(R)})^{-1} := (\mathrm{Re}\,J)^{-1}\,J\,(\mathrm{Re}\,J)^{-1}$ with $J_{ij} = \mathrm{Tr}\,\rho_{\theta_0} L_j L_i$.

Proof. Let us set $D_i := L_i$ for $1 \leq i \leq d$ in Theorem 2.19. Then $\Sigma = \tau$, and the local unbiasedness condition ${}^t F\,(\mathrm{Re}\,\tau) = I$ has a unique solution $F = (\mathrm{Re}\,\Sigma)^{-1}$, whereby $Z = (\mathrm{Re}\,J)^{-1}\,J\,(\mathrm{Re}\,J)^{-1}$. □

Note that RLDs may not exist if the model is degenerate (i.e., non-faithful). This means that $J^{(R)}$ may not be well-defined for such a model. Nevertheless we use the notation $(J^{(R)})^{-1}$ even for a degenerate model, and call it the inverse of the RLD Fisher information matrix, as long as the SLD tangent space is $\mathcal{D}$ invariant. For an idea behind this nomenclature, consult [7].

Finally, we show that the Holevo bound for the nth i.i.d. extension model is precisely $\frac{1}{n}$ times that for the base model.

Corollary 2.21. *Given a quantum statistical model $\mathcal{S} = \{\rho_\theta\,;\,\theta \in \Theta \subset \mathbb{R}^d\}$, let $\mathcal{S}^{(n)} = \{\rho_\theta^{\otimes n}\,;\,\theta \in \Theta \subset \mathbb{R}^d\}$ be the nth i.i.d. extension model. Then*

$$C_{\theta_0}\left(\rho_\theta^{\otimes n}, G\right) = \frac{1}{n}\,C_{\theta_0}(\rho_\theta, G).$$

Proof. Let us distinguish quantities that belong to models of different extension by specifying the degree n of extension in the superscript. Letting $\{L_i\}_{i=1}^d$ and $\{D_j\}_{j=1}^r$ be SLDs and a basis of $\mathcal{T}$ in Theorem 2.19, the corresponding quantities for $\mathcal{S}^{(n)}$ are given by

$$L_i^{(n)} = \sum_{k=1}^n I^{\otimes k-1} \otimes L_i \otimes I^{\otimes n-k}$$

and

$$D_j^{(n)} = \sum_{k=1}^n I^{\otimes k-1} \otimes D_j \otimes I^{\otimes n-k}.$$

Thus

$$\Sigma^{(n)} = n\Sigma^{(1)}, \quad \tau^{(n)} = n\tau^{(1)}, \quad F^{(n)} = \frac{1}{n}F^{(1)},$$

so that

$$Z^{(n)} = {}^tF^{(n)}\Sigma^{(n)}F^{(n)} = \frac{1}{n}Z^{(1)},$$

and

$$C_{\theta_0}\left(\rho_\theta^{\otimes n}, G\right) = \frac{1}{n} C_{\theta_0}\left(\rho_\theta, G\right)$$

doe to Theorem 2.19. □

Appendix 2.B Estimation of quantum Gaussian shift model

In this section, we briefly overview the estimation theory for a quantum Gaussian shift model. For a mathematically rigorous treatment, consult [24].

Lemma 2.22. *Let $(X, \phi_h) \sim N(h, J)$, where J is a $d \times d$ positive semidefinite complex matrix. Then*

$$\phi_h(X_i) = h_i \tag{2.24}$$

and

$$\phi_h((X_j - h_j)(X_i - h_i)) = J_{ij} \tag{2.25}$$

hold.

Proof. Letting $U(\xi) := \mathrm{e}^{\sqrt{-1}\xi^i X_i}$,

$$\begin{aligned}
\phi_h(U(\xi)) &= 1 + \sqrt{-1}\phi_h(\xi^i X_i) - \frac{1}{2}\phi_h(\left(\xi^i X_i\right)^2) + o(\xi^2) \\
&= 1 + \sqrt{-1}\phi_h(X_i)\xi^i - \frac{1}{2}\phi_h(X_iX_j)\xi^i\xi^j + o(\xi^2) \\
&= 1 + \sqrt{-1}\phi_h(X_i)\xi^i - \frac{1}{2}\phi_h(X_i \circ X_j)\xi^i\xi^j + o(\xi^2),
\end{aligned}$$

where $X_i \circ X_j = \frac{1}{2}\left(X_iX_j + X_JX_i\right)$. Further, letting $V = \mathrm{Re}\, J$ and $S = \mathrm{Im}\, J$,

$$\begin{aligned}
\mathrm{e}^{\sqrt{-1}\xi^i h_i - \frac{1}{2}V_{ij}\xi^i\xi^j} &= 1 + \left(\sqrt{-1}\xi^i h_i - \frac{1}{2}V_{ij}\xi^i\xi^j\right) + \frac{1}{2}\left(\sqrt{-1}\xi^i h_i - \frac{1}{2}V_{ij}\xi^i\xi^j\right)^2 + o(\xi^2) \\
&= 1 + \sqrt{-1}\xi^i h_i - \frac{1}{2}\left(V_{ij} + h_ih_j\right)\xi^i\xi^j + o(\xi^2).
\end{aligned}$$

A comparison immediately leads to (2.24) and the identity $\phi_h(X_i \circ X_j) = V_{ij} + h_ih_j$. Thus

$$\begin{aligned}
\phi_h((X_j - h_j)(X_i - h_i)) &= \phi_h(X_jX_i - h_jX_i - h_iX_j + h_ih_j) \\
&= \phi_h(X_jX_i) - h_ih_j \\
&= \phi_h\left(X_i \circ X_j - \frac{1}{2}[X_i, X_j]\right) - h_ih_j = J_{ij}.
\end{aligned}$$

□

In what follows, we treat the quantum Gaussian shift model $\left\{N(\tau h, \Sigma)\,;\ h \in \mathbb{R}^d\right\}$ on $\mathrm{CCR}\,(\mathrm{Im}\,\Sigma)$, where Σ is an $r \times r$ complex matrix such that $\Sigma \geq 0$ and $\mathrm{Re}\,\Sigma > 0$, and τ is an $r \times d$ real matrix with $d \leq r$ such that $\mathrm{rank}\,\tau = d$. Let $X = (X_1, \ldots, X_r)$ be the basic canonical observables of $\mathrm{CCR}\,(\mathrm{Im}\,\Sigma)$, and $(X, \phi_h) \sim N(\tau h, \Sigma)$.

Lemma 2.23. *Let $U(\xi) := \mathrm{e}^{\sqrt{-1}\xi^i X_i}$. The SLD L_i $(1 \leq i \leq d)$ at h defined by*

$$\frac{\partial}{\partial h_k}\phi_h(U(\xi)) = \frac{1}{2}\phi_h\left(U(\xi)L_k + L_k U(\xi)\right) \tag{2.26}$$

is given by

$$L_k = \sum_{\ell=1}^{r}\left[\left(\mathrm{Re}\,\Sigma\right)^{-1}\tau\right]_{\ell k}(X_\ell - (\tau h)_\ell I). \tag{2.27}$$

Proof. In this proof we lift Einstein's summation convention. Let $V = \mathrm{Re}\,\Sigma$ and $S = \mathrm{Im}\,\Sigma$, and fix a $k \in \{1, \ldots, d\}$ arbitrarily. Due to the Baker-Hausdorff formula,

$$U(\xi) = \mathrm{e}^{\sqrt{-1}\sum_{i=1}^r \xi^i X_i} = \exp\left(-\sqrt{-1}\sum_{i=1}^{r} S_{ki}\xi^k\xi^i\right)\exp\left(\sqrt{-1}\xi^k X_k\right)\exp\left(\sqrt{-1}\sum_{i\neq k}\xi^i X_i\right).$$

By differentiating in ξ^k, we have

$$\frac{\partial}{\partial \xi^k}U(\xi) = -\sqrt{-1}\left(\sum_{i=1}^{r} S_{ki}\xi^i - X_k\right)U(\xi).$$

Thus

$$\begin{aligned}
\phi_h((X_k - (\tau h)_k I)U(\xi)) &= \phi_h\left(\left(\sum_{i=1}^{r} S_{ki}\xi^i - \sqrt{-1}\frac{\partial}{\partial \xi^k} - (\tau h)_k I\right)U(\xi)\right)\\
&= \left(\sum_{i=1}^{r} S_{ki}\xi^i - \sqrt{-1}\frac{\partial}{\partial \xi^k} - (\tau h)_k\right)\phi_h(U(\xi))\\
&= \left(\sum_{i=1}^{r} S_{ki}\xi^i - \sqrt{-1}\frac{\partial}{\partial \xi^k} - (\tau h)_k\right)\mathrm{e}^{\sqrt{-1}\,{}^t\xi\tau h - \frac{1}{2}\,{}^t\xi V\xi}\\
&= \left(\sum_{i=1}^{r} S_{ki}\xi^i - (\tau h)_k\right)\phi_h(U(\xi)) - \sqrt{-1}\left(\sqrt{-1}(\tau h)_k - (V\xi)_k\right)\phi_h(U(\xi))\\
&= \left(S\xi + \sqrt{-1}V\xi\right)_k\phi_h(U(\xi))\\
&= \sqrt{-1}\left(\bar{J}\xi\right)_k\phi_h(U(\xi)).
\end{aligned} \tag{2.28}$$

Similarly, we obtain

$$\phi_h(U(\xi)(X_k - (\tau h)_k I)) = \sqrt{-1}\left(J\xi\right)_k\phi_h(U(\xi)). \tag{2.29}$$

By combining (2.28) and (2.29),

$$\phi_h\left(\,(X_k - (\tau h)_k I)U(\xi) + U(\xi)(X_k - (\tau h)_k I)\,\right) = 2\sqrt{-1}\left(V\xi\right)_k\phi_h(U(\xi)). \tag{2.30}$$

On the other hand, by a direct calculation

$$\frac{\partial}{\partial h_k}\phi_h(U(\xi)) = \frac{\partial}{\partial h_k}\mathrm{e}^{\sqrt{-1}\,{}^t\xi\tau h - \frac{1}{2}\,{}^t\xi V\xi} = \sqrt{-1}({}^t\xi\tau)_k\phi_h(U(\xi)). \tag{2.31}$$

A comparison between (2.30) and (2.31) yields

$$L_k = \sum_{\ell=1}^{r}\left[V^{-1}\tau\right]_{\ell k}(X_\ell - (\tau h)_\ell I).$$

□

Let $\tilde{L}_k := X_k - (\tau h)_k I$. It follows from (2.28) and (2.29) that $\mathcal{D}_{\phi_h}(\tilde{L}_i) = \sum_{i=1}^r (V^{-1}S)_{ki}\tilde{L}_k$, where $\mathcal{D}_{\phi_h}$ is the commutation operator with respect to ϕ_h defined by

$$\phi_h\left(U(\xi)\mathcal{D}_{\phi_h}(X) + \mathcal{D}_{\phi_h}(X)U(\xi)\right) = \sqrt{-1}\phi_h\left(U(\xi)X - XU(\xi)\right).$$

This means $\mathcal{T} = \mathrm{span}\left\{\tilde{L}_k\right\}_{k=1}^r$ is $\mathcal{D}_{\phi_h}$ invariant. Further, we can check from (2.27) that $\mathrm{span}\{L_i\}_{i=1}^d \subset \mathcal{T}$ and

$$\phi_h(\tilde{L}_j\tilde{L}_i) = \Sigma_{ij} \tag{2.32}$$

and

$$\mathrm{Re}\,\phi_h(L_j\tilde{L}_i) = \tau_{ij}. \tag{2.33}$$

These relations play a fundamental role in connecting a general quantum statistical model $\mathcal{S} = \left\{\rho_\theta\,;\ \theta \in \Theta \subset \mathbb{R}^d\right\}$ on $\mathcal{H}$ with a quantum Gaussian shift model $\mathcal{G} = \left\{N(\tau h, \Sigma)\,;\ h \in \mathbb{R}^d\right\}$ as follows. Let $\{L_i^{\mathcal{S}}\}_{i=1}^d$ be the SLDs of the model $\mathcal{S}$ at $\theta = \theta_0$, and let $\mathcal{T}^{\mathcal{S}}$ the minimal $\mathcal{D}^{\mathcal{S}}$ invariant extension of the SLD tangent space $\mathrm{span}\{L_i^{\mathcal{S}}\}_{i=1}^d$. Further let $\{D_j^{\mathcal{S}}\}_{j=1}^r$ be a basis of $\mathcal{T}^{\mathcal{S}}$ and let Σ and τ are $r \times r$ and $r \times d$ matrices whose (i,j)th entries are given by $\Sigma_{ij} = \mathrm{Tr}\,\rho_{\theta_0}D_jD_i$ and $\tau_{ij} = \mathrm{Re}\,\mathrm{Tr}\,\rho_{\theta_0}L_jD_i$. Based on those information, we introduce a quantum Gaussian shift model $\mathcal{G} = \left\{N(\tau h, \Sigma)\,;\ h \in \mathbb{R}^d\right\}$ on $\mathrm{CCR}\,(\mathrm{Im}\,\Sigma)$, which exhibits relations (2.32) and (2.33). Recall that the Holevo bound of a quantum statistical model is completely determined by the information Σ and τ (Theorem 2.19). We thus obtain the following important consequence.

Corollary 2.24. *The Holevo bound $C_{\theta_0}(\rho_\theta, G)$ for the model $\mathcal{S}$ at $\theta = \theta_0$ is identical to the Holevo bound $C_h\left(N(\tau h, \Sigma), G\right)$ for the Gaussian shift model $\mathcal{G}$.*

As to the achievability of the Holevo bound $C_h\left(N(\tau h, \Sigma), G\right)$ for the Gaussian shift model $\mathcal{G}$, we have the following.

Theorem 2.25. *Given a weight $G > 0$, there exist an unbiased estimator $\hat{M}$ that achieves the Holevo bound for the model $\{N(\tau h, \Sigma)\,;\, h \in \mathbb{R}^d\}$, i.e.,*

$$\mathrm{Tr}\,GV_h[\hat{M}] = C_h\left(N(\tau h, \Sigma), G\right).$$

Proof. Let F be the matrix that achieve the minimum of (2.23) for the model $\{N(\tau h, \Sigma)\}_h$, and let $Z = {}^tF\Sigma F$. Further, let $\tilde{V} = \mathrm{Re}\,Z$, $\tilde{S} = \mathrm{Im}\,Z$. $\hat{V} = \sqrt{G^{-1}}\left|\sqrt{G}\,\mathrm{Im}\,Z\sqrt{G}\right|\sqrt{G^{-1}}$, and $\hat{Z} = \hat{V} - \sqrt{-1}\tilde{S}$. We introduce an ancillary quantum Gaussian state $(Y, \psi) \sim N(0, \hat{Z})$ on another $\mathrm{CCR}\left(-\tilde{S}\right)$, and a set of canonical observables

$$\overline{X}_i := \tilde{X}_i \otimes I + I \otimes Y_i \quad (1 \le i \le d),$$

on $\mathrm{CCR}\left(\tilde{S}\right) \otimes \mathrm{CCR}\left(-\tilde{S}\right)$, where $\tilde{X}_i = F_i^k X_k$. It is important to notice that the CCR subalgebra $\mathcal{A}[\overline{X}]$ generated by $\{\overline{X}_i\}_{1\le i\le d}$ is a commutative one because

$$\frac{\sqrt{-1}}{2}[\overline{X}_i, \overline{X}_j] = \tilde{S}_{ij} - \tilde{S}_{ij} = 0$$

for $1 \le i, j \le d$. Moreover

$$(\phi_h \otimes \psi)(\mathrm{e}^{\sqrt{-1}\xi^i\overline{X}_i}) = \left[\phi_h\left(\mathrm{e}^{\sqrt{-1}\xi^i\tilde{X}_i}\right)\right]\left[\psi\left(\mathrm{e}^{\sqrt{-1}\xi^iY_i}\right)\right] = \mathrm{e}^{\sqrt{-1}\xi^ih_i - \frac{1}{2}\xi^i\xi^j(\tilde{V}+\hat{V})_{ij}}.$$

This means that the observables $\overline{X}_i$ $(1 \le i \le d)$ follow the classical Gaussian distribution $N(h, \tilde{V} + \hat{V})$. In particular,

$$E_h[\overline{X}] = h$$

for all $h \in \mathbb{R}^d$, and

$$\mathrm{Tr}\,GV_h[\overline{X}] = \mathrm{Tr}\,G(\tilde{V} + \hat{V}) = C_h\left(N(\tau h, \Sigma), G\right).$$

The claim was verified. □

Appendix 2.C Estimation theory for pure state models

Lemma 2.26. *Let ρ be a pure state and $A_1, \ldots, A_d$ observables on a finite dimensional Hilbert space $\mathcal{H}$. If $J_{ij} := \mathrm{Tr}\,\rho A_j A_i$ are all real for $1 \le i, j \le d$, there exist observables $K_1, \ldots, K_d$ such that*

$$[A_i + K_i, A_j + K_j] = 0,$$

for $1 \le i, j \le d$ and

$$K_i \rho = 0$$

for $1 \le i \le d$.

Proof. Let $\rho := |\psi\rangle\langle\psi|$, and let $|l_i\rangle := A_i|\psi\rangle$ for $1 \le i \le d$. Because $\langle\psi|l_i\rangle$ and $\langle l_i|l_j\rangle$ $(= J_{ji})$ are all real, there exist a CONS $\{|e_k\rangle\}_{k=1}^{\dim\mathcal{H}}$ of $\mathcal{H}$ such that $\langle e_k|\psi\rangle$ and $\langle e_k|l_i\rangle$ are all real, and that $\langle e_k|\psi\rangle \neq 0$ for all k. Let

$$\tilde{A}_i := \sum_{k=1}^{\dim\mathcal{H}} \frac{\langle e_k|l_i\rangle}{\langle e_k|\psi\rangle} |e_k\rangle\langle e_k|,$$

and $K_i := \tilde{A}_i - A_i$. Obviously $[A_i + K_i, A_j + K_j] = [\tilde{A}_i, \tilde{A}_j] = 0$, and

$$K_i|\psi\rangle = \left(\tilde{A}_i - A_i\right)|\psi\rangle = |l_i\rangle - |l_i\rangle = 0.$$

This means $K_i\rho = 0$. □

Theorem 2.27. *Let $\left\{\rho_\theta\,;\ \theta \in \Theta \subset \mathbb{R}^d\right\}$ be a quantum statistical model comprising pure states on a finite dimensional Hilbert space $\mathcal{H}$, and let $C_{\theta_0}(\rho_\theta, G)$ be the Holevo bound at $\theta_0 \in \Theta$ for a given weight $G > 0$. There exist a locally unbiased estimator $\hat{M}$ at $\theta_0 \in \Theta$ such that $\mathrm{Tr}\,GV[\hat{M}] = C_{\theta_0}(\rho_\theta, G)$.*

Proof. Let $\mathcal{T}$ be the minimal $\mathcal{D}$ invariant extension of the SLD tangent space $\mathrm{span}\,\{L_i\}_{i=1}^d$ of the model $\{\rho_\theta\}$ at $\theta = \theta_0$, i.e., containing all the SLDs $\{L_i\}_{i=1}^d$ of $\{\rho_\theta\}$ at θ_0, let $\{D_j\}_{j=1}^r$ be a basis of $\mathcal{T}$. Let Σ, τ be $r \times r$, $r \times d$ complex matrices defined by $\Sigma_{ij} = \mathrm{Tr}\,\rho_{\theta_0} D_j D_i$, $\tau_{ij} = \mathrm{Tr}\,\rho_{\theta_0} L_j D_i$. According to Theorem 2.19, the Holevo bound for a weight $G > 0$ can be expressed

$$\begin{aligned} C_{\theta_0}(\rho_\theta, G) \quad = \quad & \min_F \{\mathrm{Tr}\,GZ + \mathrm{Tr}\left|\sqrt{G}\,\mathrm{Im}\,Z\sqrt{G}\right|\ ;\ Z = {}^tF\Sigma F, \\ & F \text{ is an } r \times d \text{ real matrix satisfying } {}^tF\,\mathrm{Re}\,(\tau) = I\}. \end{aligned} \tag{2.34}$$

Let F be the matrix that attains the minimum in (2.34), and let $Z := {}^tF\Sigma F$, $\tilde{V} := \mathrm{Re}\,Z$, $\tilde{S} := \mathrm{Im}\,Z$, $\hat{V} = \sqrt{G^{-1}}\left|\sqrt{G}\,\mathrm{Im}\,Z\sqrt{G}\right|\sqrt{G^{-1}}$, and $\hat{Z} = \hat{V} - \sqrt{-1}\tilde{S}$. Lemma 2.13 assures that there exist a Hilbert space $\hat{\mathcal{H}}$ and a pure state σ and observables B_i $(1 \le i \le d)$ on $\hat{\mathcal{H}}$ such that $\mathrm{Tr}\,\sigma B_i = 0$ and $\mathrm{Tr}\,\sigma B_j B_i = \hat{Z}_{ij}$. Further, let

$$\overline{X}_i := \tilde{X}_i \otimes \hat{I} + I \otimes B_i \quad (1 \le i \le d),$$

where $\tilde{X}_i := F_i^k D_k$ $(1 \le i \le d)$, and $\hat{I}$ is the identity matrix on $\hat{\mathcal{H}}$. It then follows that

$$\mathrm{Tr}\,(\rho_{\theta_0} \otimes \sigma)\,\overline{X}_j\overline{X}_i = \left(\tilde{V} + \hat{V}\right)_{ij}. \tag{2.35}$$

According to Lemma 2.26, there exist observables $K_1, \ldots, K_d$ on $\mathcal{H} \otimes \hat{\mathcal{H}}$ such that $[\overline{X}_i + K_i, \overline{X}_j + K_j] = 0$ and $K_i(\rho_{\theta_0} \otimes \sigma) = 0$. Let $\hat{T}_i := \theta_0^i I \otimes \hat{I} + \left(\overline{X}_i + K_i\right)$. Then $\hat{T}_1, \ldots, \hat{T}_d$ are simultaneously measurable, and satisfy the local unbiasedness condition:

$$\mathrm{Tr}\,(\rho_{\theta_0} \otimes \sigma)\,\hat{T}_j = \theta_0^j$$

and

$$\begin{aligned}\mathrm{Tr}\ (\partial_i\rho_{\theta_0}\otimes\sigma)\hat{T}_j &= \mathrm{Tr}\,\partial_i\rho_{\theta_0}\tilde{X}_j \\ &= F_j^k\mathrm{Tr}\,\partial_i\rho_{\theta_0}D_k \\ &= F_j^k\mathrm{Re}\,\mathrm{Tr}\,\rho_{\theta_0}L_iD_k \\ &= \{F\,(\mathrm{Re}\,\tau)\}_{ji} = \delta_{ij}.\end{aligned}$$

Further

$$V_{\theta_0}[\hat{T}]_{ij} = \mathrm{Tr}\ (\rho_{\theta_0}\otimes\sigma)\left(\overline{X}_i + K_i\right)\left(\overline{X}_i + K_i\right) = \left(\tilde{V} + \hat{V}\right)_{ij}.$$

This completes the proof. □

Appendix 2.D Quantum central limit theorem

Jaksić, Pautrat, and Pillet [28] proved the following strong version of a quantum central limit theorem.

Proposition 2.28. *Given a sequence $\mathcal{H}^{(n)}$ of Hilbert space, let $\rho^{(n)}$ and $A^{(n)} = (A_1^{(n)}, \ldots, A_d^{(n)})$ be a state and a list of observables on $\mathcal{H}^{(n)}$ that enjoy the quantum central limit theorem in the sense of convergence of the quasi-characteristic function:*

$$\left(A^{(n)}, \rho^{(n)}\right) \underset{q}{\rightsquigarrow} N(h, J) \sim (X, \phi),$$

where J is a $d \times d$ positive semidefinite matrix. Then for any bounded continuous functions $f_1, \ldots, f_m$ and a noncommutative polynomial P, it follows that

$$\lim_{n\to\infty} \mathrm{Tr}\,\rho^{(n)} P\left(\overrightarrow{f(A^{(n)})}\right) = \phi\left(P\left(\overrightarrow{f(X)}\right)\right),$$

where $\overrightarrow{f(B)} := (f_1(B_1), \ldots, f_1(B_d), \ldots, f_m(B_1), \ldots, f_m(B_d))$ for a given list $B = (B_1, \ldots, B_d)$ of observables, and $P\left(\overrightarrow{f(B)}\right) := P(f_1(B_1), \ldots, f_1(B_d), \ldots, f_m(B_1), \ldots, f_m(B_d))$.

Proposition 2.28 is strong enough to prove the following, which is essential in constructing a sequence of POVMs that asymptotically achieves the Holevo bound (Theorem 2.12).

Corollary 2.29. *Under the same assumption as in Proposition 2.28, for any bounded continuous functions $g, f_1, \ldots, f_m$, and noncommutative polynomials P, Q, with P being Hermitian operator-valued, it follows that*

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Tr}\,\rho^{(n)} g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) \\ = \phi\left(g\left(P\left(\overrightarrow{f(X)}\right)\right) Q\left(\overrightarrow{f(X)}\right) g\left(P\left(\overrightarrow{f(X)}\right)\right)\right).\end{aligned}$$

Proof. Let $l := \max_{1\le i\le m}\sup_x |f_i(x)|$. There exist $l_P > 0$ and $l_Q > 0$ such that $l_P > \left\|P(\overrightarrow{B})\right\|$ and $l_Q > \left\|Q(\overrightarrow{B})\right\|$ for any list $\overrightarrow{B} = (B_1, \ldots, B_{dm})$ of observables such that $\|B_i\| \le l$. Let $l_g := \sup\{|g(x)|\ ;\ x \in [-l_P, l_P]\}$. There exist a sequence $R^{(k)}(x)$ of polynomials that uniformly converges to $g(x)$ on $[-l_P, l_P]$.

Let

$$a_{kn} := \mathrm{Tr}\,\rho^{(n)} R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right),$$

and let

$$a_n := \mathrm{Tr}\,\rho^{(n)} g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right).$$

We show that a_{kn} uniformly converges to a_n as $k \to \infty$. In fact, letting $l_R := \sup\left\{R^{(k)}(x)\,;\ k \in \mathbb{N}, x \in [-l_P, l_P]\right\}$,

$$
\begin{aligned}
&\sup_{n\in\mathbb{N}} |a_n - a_{kn}| \\
&\quad = \sup_{n\in\mathbb{N}} \Bigg| \operatorname{Tr} \rho^{(n)} g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) \\
&\quad = -\operatorname{Tr} \rho^{(n)} R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) \Bigg| \\
&\quad \le \sup_{n\in\mathbb{N}} \left| \operatorname{Tr} \rho^{(n)} g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) \left[g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) - R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right)\right]\right| \\
&\qquad + \sup_{n\in\mathbb{N}} \left| \operatorname{Tr} \rho^{(n)} \left[g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) - R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right)\right] Q\left(\overrightarrow{f(A^{(n)})}\right) R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right)\right| \\
&\quad \le l_g l_Q \sup_{n\in\mathbb{N}} \left\| g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) - R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right)\right\| \\
&\qquad + l_Q l_R \sup_{n\in\mathbb{N}} \left\| g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) - R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right)\right\| \\
&\quad \le l_Q(l_g + l_R) \sup_{x\in[-l_P,l_P]} \left| g(x) - R^{(k)}(x)\right|,
\end{aligned}
$$

which converges to zero as $k \to \infty$.

The uniform convergence $a_{kn} \rightrightarrows a_n$ as well as the existence of $\lim_{k\to\infty}\lim_{n\to\infty} a_{kn}$, which follows from Proposition 2.28, ensure that

$$
\begin{aligned}
&\lim_{n\to\infty} \operatorname{Tr} \rho^{(n)} g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) g\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) \\
&\quad = \lim_{n\to\infty}\lim_{k\to\infty} \operatorname{Tr} \rho^{(n)} R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) Q\left(\overrightarrow{f(A^{(n)})}\right) R^{(k)}\left(P\left(\overrightarrow{f(A^{(n)})}\right)\right) \\
&\quad = \lim_{n\to\infty}\lim_{k\to\infty} a_{kn} \\
&\quad = \lim_{k\to\infty}\lim_{n\to\infty} a_{kn} \\
&\quad = \lim_{k\to\infty} \phi\left(R^{(k)}\left(P\left(\overrightarrow{f(X)}\right)\right) Q\left(\overrightarrow{f(X)}\right) R^{(k)}\left(P\left(\overrightarrow{f(X)}\right)\right)\right) \\
&\quad = \phi\left(g\left(P\left(\overrightarrow{f(X)}\right)\right) Q\left(\overrightarrow{f(X)}\right) g\left(P\left(\overrightarrow{f(X)}\right)\right)\right).
\end{aligned}
$$

This proves the claim. □

Chapter 3

Efficiency of Quantum State Tomography for Qubits

Abstract

The efficiency of quantum state tomography is discussed from the point of view of quantum parameter estimation theory, in which the trace of the weighted covariance is to be minimized. It is shown that tomography is optimal only when a special weight is adopted.

3.1 Motivation

Let $\mathcal{L}(\mathcal{H})$ be the set of linear operators on a Hilbert space $\mathcal{H} = \mathbb{C}^2$, and let $\mathcal{S} := \{\tau_x \mid x = (x^1, x^2, x^3) \in \mathcal{X}\}$ be the set of strictly positive density operators on $\mathcal{H}$ parametrized by the Stokes parameters $x \in \mathcal{X} := \{x \in \mathbb{R}^3 \mid (x^1)^2 + (x^2)^2 + (x^3)^2 < 1\}$ as

$$\tau_x := \frac{1}{2}(I + x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3), \tag{3.1}$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices. Suppose we have an unknown quantum state $\tau = \tau_x \in \mathcal{S}$. We are interested in identifying the true value of the parameter x.

Let

$$\mathcal{M}^{(s)}(\mathcal{H}) := \{(M_1, M_2, \ldots, M_s) \mid M_i \in \mathcal{L}(\mathcal{H}),\, M_i \geq 0,\, \sum_{i=1}^{s} M_i = I\}$$

be the set of positive operator-valued measures (POVMs) on $\mathcal{H}$ taking values on a finite set of outcomes labeled by $\{1, 2, \ldots, s\}$, and let $\mathcal{M}(\mathcal{H}) = \bigcup_{s=1}^{\infty} \mathcal{M}^{(s)}(\mathcal{H})$. Given POVMs $M = (M_1, M_2, \ldots, M_{s_1})$, $N = (N_1, N_2, \ldots, N_{s_2})$, and a real number p between 0 and 1, we can generate a new POVM by a randomized combination of them as follows:

$$pM \oplus (1-p)N := (pM_1, \ldots, pM_{s_1}, (1-p)N_1, \ldots, (1-p)N_{s_2}) \in \mathcal{M}(\mathcal{H}).$$

We can repeat this randomization procedure inductively to obtain $\bigoplus_{i=1}^{k} p_i M^{(i)} \in \mathcal{M}(\mathcal{H})$, where $M^{(1)}, M^{(2)}, \ldots, M^{(k)} \in \mathcal{M}(\mathcal{H})$ and $p_i \geq 0\,(1 \leq i \leq k)$ such that $\sum_{i=1}^{k} p_i = 1$. We shall call $\bigoplus_{i=1}^{k} p_i M^{(i)}$ a *random measurement* when $M^{(1)}, M^{(2)}, \ldots, M^{(k)} \in \mathcal{M}(\mathcal{H})$ are all projection-valued measurements (PVMs). Applying a random measurement means applying one of the projection-valued measurement $\{M^{(i)}\}_{1 \leq i \leq k}$ chosen at random according to the probability distribution $p = (p_i)_{1 \leq i \leq k}$[1].

Let $M^{(1)}, M^{(2)}, M^{(3)}$ be projection-valued measurements given by the spectral decomposition of the observables $\sigma_1, \sigma_2, \sigma_3$, respectively, and let $M^{(T)} := \frac{1}{3}(M^{(1)} \oplus M^{(2)} \oplus M^{(3)})$ be their random

[1] Helstrom [21] defined a random measurement based on a different type of convex structure of $\mathcal{M}(\mathcal{H})$ as $(pM_1 + (1-p)N_1, \ldots, pM_s + (1-p)N_s)$. Our definition of random measurement is seemingly different from his.

measurement according to the uniform distribution. Suppose that, among m applications of $M^{(T)}$ to the unknown state τ_x, the μth PVM $M^{(\mu)}$ has been chosen m_μ times and the outcomes ± 1 have been observed $m_\mu^\pm$ times, where $m = m_1 + m_2 + m_3$ and $m_\mu = m_\mu^+ + m_\mu^-$ for $\mu \in \{1, 2, 3\}$. We can construct an unbiased estimator for the Stokes parameters $x = (x^1, x^2, x^3)$ as

$$\hat{x}^\mu := \frac{m_\mu^+ - m_\mu^-}{m_\mu}, \qquad \mu \in \{1, 2, 3\}. \tag{3.2}$$

We shall call this estimator a *tomography* in this chapter. Note that the tomography can be regarded as a maximum likelihood estimator. In fact, since the probability distribution for the outcomes ± 1 of the μth PVM

$$M^{(\mu)} = \left(\frac{1}{2}(I + \sigma_\mu),\ \frac{1}{2}(I - \sigma_\mu) \right), \tag{3.3}$$

applied to the state $\tau_x \in \mathcal{S}$ is given by $p_{\tau_x}^{M^{(\mu)}} = (\frac{1+x^\mu}{2}, \frac{1-x^\mu}{2})$, the probability distribution for the outcome of the tomography $M^{(T)}$ is

$$p_{\tau_x}^{M^{(T)}} = \frac{1}{6}(1 + x^1, 1 - x^1, 1 + x^2, 1 - x^2, 1 + x^3, 1 - x^3). \tag{3.4}$$

As a consequence, the likelihood function for the outcomes $(m_\mu^\pm)_{1 \le \mu \le 3}$ obtained by m applications of $M^{(T)}$ is

$$l_m(x) = \sum_{\mu=1}^{3} \left(m_\mu^+ \log \frac{1 + x^\mu}{6} + m_\mu^- \log \frac{1 - x^\mu}{6} \right),$$

and it is easy to see that $\frac{\partial}{\partial x^\mu} l_m = 0$ is equivalent[2] to (3.2).

In order to investigate the optimality of the tomography, let us recall some basic facts from quantum parameter estimation theory. Let $\{\rho_\theta \mid \theta = (\theta^1, \ldots, \theta^d) \in \Theta\}$ be a smooth parametric family of density operators on a Hilbert space $\mathcal{H}$ with parameter space $\Theta \subset \mathbb{R}^d$. An estimator is represented by a pair $(M, \hat{\theta})$ of a POVM $M \in \mathcal{M}(\mathcal{H})$ and a map $\hat{\theta} : \mathbb{N} \to \Theta$ that gives the estimated value $\hat{\theta}(n)$ from each observed data $n \in \mathbb{N}$. An estimator $(M, \hat{\theta})$ is called *unbiased* if

$$E_\theta[M, \hat{\theta}] := \sum_{n \in \mathbb{N}} \hat{\theta}(n) \operatorname{Tr} \rho_\theta M_n = \theta \tag{3.5}$$

is satisfied for all $\theta \in \Theta$. An estimator $(M, \hat{\theta})$ is called *locally unbiased* [24] at a given point $\theta_0 \in \Theta$ if the condition (3.5) is satisfied around $\theta = \theta_0$ up to the first order of the Taylor expansion. It is well known that an estimator $(M, \hat{\theta})$ that is locally unbiased at θ_0 satisfies the following series of inequalities [24, 21]:

$$V_{\theta_0}[M, \hat{\theta}] \ge (g_{\theta_0}(M))^{-1} \ge (J_{\theta_0})^{-1}, \tag{3.6}$$

where $V_\theta[\cdot]$ denotes the covariance matrix, and $g_\theta(M)$ is the classical Fisher information matrix at θ with respect to $M \in \mathcal{M}(\mathcal{H})$ defined by

$$g_\theta(M) := \left[\sum_n \frac{(\frac{\partial}{\partial \theta^i} \operatorname{Tr} \rho_\theta M_n)(\frac{\partial}{\partial \theta^j} \operatorname{Tr} \rho_\theta M_n)}{\operatorname{Tr} \rho_\theta M_n} \right]_{1 \le i,j \le d}.$$

Further, J_θ is the quantum Fisher information matrix at θ given by

$$J_\theta := \left[\operatorname{Tr} (\frac{\partial}{\partial \theta^i} \rho_\theta) L_j \right]_{1 \le i,j \le d} = \left[\frac{1}{2} \operatorname{Tr} \rho_\theta (L_i L_j + L_j L_i) \right]_{1 \le i,j \le d},$$

where L_i is the ith symmetric logarithmic derivative (SLD) defined by the selfadjoint operator satisfying the equation

$$\frac{\partial}{\partial \theta^i} \rho_\theta = \frac{1}{2}(L_i \rho_\theta + \rho_\theta L_i). \tag{3.7}$$

[2] There are possibilities that $\hat{x} \notin \mathcal{X}$. However it follows from the law of large numbers of the tomography that $\hat{x} \in \mathcal{X}$ for sufficiently large m almost surely.

The inequality $V_{\theta_0}[M,\hat{\theta}] \geq (J_{\theta_0})^{-1}$ is called the *quantum Cramér-Rao inequality*. The first inequality in (3.6) is saturated when $\hat{\theta}^i(n) = \theta^i + \sum_j (g_\theta(M)^{-1})^{ij} \frac{\partial}{\partial \theta^j} (\log \mathrm{Tr}\, \rho_\theta M_n)$ is adopted. However the second inequality in (3.6) cannot be saturated in general because of the non-commutativity of the SLDs. To avoid this difficulty, we often adopt an alternative strategy to seek the estimator which minimizes $\mathrm{Tr}\, H_{\theta_0} V_{\theta_0}[M,\hat{\theta}]$, where H_θ is a given $d \times d$ real positive definite matrix for each θ called a *weight* [24, 21]. Thus the problem of finding the optimal estimator boils down to the problem of finding $M \in \mathcal{M}(\mathcal{H})$ which minimizes $\mathrm{Tr}\, H_{\theta_0} g_{\theta_0}(M)^{-1}$.

It is known that when $\dim \mathcal{H} = 2$, there is a definitive answer to the optimality of estimators, which is summarized in the following Propositions.

Proposition 3.1. *For a given weight H_θ,*

$$\min \left\{ \mathrm{Tr}\, H_\theta g_\theta(M)^{-1} \mid M \in \mathcal{M}(\mathcal{H}) \right\} = \left(\mathrm{Tr}\, R_\theta \right)^2, \tag{3.8}$$

where $R_\theta := \sqrt{\sqrt{J_\theta^{-1}} H_\theta \sqrt{J_\theta^{-1}}}$. The minimum is attained if and only if $M \in \mathcal{M}(\mathcal{H})$ satisfies

$$g_\theta(M) = \frac{\sqrt{J_\theta} R_\theta \sqrt{J_\theta}}{\mathrm{Tr}\, R_\theta}. \tag{3.9}$$

Proposition 3.1 was first proved by Nagaoka [37] (cf. [7]) when $d = 2$. The case $d = 3$ is proved by Hayashi [17], and independently by Gill and Massar [13]. Further, Nagaoka constructed explicitly a measurement which attains the minimum when $d = 2$. His construction of an optimal estimator can be generalized as follows.

Proposition 3.2. *Given a weight H_θ, let us diagonalize R_θ as $R_\theta = USU^{-1}$ where $S = diag(S_1, \ldots, S_d)$ is a diagonal matrix and $U \in O(d)$, and let $M^{(i)}$ be a projection-valued measurement given by the spectral decomposition of the operator*

$$\hat{L}^i := \sum_{k=1}^{d} K^{ik} L_k, \tag{3.10}$$

where $K^{ik} := (U^{-1} \sqrt{J_\theta^{-1}})^{ik}$. Then the random measurement

$$M := p_1 M^{(1)} \oplus \cdots \oplus p_d M^{(d)} \tag{3.11}$$

satisfies (3.9)*, where $p_i := S_i/(S_1 + \cdots + S_d)$.*

Note that the optimal measurement (3.11) depends on the true value of $\theta \in \Theta$ in general. In such a case, we necessary invoke an adaptive estimation scheme [9] to achieve the minimum (3.8).

Now it is natural to inquire whether the tomography is optimal in view of Propositions 3.1 and 3.2. The answer is given by the following.

Theorem 3.3. *Tomography is optimal if and only if the weight H_x is proportional to the following special one:*

$$H_x^{(T)} := \begin{pmatrix} \frac{1}{1-(x^1)^2} & -\frac{(x^1)(x^2)}{(1-(x^1)^2)(1-(x^2)^2)} & -\frac{(x^3)(x^1)}{(1-(x^3)^2)(1-(x^1)^2)} \\ -\frac{(x^1)(x^2)}{(1-(x^1)^2)(1-(x^2)^2)} & \frac{1}{1-(x^2)^2} & -\frac{(x^2)(x^3)}{(1-(x^2)^2)(1-(x^3)^2)} \\ -\frac{(x^3)(x^1)}{(1-(x^3)^2)(1-(x^1)^2)} & -\frac{(x^2)(x^3)}{(1-(x^2)^2)(1-(x^3)^2)} & \frac{1}{1-(x^3)^2} \end{pmatrix}. \tag{3.12}$$

Note that $H_x^{(T)}$ is not rotationally symmetric. This implies that the optimal weight depends on the choice of the coordinate axes. Theorem 3.3 also implies that the tomography is not optimal for a rotationally symmetric weight that is natural for a physical point of view.

This chapter is organized as follows. Theorem 3.3 is proved in Section 3.2, and the non-optimality of the tomography for a rotationally symmetric weight is discussed and numerically demonstrated in Section 3.3. An extension to the case when $\dim \mathcal{H} \geq 3$ is also discussed there. For the reader's convenience, simple proofs of Propositions 3.1 and 3.2 are given in Appendix.

3.2 Proof of Theorem 3.3

We prove Theorem 3.3 in a series of Lemmas.

Lemma 3.4. *Let L_μ be the SLD of $\frac{\partial}{\partial x^\mu}$ for $\mu \in \{1,2,3\}$. Then*

$$L_\mu = \sigma_\mu - \frac{x^\mu}{2\det\tau}(I-\tau).$$

Proof. We need only verify that L_μ satisfies equation (3.7).

$$L_\mu\tau = \sigma_\mu\tau - \frac{x^\mu}{2\det\tau}\tau(I-\tau) = \sigma_\mu\tau - \frac{x^\mu}{2}I.$$

Therefore

$$\begin{aligned}\frac{1}{2}(L_\mu\tau+\tau L_\mu) &= \frac{1}{2}(\{\tau,\sigma_\mu\}-x^\mu I) = \frac{1}{2}(\{\frac{1}{2}I,\sigma_\mu\}+\{\frac{x^\mu}{2}\sigma_\mu,\sigma_\mu\}-x^\mu I)\\ &= \frac{1}{2}(\sigma_\mu + x^\mu I - x^\mu I) = \frac{\sigma_\mu}{2} = \frac{\partial}{\partial x^\mu}\tau\end{aligned}$$

where $\{A,B\} := AB+BA$ for $A,B\in\mathcal{L}(\mathcal{H})$. □

Lemma 3.5. *Let J_x be the SLD Fisher information matrix at x. Then*

$$J_x = (I - |x\rangle\langle x|)^{-1}$$

where $|x\rangle = \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}$.

Proof. We calculate the elements of J_x.

$$(J_x)_{\mu\nu} = \operatorname{Tr}\frac{\partial\tau}{\partial x^\mu}L_\nu = \operatorname{Tr}\frac{\sigma_\mu}{2}\left(\sigma_\mu - \frac{x^\mu}{2\det\tau}(I-\tau)\right) = \delta_{\mu\nu} + \frac{x^\mu x^\nu}{4\det\tau}.$$

Thus

$$J_x = I + \frac{1}{4\det\tau}|x\rangle\langle x| = I + \frac{1}{1-r^2}|x\rangle\langle x|.$$

Then

$$(I-|x\rangle\langle x|)\left(I+\frac{1}{1-r^2}|x\rangle\langle x|\right) = I + \frac{1}{1-r^2}|x\rangle\langle x| - |x\rangle\langle x| - \frac{r^2}{1-r^2}|x\rangle\langle x| = I,$$

where $r=\sqrt{\langle x|x\rangle}$. Therefore $I+\frac{1}{1-r^2}|x\rangle\langle x| = (I-|x\rangle\langle x|)^{-1}$. □

Lemma 3.6. *Given $F_x \in g_x(\mathcal{M}(\mathcal{H}))$ with $F_x > 0$. There exists a weight H_x such that*

$$\min_{M\in\mathcal{M}(\mathcal{H})}\left\{\operatorname{Tr} H_x g_x(M)^{-1}\right\} = \operatorname{Tr} H_x F_x^{-1} \tag{3.13}$$

if and only if

$$\operatorname{Tr} J_x^{-1}F_x = 1. \tag{3.14}$$

Further, when (3.14) *is satisfied,*

$$H_x = kF_x\,J_x^{-1}F_x \tag{3.15}$$

is the only weight which satisfies (3.13) *where k is an arbitrary real positive number.*

Proof. We first assume that there exists a weight H_x which satisfies (3.13). Let $R_x := \sqrt{\sqrt{J_x^{-1}} H_x \sqrt{J_x^{-1}}}$. According to Proposition 3.1, F_x must be

$$F_x = \frac{\sqrt{J_x} R_x \sqrt{J_x}}{\operatorname{Tr} R_x},$$

so that

$$\operatorname{Tr} J_x^{-1} F_x = \operatorname{Tr} J_x^{-1} \frac{\sqrt{J_x} R_x \sqrt{J_x}}{\operatorname{Tr} R_x} = 1.$$

Then we conclude (3.14).

We next assume that (3.14) is satisfied. Let $H_x = k F_x J_x^{-1} F$. It follows from Proposition 3.1 that

$$\begin{aligned} \min_{M \in \mathcal{M}(\mathcal{H})} \operatorname{Tr} H_x g_x(M)^{-1} &= \left(\operatorname{Tr} \sqrt{k \sqrt{J_x^{-1}} F_x J_x^{-1} F_x \sqrt{J_x^{-1}}} \right)^2 \\ &= k \left(\operatorname{Tr} J_x^{-1} F_x \right)^2 = k \left(\operatorname{Tr} J_x^{-1} F_x \right) \\ &= \operatorname{Tr} \left(k F_x J_x^{-1} F_x \right) F_x^{-1} = \operatorname{Tr} H_x F_x^{-1}. \end{aligned}$$

Further, the weight of the form (3.15) are the only weights which satisfy (3.13) because the mapping

$$M^{(1)}(d, \mathbb{R}) \ni H_x \mapsto \frac{\sqrt{J_x} \sqrt{\sqrt{J_x^{-1}} H_x \sqrt{J_x^{-1}}} \sqrt{J_x}}{\operatorname{Tr} \sqrt{\sqrt{J_x^{-1}} H_x \sqrt{J_x^{-1}}}} = \frac{\sqrt{J_x} R_x \sqrt{J_x}}{\operatorname{Tr} R_x} \in g_x(\mathcal{M}(\mathcal{H}))$$

is injective where $M^{(1)}(d, \mathbb{R}) := \{G \mid G \text{ is } d \times d \text{ real positive definite matrix, } \operatorname{Tr} G = 1\}$. □

Proof of Theorem 3.3. We can calculate the classical Fisher information matrix with respect to $M^{(T)}$ from (3.4) as follow:

$$g_x(M^{(T)}) = \frac{1}{3} \begin{pmatrix} \frac{1}{1-(x^1)^2} & 0 & 0 \\ 0 & \frac{1}{1-(x^2)^2} & 0 \\ 0 & 0 & \frac{1}{1-(x^3)^2} \end{pmatrix}. \tag{3.16}$$

Then

$$\begin{aligned} \operatorname{Tr} J_x^{-1} g_x(M^{(T)}) &= \operatorname{Tr} \frac{1}{3} (I - |r\rangle \langle r|) \begin{pmatrix} \frac{1}{1-(x^1)^2} & 0 & 0 \\ 0 & \frac{1}{1-(x^2)^2} & 0 \\ 0 & 0 & \frac{1}{1-(x^3)^2} \end{pmatrix} \\ &= \frac{1}{3} \left(\frac{1}{1-(x^1)^2} + \frac{1}{1-(x^2)^2} + \frac{1}{1-(x^3)^2} - \frac{(x^1)^2}{1-(x^1)^2} - \frac{(x^2)^2}{1-(x^2)^2} - \frac{(x^3)^2}{1-(x^3)^2} \right) \\ &= 1 \end{aligned}$$

We see from Lemma 3.6 that $H_x := k\, g_x(M^{(T)})\, J_x^{-1}\, g_x(M^{(T)})$ are the only weights which satisfy

$$\min_{N \in \mathcal{M}(\mathcal{H})} \left\{ \operatorname{Tr} H_x\, g_x(N)^{-1} \right\} = \operatorname{Tr} H_x\, g_x(M^{(T)})^{-1}.$$

Then

$$\begin{aligned} & k\, g_x(M^{(T)})\, J_x^{-1} g_x(M^{(T)}) \\ = k\, g_x(M^{(T)})\, (I - |x\rangle \langle x|) g_x(M^{(T)}) &= k \left(g_x(M^{(T)})^2 - g_x(M^{(T)})\, |x\rangle \langle x|\, g_x(M^{(T)}) \right) \\ = 9k \begin{pmatrix} \frac{1}{1-(x^1)^2} & -\frac{(x^1)(x^2)}{(1-(x^1)^2)(1-(x^2)^2)} & -\frac{(x^3)(x^1)}{(1-(x^3)^2)(1-(x^1)^2)} \\ -\frac{(x^1)(x^2)}{(1-(x^1)^2)(1-(x^2)^2)} & \frac{1}{1-(x^2)^2} & -\frac{(x^2)(x^3)}{(1-(x^2)^2)(1-(x^3)^2)} \\ -\frac{(x^3)(x^1)}{(1-(x^3)^2)(1-(x^1)^2)} & -\frac{(x^2)(x^3)}{(1-(x^2)^2)(1-(x^3)^2)} & \frac{1}{1-(x^3)^2} \end{pmatrix} \\ &= 9k\, H_x^{(T)}. \end{aligned}$$

□

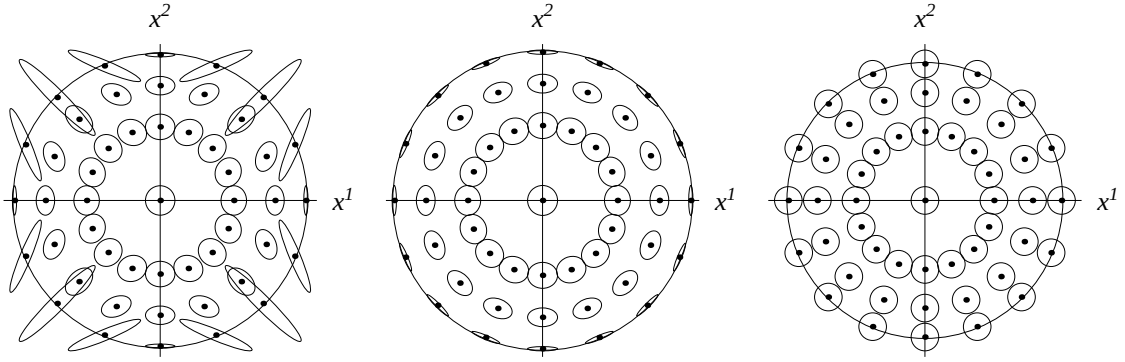

Figure 3.1: Indicatrices for several typical weights H_x, where $H_x = H_x^{(T)}$ (left), $H_x = J_x$ (middle), and $H_x = I$ (right).

3.3 Discussions

Let us investigate the properties of the weight $H_x^{(T)}$ that is optimal for the tomography. We first regard a weight H_x as a metric tensor on the tangent space $\mathcal{T}_{\tau_x}\mathcal{S}$ at $x \in \mathcal{X}$, and let us plot the indicatrix, the set of end points of tangent vectors $\boldsymbol{v} \in \mathcal{T}_x\mathcal{S}$ centered at x satisfying ${}^t\boldsymbol{v}H_x\boldsymbol{v} = 1$. Figure 3.1 shows the indicatrices on the x^1x^2-plane for $H_x = H_x^{(T)}$ (left), $H_x = J_x$ (middle), and $H_x = I$ (right). Obviously $H_x^{(T)}$ is not rotationally symmetric, and is awkwardly distorted when $x = (x^1, x^2, x^3) \in \mathcal{X}$ is off the coordinate axes. This means that the tomography depends highly on the choice of the coordinate axes. Actually, an estimation scheme should be independent of the choice of the coordinate axes because their choice is completely arbitrary. It is therefore natural to adopt a rotationally symmetric weight H_x which satisfies $U^*H_{(Ux)}U = H_x$ for $U \in SO(3)$.

Any rotationally symmetric weight can be represented by

$$H_x^{(f,g)} := f(r)I + (g(r) - f(r))\frac{1}{r^2}\left|x\right\rangle\left\langle x\right|, \tag{3.17}$$

for $x \neq 0$ where f, g are functions on $(0, 1)$ such that $f(r) > 0$ and $g(r) > 0$ (see Appendix 3.B). Given a weight $H_x = H_x^{(f,g)}$, let $M^{(f,g)} \in \mathcal{M}(\mathcal{H})$ be the corresponding optimal measurement given by (3.11), and let $c_x := \operatorname{Tr} H_x^{(f,g)} g_x (M^{(f,g)})^{-1}$ and $c_x^{(T)} := \operatorname{Tr} H_x^{(f,g)} g_x (M^{(T)})^{-1}$. It then follows from (3.8) and (3.16) that

$$\begin{aligned} c_x &= \left(\operatorname{Tr}\sqrt{\sqrt{J_x^{-1}}H_x^{(f,g)}\sqrt{J_x^{-1}}}\right)^2 \\ &= \left(2\sqrt{f(r)} + \sqrt{(1-r^2)g(r)}\right)^2, \end{aligned} \tag{3.18}$$

and

$$c_x^{(T)} = 3(2f(r) + (1-r^2)g(r)) + 3tr^2(g(r) - f(r)), \tag{3.19}$$

where $t := 1 - \frac{(x^1)^4 + (x^2)^4 + (x^3)^4}{r^4}$. Note that $0 \leq t \leq \frac{2}{3}$, and that $t = 0$ if and only if x is on one of the coordinate axes, and $t = \frac{2}{3}$ if and only if x is parallel to one of the vectors $(1, 1, 1)$, $(-1, 1, 1)$, $(1, -1, 1)$, and $(1, 1, -1)$. In addition,

$$c_x^{(T)} - c_x = 2\left(\sqrt{(1-r^2)g(r)} - \sqrt{f(r)}\right)^2 + 3r^2\left(g(r) - f(r)\right)t \tag{3.20}$$

$$= 2\left(\sqrt{(1-r^2)f(r)} - \sqrt{g(r)}\right)^2 + 3r^2\left(f(r) - g(r)\right)\left(\frac{2}{3} - t\right). \tag{3.21}$$

By considering the cases $g(r) \geq f(r)$ and $f(r) > g(r)$ separately, we conclude that $c_x^{(T)} \geq c_x$ for any rotationally symmetric weight $H_x^{(f,g)}$.

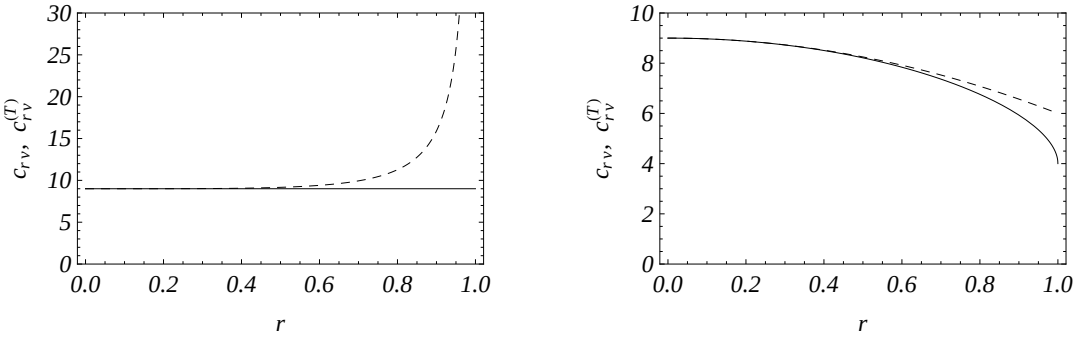

Figure 3.2: The behavior of $c_{r\boldsymbol{v}}$ (solid) and $c^{(T)}_{r\boldsymbol{v}}$ (dashed) as functions of radius r in the direction $\boldsymbol{v} = \frac{1}{\sqrt{3}}(1,1,1)^t$ for $H_x^{(f,g)} = J_x$ (left) and $H_x^{(f,g)} = I$ (right).

For example, when $H_x^{(f,g)} = J_x$, for which $f(r) = 1$ and $g(r) = \frac{1}{1-r^2}$, we see that $g(r) - f(r) \to \infty$ as $r \to 1$, so that $c_x^{(T)}$ becomes much larger than c_x. On the other hand, when $H_x^{(f,g)} = I$, for which $f(r) = g(r) = 1$, the second terms in (3.20) and (3.21) vanish, and the difference $c_x^{(T)} - c_x$ becomes relatively small. Figure 3.2 shows the behavior of $c_{r\boldsymbol{v}}$ (solid) and $c^{(T)}_{r\boldsymbol{v}}$ (dashed) as functions of radius r in the direction $\boldsymbol{v} = \frac{1}{\sqrt{3}}(1,1,1)^t$ for $H_x^{(f,g)} = J_x$ (left) and $H_x^{(f,g)} = I$ (right). When $H_x^{(f,g)} = J_x$, we see that $c^{(T)}_{r\boldsymbol{v}}$ diverges as $r \to 1$, while $c_{r\boldsymbol{v}}$ converges to 9. When $H_x^{(f,g)} = I$, on the other hand, $c^{(T)}_{r\boldsymbol{v}}$ and $c_{r\boldsymbol{v}}$ converge to 6 and 4 respectively as $r \to 1$, and their difference is relatively small.

Now let us make a numerical simulation to compare the asymptotic performance of the tomography and the optimal adaptive estimation schemes for $H_x = J_x$ and $H_x = I$. We set the qubit state to be estimated as τ_{x_0} with $x_0 = (0.55, 0.55, 0.55)$. Since the optimal estimator given in Proposition 3.2 depends on the true value of $x \in \mathcal{X}$, we shall invoke an adaptive estimation scheme in evaluating $\mathrm{Tr}\, H_x g_x(M(x))^{-1}$, with $M(x)$ being the optimal POVM for $x \in \mathcal{X}$, as follows [36, 9]: We begin by choosing $\hat{x}^{(0)} \in \mathcal{X}$ arbitrarily. Suppose that $M(\hat{x}^{(0)})$ is applied and that the outcome $n_1 \in \{1, 2, \ldots, s\}$ is obtained. The maximum likelihood estimator is given by

$$\hat{x}^{(1)} := \operatorname*{argmax}_{x \in \mathcal{X}} l_1(x),$$

where

$$l_1(x) := \log \mathrm{Tr}\, \tau(x)\, M_{n_1}(\hat{x}^{(0)}).$$

At the mth stage ($m \geq 2$), suppose that $M(\hat{x}^{(m-1)})$ is applied and that the outcome $n_m \in \{1, 2, \ldots, s\}$ is obtained. The maximum likelihood estimator at the mth stage is given by

$$\hat{x}^{(m)} := \operatorname*{argmax}_{x \in \mathcal{X}} l_m(x),$$

where

$$l_m(x) := \sum_{i=1}^{m} \log \mathrm{Tr}\, \tau(x)\, M_{n_i}(\hat{x}^{(i-1)}).$$

Because of the strong consistency and the asymptotic efficiency of the adaptive estimation [9], the sequence $m \times \mathrm{Tr}\, H_{x_0} V[\hat{x}^{(m)}]$ of the weighted covariances multiplied by m converges to $\mathrm{Tr}\, H_{x_0} g_{x_0}(M(x_0))^{-1}$ as $m \to \infty$. Let us demonstrate this behavior by a numerical simulation. We have performed two kinds of numerical simulations in which the weight H_x has been set as $H_x = J_x$ and $H_x = I$. These results are shown in the left and the right figure in Figure 3.3, where the solid and dashed curves correspond to the adaptive estimation and the tomography, and the solid and dashed horizontal lines correspond to the theoretical limits. As figures of merit, we have plotted in Figure 3.3 the sample averages of $2m \times B(\tau_{x_0}, \tau_{\hat{x}^{(m)}})$, where $B(\cdot, \cdot)$ is

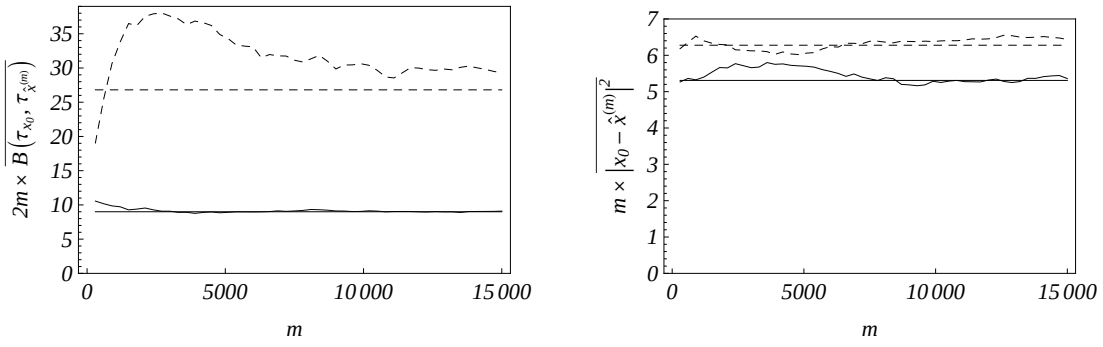

Figure 3.3: A numerical comparison between the tomography and the optimal adaptive estimation for the weight H_x, where H_x has been set as $H_x = J_x$ (left) or $H_x = I$ (right). The solid and dashed curves correspond to the adaptive estimation and the tomography, respectively, and the solid and dashed horizontal lines correspond to the theoretical limit. As a figure of merit, we have plotted the sample averages of $2m \times B(\tau_{x_0}, \tau_{\hat{x}^{(m)}})$ or $m \times |x_0 - \hat{x}^{(m)}|^2$ instead of $m \times \mathrm{Tr}\, H_{x_0} V[\hat{x}^{(m)}]$.

the Bures distance, or $m \times |x_0 - \hat{x}^{(m)}|^2$ instead of $m \times \mathrm{Tr}\, J_{x_0} V[\hat{x}^{(m)}]$ or $m \times \mathrm{Tr}\, V[\hat{x}^{(m)}]$ because they are asymptotically equivalent (See Appendix 3.C). The sample averages are calculated by repeating the estimation schemes 1000 times. We see that the sample average of each estimation scheme approaches the corresponding theoretical value, as m becomes large. We further observe that the adaptive estimation scheme is more efficient than the tomography, and the difference of their performances is noticeable when $H_x = J_x$. We could conclude that the tomography is not efficient for a rotationally symmetric weight that is natural in estimating an unknown qubit state.

Finally we shall touch upon a generation to a higher dimensional Hilbert space $\mathcal{H}$. Let $q = \dim \mathcal{H} (\geq 3)$ and let $\{\left|e_i^{(\alpha)}\right\rangle\}_{i=1}^{q}$ be an orthonormal basis for each $\alpha = 1, \ldots, q+1$ satisfying $|\left\langle e_i^{(\alpha)} | e_j^{(\beta)} \right\rangle|^2 = \frac{1}{q}$ $(\alpha \neq \beta)$ for all i, j. A finite subset $\{\left|e_i^{(\alpha)}\right\rangle\}_{\alpha,i}$ of the Hilbert space $\mathcal{H}$ is called a full set of *mutually unbiased bases*. It is known that a full set of mutually unbiased bases exists when q is a prime number or the power of a prime [1]. As before, we regard the uniform combination

$$M^{(T)} := \frac{1}{q+1} \bigoplus_{\alpha=1}^{q+1} M^{(\alpha)}$$

of the PVMs $M^{(\alpha)} := (\left|e_1^{(\alpha)}\right\rangle \left\langle e_1^{(\alpha)}\right|, \cdots, \left|e_q^{(\alpha)}\right\rangle \left\langle e_q^{(\alpha)}\right|) \in \mathcal{M}(\mathcal{H})$ as a tomography on $\mathcal{H}$. Let $\mathcal{S}$ be the set of strictly positive density operators on $\mathcal{H}$, and let $x = \{x_{\alpha,i}\}$ be an affine parametrization of $\mathcal{S}$ given by

$$\tau_x = \frac{1}{q} I + \sum_{\alpha=1}^{q+1} \sum_{i=1}^{q-1} x_{\alpha,i} (\left|e_i^{(\alpha)}\right\rangle \left\langle e_i^{(\alpha)}\right| - \frac{1}{q} I).$$

Figure 3.4 shows the behavior of $c_{r\boldsymbol{v}}$ (solid) and $c_{r\boldsymbol{v}}^{(T)}$ (dashed) as functions of r in the direction $\boldsymbol{v} \in \mathbb{R}^{q^2-1}$ where

$$\begin{aligned} c_x &:= \min\{\mathrm{Tr}\, J_x g_x(M)^{-1} \mid M \in \mathcal{M}(\mathcal{H})\}, \\ c_x^{(T)} &:= \mathrm{Tr}\, J_x g_x(M^{(T)})^{-1} \end{aligned}$$

with $v_{11} = 1$ and $v_{\alpha i} = 0$ ($\alpha \neq 1$ or $i \neq 1$) for $\dim \mathcal{H} = 3$ (left) and $\dim \mathcal{H} = 4$ (right). We see that the behavior for $\dim \mathcal{H} = 3$ and 4 are almost the same as that for $\dim \mathcal{H} = 2$ plotted in Figure 3.2. This observation suggests that the same non-optimality result would hold for $\dim \mathcal{H} \geq 3$.

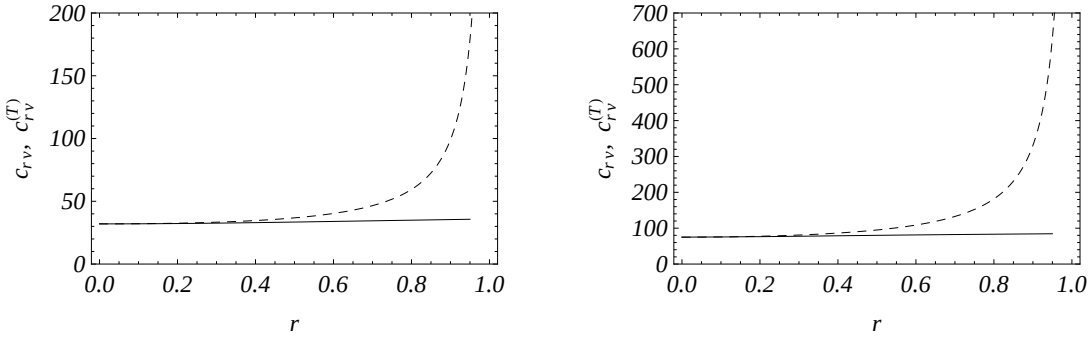

Figure 3.4: The behavior of $c_{r\boldsymbol{v}}$ (solid) and $c^{(T)}_{r\boldsymbol{v}}$ (dashed) as functions of r in the direction $\boldsymbol{v} \in \mathbb{R}^{q^2-1}$ where $v_{11} = 1$ and $v_{\alpha i} = 0$ ($\alpha \neq 1$ or $i \neq 1$) for $\dim \mathcal{H} = 3$ (left) and $\dim \mathcal{H} = 4$ (right).

Appendices

Appendix 3.A Proofs of Propositions 3.1 and 3.2

In this appendix, we give simple proofs of Propositions 3.1 and 3.2 for the reader's convenience. We start with some lemmas which hold for an arbitrary finite dimensional Hilbert space $\mathcal{H}$. Let us define the inner product $\langle \cdot, \cdot \rangle_\theta$ on $\mathcal{L}(\mathcal{H})$, as

$$\langle A, B \rangle_\theta := \frac{1}{2} \mathrm{Tr}\, \rho_\theta (A^* B + B A^*).$$

Then we can rewrite $g_\theta(M)$ by SLD as follows:

$$g_\theta(M) = \left[\sum_x \frac{\langle L_i, M_x \rangle_\theta \langle L_j, M_x \rangle_\theta}{\langle I, M_x \rangle_\theta} \right]_{1 \leq i,j \leq d},$$

Further we can also rewrite J_θ as $J_\theta = [\langle L_i, L_j \rangle]_{ij}$. Let us define $\hat{L}^i$ as (3.10). Let us define

$$\hat{g}_\theta(M) = \left[\sum_x \frac{\left\langle \hat{L}^i, M_x \right\rangle_\theta \left\langle \hat{L}^j, M_x \right\rangle_\theta}{\langle I, M_x \rangle_\theta} \right]_{1 \leq i,j \leq d}.$$

Lemma 3.7. *$\{\hat{L}^i\}_i \cup \{I\}$ is orthonormal with respect to $\langle \cdot, \cdot \rangle_\theta$.*

Proof.

$$\left\langle \hat{L}^i, \hat{L}^j \right\rangle_\theta = \sum_{s,t} K^{is} K^{jt} \langle L_s, L_t \rangle_\theta = \sum_{s,t} K^{is} J_{\theta,st} (K^*)^{tj} = (U^{-1} \sqrt{J_\theta^{-1}} J_\theta \sqrt{J_\theta^{-1}} U)^{ij} = \delta^{ij}.$$

Further

$$\left\langle \hat{L}^i, I \right\rangle_\theta = \sum_s K^{is} \langle L_s, I \rangle_\theta = \sum_s K^{is} \mathrm{Tr}\, \rho_\theta L_s = \sum_s K^{is} \mathrm{Tr}\, \partial_i \rho_\theta = 0,$$

then

$$\langle I, I \rangle_\theta = \mathrm{Tr}\, \rho_\theta = 1.$$

□

Lemma 3.8. *It holds that*

$$\mathrm{Tr}\, \hat{g}_\theta(M) \leq \dim \mathcal{H} - 1,$$

for all $M \in \mathcal{M}(\mathcal{H})$.

Proof.

$$\begin{aligned}
\operatorname{Tr}\hat{g}_\theta(M) &= \sum_x \frac{\sum_{i=1}^d \left\langle \hat{L}^i,\, M_x\right\rangle_\theta^2}{\langle I,\, M_x\rangle_\theta} = \sum_x \left(\sum_{i=1}^d \frac{\left\langle \hat{L}^i,\, M_x\right\rangle_\theta^2 + \langle I,\, M_x\rangle_\theta^2}{\langle I,\, M_x\rangle_\theta} - \langle I,\, M_x\rangle_\theta \right) \\
&\leq \sum_x \left(\frac{\langle M_x,\, M_x\rangle_\theta}{\langle I,\, M_x\rangle_\theta} - \langle I,\, M_x\rangle_\theta \right) \qquad (3.22) \\
&= \sum_x \frac{\langle M_x,\, M_x\rangle_\theta}{\langle I,\, M_x\rangle_\theta} - 1 \leq \sum_x \operatorname{Tr} M_x - 1 \qquad (3.23) \\
&= \operatorname{Tr} I - 1 = \dim \mathcal{H} - 1.
\end{aligned}$$

Inequality (3.22) follows from Bessel's inequality, and inequality (3.23) from

$$\langle I,\, M_x\rangle_\theta \operatorname{Tr} M_x = (\operatorname{Tr}\rho_\theta M_x)(\operatorname{Tr} M_x) \geq \operatorname{Tr}\rho_\theta M_x^2 = \langle M_x,\, M_x\rangle_\theta .$$

□

Lemma 3.9. *Let $g_\theta(\mathcal{M}(\mathcal{H})) := \{g_\theta(M) \mid M \in \mathcal{M}(\mathcal{H})\}$. Then $g_\theta(\mathcal{M}(\mathcal{H}))$ is a convex set. Similarly, $\hat{g}_\theta(\mathcal{M}(\mathcal{H}))$ is also a convex set.*

Proof. Let $M^{(1)}, M^{(2)} \in \mathcal{M}(\mathcal{H})$ and let $0 \leq p \leq 1$. Then we see

$$\begin{aligned}
&g_\theta(pM^{(1)} \oplus (1-p)M^{(2)})_{ij} \\
&\quad = \sum_x \frac{p^2 \left\langle L_i,\, M_x^{(1)}\right\rangle_\theta \left\langle L_j,\, M_x^{(1)}\right\rangle_\theta}{p\left\langle I,\, M_x^{(1)}\right\rangle_\theta} + \sum_y \frac{(1-p)^2 \left\langle L_i,\, M_y^{(2)}\right\rangle_\theta \left\langle L_j,\, M_y^{(2)}\right\rangle_\theta}{(1-p)\left\langle I,\, M_y^{(2)}\right\rangle_\theta} \\
&\quad = \sum_x p\frac{\left\langle L_i,\, M_x^{(1)}\right\rangle_\theta \left\langle L_j,\, M_x^{(1)}\right\rangle_\theta}{\left\langle I,\, M_x^{(1)}\right\rangle_\theta} + \sum_y (1-p)\frac{\left\langle L_i,\, M_y^{(2)}\right\rangle_\theta \left\langle L_j,\, M_y^{(2)}\right\rangle_\theta}{\left\langle I,\, M_y^{(2)}\right\rangle_\theta} \\
&\quad = pg_\theta(M^{(1)})_{ij} + (1-p)g_\theta(M^{(2)})_{ij}. \qquad (3.24)
\end{aligned}$$

This implies that any convex combination of $g_\theta(M^{(1)})$ and $g_\theta(M^{(2)})$ belongs to $g_\theta(\mathcal{M}(\mathcal{H}))$. □

Now we restrict ourselves to the case when $\dim\mathcal{H} = 2$. In this case it is necessary that $1 \leq d \leq 3$.

Lemma 3.10. *Given $\boldsymbol{v} = (v_1, \ldots, v_d)^t \in \mathbb{R}^d$ such that $|\boldsymbol{v}| = 1$, then*

$$\hat{g}_\theta(M^{(\boldsymbol{v})}) = |\boldsymbol{v}\rangle\langle\boldsymbol{v}|, \qquad (3.25)$$

where $M^{(\boldsymbol{v})}$ is a projection-valued measurement given by the spectral decomposition of $L_{\boldsymbol{v}} := \sum_{i=1}^d v_i \hat{L}^i$.

Proof.

$$\begin{aligned}
\langle\boldsymbol{v}|\,\hat{g}_\theta(M^{(\boldsymbol{v})})\,|\boldsymbol{v}\rangle &= \textstyle\sum_x \sum_{i,j} v_i v_j \frac{\left\langle \hat{L}^i, M_x^{(\boldsymbol{v})}\right\rangle_\theta \left\langle \hat{L}^j, M_x^{(\boldsymbol{v})}\right\rangle_\theta}{\left\langle I, M_x^{(\boldsymbol{v})}\right\rangle_\theta} \\
&= \textstyle\sum_x \frac{\left\langle L_{\boldsymbol{v}}, M_x^{(\boldsymbol{v})}\right\rangle_\theta^2}{\left\langle I, M_x^{(\boldsymbol{v})}\right\rangle_\theta} = \sum_x \frac{\left\langle L_{\boldsymbol{v}}, M_x^{(\boldsymbol{v})}\right\rangle_\theta^2}{\left\langle M_x^{(\boldsymbol{v})}, M_x^{(\boldsymbol{v})}\right\rangle_\theta} = \sum_x \left\langle L^{\boldsymbol{v}},\, \tilde{M}_x^{(\boldsymbol{v})}\right\rangle_\theta^2 \\
&\leq \langle L_v,\, L_v\rangle_\theta = 1, \qquad (3.26)
\end{aligned}$$

where $\tilde{M}_x^{(\boldsymbol{v})} := M_x^{(\boldsymbol{v})} / \sqrt{\left\langle M_x^{(\boldsymbol{v})},\, M_x^{(\boldsymbol{v})}\right\rangle}$. Because $\{\tilde{M}_x^{(\boldsymbol{v})}\}_x$ is orthonormal with respect to $\langle\cdot,\,\cdot\rangle_\theta$, the inequality (3.26) follows from Bessel's inequality. Further by definition, $L_{\boldsymbol{v}} \in \operatorname{span}\{\tilde{\mathrm{M}}_{\mathrm{x}}^{(\boldsymbol{v})}\}_{\mathrm{x}}$. Therefor

$$\langle\boldsymbol{v}|\,\hat{g}_\theta(M^{(\boldsymbol{v})})\,|\boldsymbol{v}\rangle = 1. \qquad (3.27)$$

According to Lemma 3.8,

$$\operatorname{Tr} \hat{g}_\theta(M^{(\boldsymbol{v})}) \leq \dim \mathcal{H} - 1 = 1. \tag{3.28}$$

We can conclude (3.25) from (3.27) and (3.28) and $\hat{g}_\theta(M^{(\boldsymbol{v})}) \geq 0$. □

Lemma 3.11. *Let $M^+(d,\mathbb{R})$ be the set of $d \times d$ real positive semi definite matrices. Then*

$$\hat{g}_\theta(\mathcal{M}(\mathcal{H})) = \left\{ G \in M^+(d,\mathbb{R}) \mid \operatorname{Tr} G \leq 1 \right\}.$$

Proof. According to Lemma 3.10, for any $\boldsymbol{v} = (v_1, \ldots, v_d)^t \in \mathbb{R}^d$ such that $|\boldsymbol{v}| = 1$,

$$|v\rangle \langle v| \in \hat{g}_\theta(\mathcal{M}(\mathcal{H})).$$

We further observe that $0 \in \hat{g}_\theta(\mathcal{M}(\mathcal{H}))$ because the POVM $M^{(0)} := (I)$ provides no information. Then we see from Lemma 3.9 that

$$\hat{g}_\theta(\mathcal{M}(\mathcal{H})) \supset co(\left\{ |v\rangle \langle v| \;\middle|\; v \in \mathbb{R}^d, |\boldsymbol{v}| = 1 \right\} \cup \{0\}) = \left\{ G \in M^+(d,\mathbb{R}) | \operatorname{Tr} G \leq 1 \right\}.$$

The converse inclusion follows from Lemma 3.8. □

Lemma 3.12.

$$g_\theta(\mathcal{M}(\mathcal{H})) = \left\{ \sqrt{J_\theta} G \sqrt{J_\theta} \;\middle|\; G \in M^+(d,\mathbb{R}), \operatorname{Tr} G \leq 1 \right\}.$$

Proof.

$$\hat{g}_\theta(M)_{ij} = \sum_{st} K^{is} K^{jt} g_\theta(M)_{st} = \sum_{st} K^{is} g_\theta(M)_{st} (K^*)^{tj},$$

thus

$$\hat{g}_\theta(M) = U^{-1} \sqrt{J_\theta^{-1}}\, g_\theta(M) \sqrt{J_\theta^{-1}} U.$$

Therefore

$$\sqrt{J_\theta} U \hat{g}_\theta(M) U^{-1} \sqrt{J_\theta} = g_\theta(M).$$

It follows from lemma 3.11 that

$$g_\theta(\mathcal{M}(\mathcal{H})) = \left\{ \sqrt{J_\theta} U G U^{-1} \sqrt{J_\theta} \;\middle|\; G \in \hat{g}_\theta(\mathcal{M}(\mathcal{H})) \right\} = \left\{ \sqrt{J_\theta} G \sqrt{J_\theta} \;\middle|\; G \in M^+(d,\mathbb{R}), \operatorname{Tr} G \leq 1 \right\}.$$

□

Lemma 3.13. *Given $S \in M^+(d,\mathbb{R})$ such that $S > 0$,*

$$\min \left\{ \operatorname{Tr} S G^{-1}; G \in M^+(d,\mathbb{R}),\ \operatorname{Tr} G = 1 \right\} = (\operatorname{Tr} \sqrt{S})^2.$$

Only if $G = \sqrt{S}/(\operatorname{Tr} \sqrt{S})$ then $\operatorname{Tr} S G^{-1} = (\operatorname{Tr} \sqrt{S})^2$.

Proof. For $G = (g_{ij})_{1 \leq i,j \leq d}$, let $f(G) := \operatorname{Tr}(S G^{-1}) + \lambda(\operatorname{Tr} G - 1)$ where λ is a Lagrange multiplier. Then

$$\frac{\partial f}{\partial G_{ij}} = \operatorname{Tr} \left[S(-G^{-1} \frac{\partial G}{\partial G_{ij}} G^{-1}) \right] + \lambda \delta_{ij} = - \langle e_j | G^{-1} S G^{-1} | e_i \rangle + \lambda \delta_{ij} = 0$$

where $\{e_i\}_{1 \leq i \leq d}$ is the standard CONS of $\mathbb{R}^d$. Thus

$$G^{-1} S G^{-1} \;=\; \lambda I$$

from which

$$G \;=\; \frac{\sqrt{S}}{\sqrt{\lambda}}$$

and

$$\lambda = \left(\operatorname{Tr}\sqrt{S}\right)^2$$

because of $\operatorname{Tr} G = 1$. As a consequence

$$\min_G \operatorname{Tr}(SG^{-1}) = \operatorname{Tr}(\lambda G) = \lambda = (\operatorname{Tr}\sqrt{S})^2.$$

□

Proof of Proposition 3.1. According to Lemma 3.12 and Lemma 3.13,

$$\begin{aligned}\min_{M\in\mathcal{M}(\mathcal{H})} \operatorname{Tr} H_\theta\, g_\theta(M)^{-1} &= \min\{\operatorname{Tr} H_\theta \sqrt{J_\theta^{-1}} G^{-1} \sqrt{J_\theta^{-1}} \mid G \in M^+(d,\mathbb{R}), \operatorname{Tr} G = 1\} \\ &= \min\{\operatorname{Tr} \sqrt{J_\theta^{-1}} H_\theta \sqrt{J_\theta^{-1}} G^{-1} \mid G \in M^+(d,\mathbb{R}), \operatorname{Tr} G = 1\} \\ &= (\operatorname{Tr} R_\theta)^2 .\end{aligned}$$

When $\operatorname{Tr} H_\theta\, g_\theta(M)^{-1}$ achieves the minimum,

$$G = \frac{R_\theta}{\operatorname{Tr} R_\theta}.$$

thus

$$g_\theta(M) = \sqrt{J_\theta} G \sqrt{J_\theta} = \frac{\sqrt{J_\theta} R_\theta \sqrt{J_\theta}}{\operatorname{Tr} R_\theta}.$$

□

Proof of Proposition 3.2. Assume that $d = 3$. According to (3.24) and Lemma 3.10,

$$\begin{aligned} g_\theta(M) &= \sqrt{J_\theta} U \hat{g}_\theta(M) U^{-1} \sqrt{J_\theta} = \sqrt{J_\theta} U \{p_1 \hat{g}_\theta(M^{(1)}) + p_2 \hat{g}_\theta(M^{(2)}) + p_3 \hat{g}_\theta(M^{(3)})\} U^{-1} \sqrt{J_\theta} \\ &= \sqrt{J_\theta} U \{p_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + p_3 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\} U^{-1} \sqrt{J_\theta} \\ &= \sqrt{J_\theta} U \frac{S}{\operatorname{Tr} S} U^{-1} \sqrt{J_\theta} = \sqrt{J_\theta} \frac{USU^{-1}}{\operatorname{Tr} S} \sqrt{J_\theta} = \frac{\sqrt{J_\theta} R_\theta \sqrt{J_\theta}}{\operatorname{Tr} R_\theta}. \end{aligned}$$

When $d = 1$ or 2, we can prove this in a similar way. □

Appendix 3.B Rotationally symmetric weight

In this appendix, we show that any rotationally symmetric weight is represented in the form

$$H_x := f(r) I + (g(r) - f(r)) \frac{1}{r^2} |x\rangle \langle x| , \tag{3.29}$$

for $x \neq 0$ where f, g are functions on $(0,1)$ such that $f(r) > 0$ and $g(r) > 0$.

Given $x \in \mathcal{X}$ $(x \neq 0)$ arbitrarily, let e_1, e_2, e_3 be an orthonormal basis of $\mathbb{R}^3$ with $e_3 = \frac{|x\rangle}{|x|}$, and let $V \in SO(3)$ be any rotation about e_3-axis. Since

$$V^* H_x V = V^* H_{(Vx)} V = H_x, \tag{3.30}$$

H_x and V are simultaneously diagonalized, and e_3 is one of their common eigenvectors. Other eigenvalues of H_x must be degenerate because V is any rotation about e_3-axis. Then H_x should be represented as

$$\begin{aligned} H_x &= \hat{f}(x) |e_1\rangle \langle e_1| + \hat{f}(x) |e_2\rangle \langle e_2| + \hat{g}(x) |e_3\rangle \langle e_3| \\ &= \hat{f}(x) I + (\hat{g}(x) - \hat{f}(x)) \frac{1}{r^2} |x\rangle \langle x| . \end{aligned} \tag{3.31}$$

Let $U \in SO(3)$ be any rotation. It follows that

$$\begin{aligned} U^* H_{(Ux)} U &= U^* \left[\hat{f}(Ux) I + (\hat{g}(Ux) - \hat{f}(Ux)) \frac{1}{r^2} |Ux\rangle \langle Ux| \right] U \\ &= \hat{f}(Ux) I + (\hat{g}(Ux) - \hat{f}(Ux)) \frac{1}{r^2} |U^* Ux\rangle \langle U^* Ux| \\ &= \hat{f}(Ux) I + (\hat{g}(Ux) - \hat{f}(Ux)) \frac{1}{r^2} |x\rangle \langle x| . \end{aligned} \tag{3.32}$$

We see that it follows $\hat{f}(x) = \hat{f}(Ux)$ and $\hat{g}(x) = \hat{g}(Ux)$ for any $U \in SO(3)$ by comparing (3.31) and (3.32). Therefore $\hat{f}$ and $\hat{g}$ must be represented by $\hat{f}(x) = f(|x|)$ and $\hat{g}(x) = g(|x|)$.

Appendix 3.C Bures distance and quantum Fisher information matrix

The Bures distance between two states ρ and σ is defined by

$$B(\rho, \sigma) := 4 \left(1 - \mathrm{Tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right) .$$

It is known that

$$B(\tau_x, \tau_{x+dx}) = \frac{1}{2} \sum_{ij} J_{x,ij} dx^i dx^j + O(|dx|^3) \tag{3.33}$$

when $|dx|$ is sufficiently small. Given an estimator $(M, \hat{x})$ that is locally unbiased at $x_0 \in \mathcal{X}$, it follows from (3.33) that

$$\begin{aligned} \mathrm{Tr}\, J_{x_0} V_{x_0}[M, \hat{x}] &= E_{x_0}[M, \sum_{ij} J_{x_0,ij} (\hat{x}^i - x_0^i)(\hat{x}^j - x_0^j)] \\ &= E_{x_0}[M, 2B(\tau_{x_0}, \tau_{\hat{x}}) + O(|\hat{x} - x_0|^3)]. \end{aligned}$$

Chapter 4

Experimental Demonstration of Adaptive Quantum State Estimation

Abstract

The first experimental demonstration of an adaptive quantum state estimation (AQSE) is reported. The strong consistency and asymptotic efficiency of AQSE have been mathematically proven [J. Phys. A:Math. Gen. 39 12489 (2006)]. In this chapter, the angle of linear polarization of single photons, or the phase parameter between the right and the left circularly polarization, is estimated using AQSE, and the strong consistency and asymptotic efficiency are experimentally verified. AQSE will provide a general useful method in both quantum information processing and metrology.

4.1 Motivation

Quantum theory is inherently statistical. This entails repetition of experiments over a number of identically prepared quantum objects, for example, quantum states, if one wants to know the "true state" or the "true value" of the parameter that specifies the quantum state [29, 42, 3, 22]. Such an estimation procedure is particularly important for quantum communication and quantum computation [39], and is also indispensable to quantum metrology [14, 38, 41, 48, 30]. In applications, one needs to design the estimation procedure in such a way that the estimated value of the parameter should be close to the true value (consistency), and that the uncertainty of the estimated value should be as small as possible (efficiency) for a given limited number of samples. In order to realize these requirements, Nagaoka advocated an adaptive quantum state estimation (AQSE) procedure [36], and recently Fujiwara proved the strong consistency and asymptotic efficiency for AQSE [9].

In this chapter, we report the first experimental demonstration of AQSE using photons. The angle of a half wave plate (HWP) that initializes the linear polarization of input photons is estimated using AQSE. A sequence of AQSE is carried out with 300 input photons, and the sequence is repeated 500 times for four different settings of HWP. The statistical analysis of these results verifies the strong consistency and asymptotic efficiency of AQSE. Recently, it has been mathematically proven that the precision of AQSE outperforms the conventional state tomography [49]. It is thus expected that AQSE will provide a useful methodology in the broad area of quantum information processing, communication, and metrology.

4.2 Adaptive Quantum State Estimation

Let us first explain AQSE in detail. For simplicity, we restrict ourselves to one-dimensional *quantum statistical model* $\mathcal{S} = \{\rho_\theta;\, \theta \in \Theta\,(\subset \mathbb{R})\}$, a smooth parametric family of density operators on a Hilbert space $\mathcal{H}$ having a one-dimensional parameter θ. Our aim is to estimate the true value of θ by means of a certain quantum estimation scheme. An *estimator* is represented by a pair $(M, \check{\theta})$, where $M = \{M(x);\, x \in \mathcal{X}\}$ is a positive operator-valued measure (POVM) that takes values on a set $\mathcal{X}$, and $\check{\theta} : \mathcal{X} \to \Theta$ is a map that gives the estimated value $\check{\theta}(x)$ from each observed data $x \in \mathcal{X}$. The observed data $x \in \mathcal{X}$ has probability density

$$f(x; \theta, M) := \operatorname{Tr} \rho_\theta M(x), \tag{4.1}$$

which depends on both the parameter θ and the measurement M.

In traditional statistics, it is often the case to confine our attention to unbiased estimators. An estimator $(M, \check{\theta})$ is called *unbiased* if

$$E_\theta[M, \check{\theta}] = \theta \tag{4.2}$$

is satisfied for all $\theta \in \Theta$, where $E_\theta[\,\cdot\,]$ denotes the expectation with respect to the density (4.1). It is well known [21] that an unbiased estimator $(M, \check{\theta})$ satisfies the quantum Cramér-Rao inequality $V_\theta[M, \check{\theta}] \geq (J_\theta)^{-1}$, where $V_\theta[\,\cdot\,]$ denotes the variance, and J_θ is the quantum Fisher information of the model $\mathcal{S}$ defined by $J_\theta := \operatorname{Tr} \rho_\theta L_\theta^2$, where L_θ is the symmetric logarithmic derivative (SLD) defined by the self-adjoint operator satisfying the equation $\frac{d\rho_\theta}{d\theta} = \frac{1}{2}\left(L_\theta \rho_\theta + \rho_\theta L_\theta\right)$.

In quantum statistics, however, it is regarded that unbiasedness is too restrictive a requirement, and we usually weaken the condition to a "local" one. An estimator $(M, \check{\theta})$ is called *locally unbiased* [24] at a given point $\theta_0 \in \Theta$ if the condition (4.2) is satisfied around $\theta = \theta_0$ up to the first order of the Taylor expansion, that is, if $E_{\theta_0}[M, \check{\theta}] = \theta_0$ and $\frac{d}{d\theta} E_\theta[M, \check{\theta}]\big|_{\theta=\theta_0} = 1$ hold. Clearly, an estimator is unbiased if and only if it is locally unbiased at all $\theta \in \Theta$. A crucial observation is that an estimator $(M, \check{\theta})$ that is locally unbiased at θ_0 also satisfies the quantum Cramér-Rao inequality

$$V_{\theta_0}[M, \check{\theta}] \geq (J_{\theta_0})^{-1} \tag{4.3}$$

at $\theta = \theta_0$, and that the lower bound in (4.3) is achievable for any one-dimensional quantum statistical model $\mathcal{S}$. To put it differently, the best locally unbiased estimator (LUE) for the parameter θ at $\theta = \theta_0$ is the one that satisfies $V_{\theta_0}[M, \check{\theta}] = (J_{\theta_0})^{-1}$.

Here we encounter a difficulty which often becomes the target of criticism: since the best LUE for estimating the parameter θ depends, in general, on the unknown parameter θ itself, the estimation strategy based on LUEs would be infeasible. In a different yet analogous context, Cochran [5] ingeniously described this kind of dilemma as follows: "You tell me the value of θ and I promise to design the best experiment for estimating θ."

To surmount this difficulty, Nagaoka [36] advocated an adaptive quantum state estimation (AQSE) scheme as follows. Suppose that, by prior investigation of the quantum statistical model $\mathcal{S}$, one has the list of optimal LUEs $\big(M(\,\cdot\,;\theta), \check{\theta}(\,\cdot\,;\theta)\big)$ for each $\theta \in \Theta$. One begins with an arbitrary initial guess $\hat{\theta}_0 \in \Theta$, and applies the measurement $M(\,\cdot\,;\hat{\theta}_0)$ that is optimal at $\hat{\theta}_0$. Suppose the data x_1 is observed, one then applies the maximum likelihood method to the likelihood function $L_1(\theta) = f(x_1; \theta, M(\,\cdot\,;\hat{\theta}_0))$, to obtain the next guess $\hat{\theta}_1$. At stage n (≥ 2), one applies the measurement $M(\,\cdot\,;\hat{\theta}_{n-1})$, where $\hat{\theta}_{n-1}$ is the maximum likelihood estimator (MLE) obtained at the previous stage. The likelihood function is then given by $L_n(\theta) := \prod_{i=1}^n f(x_i; \theta, M(\,\cdot\,;\hat{\theta}_{i-1}))$, where x_i is the observed data at stage i, and one obtains the nth MLE $\hat{\theta}_n$ that maximizes $L_n(\theta)$. It is quite natural to expect that the sequence $\hat{\theta}_n$ of MLEs would converge to the true value of the parameter θ. In fact, under certain regularity conditions, it can be shown that the sequence $\hat{\theta}_n$ is strongly consistent and asymptotically efficient [9].

4.3 Experimental setup

Now let us discuss the implementation of AQSE using photons (Fig. 1). Here the unknown parameter is the angle θ of HWP0, which determines the phase ϕ between right and left circularly

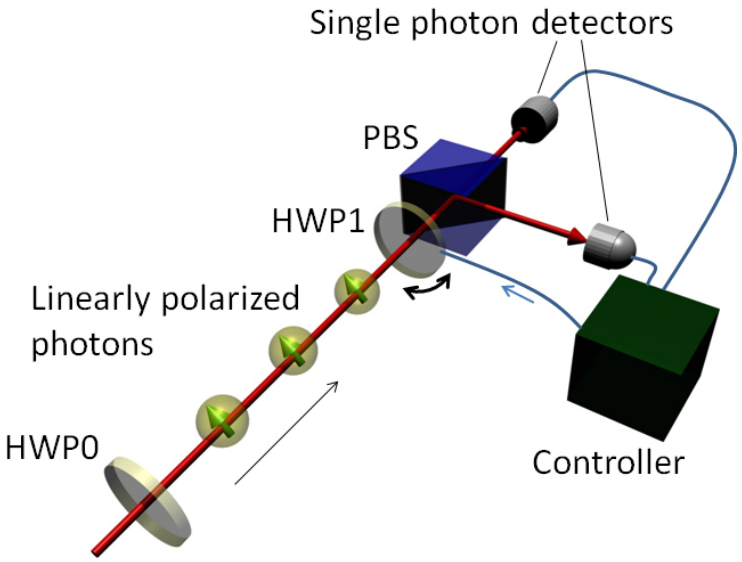

Figure 4.1: Schematic of adaptive quantum state estimation. Photons are linearly polarized with a polarization direction determined by HWP0. The polarization is analyzed by HWP1 and the polarizing beam splitter (PBS). The controller sets HWP1 to an angle calculated on the basis of the photon measurement results.

polarizations of input photons by the relation $\phi = 4\theta$. An arbitrary linear polarization can be described using right and left circular polarizations as follows:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|R\rangle + e^{i\phi}|L\rangle) = \cos(\frac{\phi}{2})|H\rangle + \sin(\frac{\phi}{2})|V\rangle . \tag{4.4}$$

By changing the angle of the half wave plate (HWP1), we can adjust the measurement basis. For such measurement, the POVM having optimal estimation capability is given by

$$M(\theta) = (M(1;\theta), M(2;\theta)) = (|\xi\rangle\langle\xi|, I - |\xi\rangle\langle\xi|), \tag{4.5}$$

where $\langle\xi| = \left(\cos\left(2\theta + \frac{\pi}{4}\right), \sin\left(2\theta + \frac{\pi}{4}\right)\right)$. By applying the POVM $M(\theta)$ to the input state $|\psi(\theta)\rangle := |\psi\rangle$, one obtains the probability distribution on $\mathcal{X} := \{1, 2\}$ which is isomorphic to the fair coin flipping.

The drawback to realizing this measurement is that the optimal POVM $M(\theta)$ depends on the unknown value of the parameter θ[1]. We can avoid this drawback by adopting an AQSE as follows. We begin by setting the initial log-likelihood function to be $l_0(\theta) = 0$, and then start inputting and detecting photons one by one. For nth photon, we apply the measurement $M(\hat{\theta}_{n-1})$ which depends on the latest MLE $\hat{\theta}_{n-1}$. Let $x_n \in \mathcal{X}$ be the outcome indicating which detector has been lit. The log-likelihood function is then updated by the formula

$$l_n(\theta) := l_{n-1}(\theta) + \log\langle\psi(\theta)|M(x_n;\hat{\theta}_{n-1})|\psi(\theta)\rangle , \tag{4.6}$$

and the nth MLE is given by $\hat{\theta}_n = \arg\max_\theta l_n(\theta)$. Let us denote the true value of the parameter θ by θ^t. It is known [9] that the sequence $\hat{\theta}_n$ of MLEs converges to the true value θ^t with probability one (strong consistency) and that the distributions of the random variables $\sqrt{n}(\hat{\theta}_n - \theta^t)$ converge to the normal distribution $N(0, J_{\theta^t}^{-1})$ (asymptotic efficiency), where J_θ denotes the quantum Fisher information of the parameter θ, which turns out to be 16 for our model (4.4).

[1]Note that any fixed POVM of the form (4.5) is optimal for almost all values of the parameter ϕ if we treat only the pure state model (4.4). However, if we treat mixed state models, the dependence of optimal POVM to the parameter becomes crucial [41].

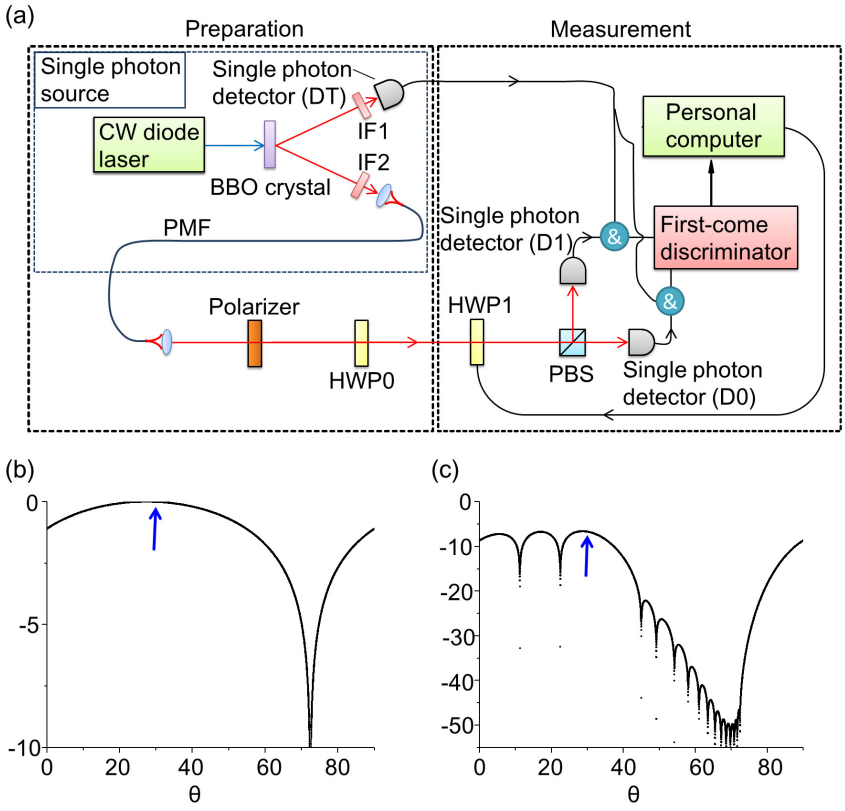

Figure 4.2: (a) Schematic of the experimental setup. (b)(c) An example showing the update of a log-likelihood function. The second term $\log \langle\psi(\theta)| M(x_n; \hat{\theta}_{n-1}) |\psi(\theta)\rangle$ in eq. (4.6) is shown in panel (b), and the updated $l_n(\theta)$ is shown in panel (c). The blue arrows indicate the true value θ^t.

The experimental setup is shown in Fig. 2(a). Single photons at 780nm are generated from a heralded single photon source [25], consisting of a CW diode pump laser (wavelength: 402 nm) and a 3 mm long BBO crystal (Type I). A pair of a signal photon (780 nm) and a trigger photon (830 nm) is created via spontaneous parametric down conversion. The detector (DT, SPCM-AQR, Perkin Elmer) after an interference filter (IF1, center wavelength 830nm) outputs an electric pulse (width 30ns) when it detects a trigger photon and the electric pulse heralds the generation of a signal photon, which is coupled to a polarization maintaining fiber (PMF) after an interference filter (IF2, center wavelength 780 nm, width 4 nm). The polarization of photons are then initialized to be horizontal using a polarizer (extinction ratio 10^{-5}). The target parameter θ^t was set using HWP0. The polarization state of the photon was analyzed by HWP1 and a polarizing beam splitter (PBS). After passing through the PBS, photons are guided to single photon detectors (D0 and D1, SPCM-AQR, Perkin Elmer) on each PBS output port. The outputs of single photon detectors are gated by the rise of the heralding signal and connected to the "first-come discriminator," consisting of a home-made electric circuit. When the discriminator receives the first signal from one of the detectors (D0 or D1) after the measurement for $(n-1)$th photon starts, the discriminator informs which detector has been clicked. The minimum pulse interval of 2.5ns can be discriminated. Note that the discriminator ignores the case when it receives the pulses from both the detectors within 2.5ns. The angle of HWP1 for measuring the nth photon is determined by calculating the discretized MLE $\hat{\theta}_n$, the maximizer of the log-likelihood function (4.6) chosen from among the 10000 points that divide the domain $[0, \pi/2)$ of the parameter θ into equal parts (Figs. 2(b) and 2(c)). When the change of HWP1 angle is completed, the measurement for the next (nth) photon will be started. In a sequence of AQSE, the above mentioned procedure is carried out up to 300 input photons (n=300). For four different HWP0 angles $\theta = 0, 30, 60$, and 78.3 [deg], we repeated the sequence for 500 times (r=500).

Let us first observe the strong consistency for the sequence $\hat{\theta}_n$ of MLEs for the parameter θ of HWP0. Fig. 3 (a) shows 500 trajectories of estimated HWP0 angle $\hat{\theta}_n$ against the number n of photons when the true value θ^t of the parameter is set to be 60 degree. The curves correspond to independent runs of adaptive estimation. Evidently, each curve of $\hat{\theta}_n$ approaches the true value θ^t, which is in accord with the mathematical result that $\hat{\theta}_n \to \theta^t$ almost surely as $n \to \infty$, even though the curves are dissimilar to each other reflecting the genuine statistical nature of quantum system. The convergence to the true value is clear in Fig. 3(b) where first 10 trajectories in Fig. 3(a) are superposed.

4.4 Experimental results

We next test the hypothesis that the MLE $\hat{\theta}_n$ follows a normal distribution for large n. More concretely, we will investigate if the random variable $\sqrt{nJ_\theta}\,(\hat{\theta}_n - \overline{\theta})$ follows the standard normal distribution $N(0,1)$, i.e., $\sqrt{nJ_\theta}\,(\hat{\theta}_n - \overline{\theta}) \sim N(0,1)$, where $\overline{\theta}$ is the sample average of MLEs $\hat{\theta}_n$ over sufficiently many independent trials. A goodness of fit test [23] was carried out as follows:

1) The real axis was divided into 23 intervals (bins) $\{I_b\}_{b=0}^{22}$, where $I_1, \ldots, I_{21}$ are disjoint partitions of the interval $[-3.5, 3.5]$ of equal width, and $I_0 = (-\infty, -3.5)$, $I_{22} = (3.5, +\infty)$. In reality, these bins were slightly shifted by $\delta/10000$, where $\delta := \sqrt{nJ_\theta}\,\pi/20000$ is the scaled resolution of the estimator $\hat{\theta}_n$, so that the data $\sqrt{nJ_\theta}\,(\hat{\theta}_n - \overline{\theta})$ did not fall on the boundaries of the bins.

2) The test-statistic $X^2 := \sum_{b=0}^{22} \frac{(N_b - r\,p_b)^2}{r\,p_b}$ was calculated, where N_b is the number of observed data which fell into bth bin, p_b the theoretical probability of falling a datum into bth bin under the null hypothesis $N(0,1)$, and r the number of repetitions of adaptive estimation procedure.

3) The test-statistic X^2 was analyzed using the chi-square distribution χ^2_{23-p} of degree $23-p$, where $p = 2$ degrees of freedom ought to be subtracted because of the normalization and the use of sample average $\overline{\theta}$.

Figure 4.4 shows the histogram of the observed data obtained by $r = 500$ independent experiments of adaptive estimation scheme, each using $n = 300$ photons. The true values θ^t of

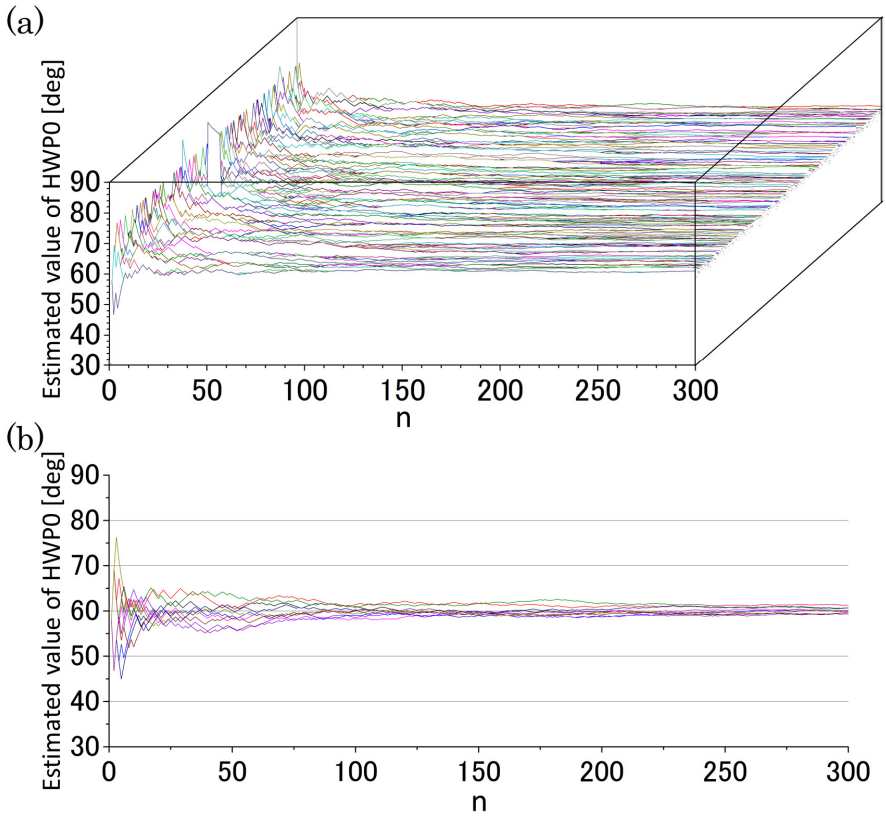

Figure 4.3: (a) Trajectories of estimated HWP0 angles against the number n of photons for $r = 500$ repetitions is shown in a three dimensional plot. (b) The first 10 curves are superposed in a two dimensional graph.

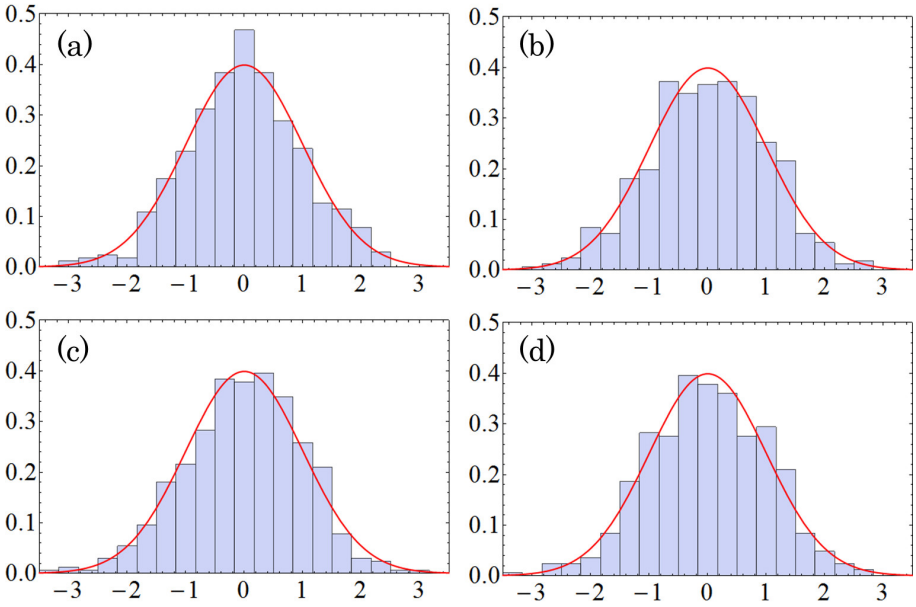

Figure 4.4: Histogram of the observed data obtained by $r = 500$ independent experiments of adaptive estimation scheme, each using $n = 300$ photons. These histograms were taken for four different true values of (a) 0 [deg], (b) 30 [deg], (c) 60 [deg] and (d) 78.3 [deg].

Table 4.1: Confidence intervals for the mean μ and the variance v. CL means confidence level.

θ^t [deg]	μ [deg] (90% CL)	v (90% CL)
0.0	-0.15 ± 0.06	[0.054, 0.067]
30.0	29.90 ± 0.06	[0.055, 0.067]
60.0	60.00 ± 0.06	[0.056, 0.068]
78.3	78.27 ± 0.06	[0.055, 0.068]

the parameter θ of HWP0 are set to be 0, 30, 60, and 78.3 degrees. The density function of the standard normal distribution $N(0,1)$ is also plotted as the solid curve. All the experimental data agree with the standard normal distribution. To be precise, the values of the test statistic X^2 are (a) 16.8 (b) 15.7 (c) 12.8 (d) 16.2, and the null hypothesis is accepted with 10% significance level in each case.

Having obtained the strong evidence that the distribution of the MLE has converged quite well to a normal distribution at $n = 300$, we finally proceed to the estimation of confidence intervals [23] for the mean μ and variance v, assuming that $\sqrt{n}\,(\hat{\theta}_n - \mu) \sim N(0, v)$. The confidence intervals for μ and v are obtained by the standard procedure based on the statistical laws that $\sqrt{\frac{r}{\overline{V}}}\,(\overline{\theta} - \mu) \sim T_{r-1}$ and $\frac{r-1}{(v/n)}\,\overline{V} \sim \chi^2_{r-1}$. Here $\overline{V}$ is the unbiased variance of MLEs $\hat{\theta}_n$ over r trials, and T_{r-1} the t-distribution of degree $r-1$.

Table 1 summarizes the results for $r = 500$ with 90% confidence level. Recall that the asymptotic efficiency asserts that $\mu \simeq \theta^t$ and $v \simeq J^{-1}_{\theta^t} (= 0.0625)$. Since the precision of the present experiment is about ± 0.2 degree[2], we conclude that the estimated values of μ and v listed in Table I are in excellent agreement with the theoretical values.

It should be noted that the purpose of our AQSE is completely different from 'adaptive measurements' proposed by Berry and Wiseman [2]. Their scheme was devised to estimate

[2]The precision of the rotation stage for HWP1 and the accuracy of the polarization basis states limited the total precision of the experimental setup to ±0.2 degree.

the phase difference between the two arms of an interferometer using a special N-photon two-mode state, approximating the canonical measurement proposed by Sanders and Milburn [46], and is not applicable to general quantum state estimation problems. By contrast, our AQSE is a general-purpose estimation scheme applicable to any quantum statistical model using n identical copies of an unknown state. AQSE may also be used in verifying the achievability of the Cramér-Rao version of the Heisenberg limit $O(1/N^2)$ [27] by applying the scheme to the n-i.i.d. extension $\rho_\theta^{\otimes n}$ of an N-photon phase-shift model ρ_θ on $\mathcal{H} \simeq (\mathbb{C}^2)^{\otimes N}$. (See also [10] for estimating a unitary channel under noise.) Incidentally, AQSE is based on the Cramér-Rao type point estimation theory and is free from the choice of *a priori* distribution which matters in Bayesian statistics such as adaptive Bayesian quantum tomography [26].

4.5 Concluding remarks

In summary, we have verified both the strong consistency and asymptotic efficiency of AQSE by experimentally estimating the angle of linear polarization of photons. Since AQSE has been mathematically proven to outperform the conventional estimation scheme such as the state tomography [49], we plan to apply AQSE to multi-parameter cases and compare the performance with other protocols using fixed measurement basis [4]. It will also be intriguing to apply AQSE to enhance the performance of quantum metrological experiments beating the standard quantum limit [14, 38, 41, 48].

Chapter 5

Conclusions

In the present dissertation, we explored asymptotic quantum state estimation theory and its applications.

We first investigated the ultimate limit of estimation precision for the case when any collective measurements are available. We developed a theory of quantum local asymptotic normality based on a new quantum log-likelihood ratio, which is applicable to any quantum statistical models satisfying mild regularity conditions. We also derived a quantum analogue of Le Cam's third lemma, and proved the asymptotic achievability of the Holevo bound for the local shift parameter on a dense subset of the parameter space. There are of course many open problems left. Among others, extending the representation theorem, convolution theorem, and local asymptotic minimax theorem to a quantum statistical framework would be the most important ones to be addressed. The difficulty with those problems lies in the fact that many standard tools in the classical statistics do not work in a quantum case. For example, suppose that random variables X_n converge in distribution to a random variable X, and that Y_n converge in distribution to a constant c, then the pairs (X_n, Y_n) converge to (X, c) in distribution. However its obvious extension to a quantum case is not always true. Before tackling the above mentioned open problems, we need to establish a theory of "quantum convergence in law."

We next investigated a more realistic situation in which only separable measurements are available. We scrutinized the case when $\dim \mathcal{H} = 2$, and showed that the quantum state tomography is optimal if and only if a physically unnatural weight is adopted. Unfortunately, we do not know anything definitive about the optimality of estimators when $\dim \mathcal{H} \geq 3$, although numerical evaluation of the minimal values of the weighted covariance matrices is possible as in Figure 3.4. Incidentally, investing the theory of quantum local asymptotic normality for separable measurements, or even for a given restricted class of measurements, would be an important subject from the viewpoint of applications.

We further reported the first experimental demonstration of an adaptive quantum state estimation (AQSE). The angle of linear polarization of single photons, or the phase parameter between the right and the left circularly polarization, was estimated using AQSE, and the strong consistency and asymptotic efficiency were experimentally verified. Experimental demonstration of AQSE for two or three dimensional qubit models is now in progress and will be reported elsewhere.

Bibliography

[1] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury and F. Vatan, "A new proof of the existence of mutually unbiased bases," *Algorithmica*, **34**, 512 (2002).

[2] D. W. Berry and H. M. Wiseman, "Optimal states and almost optimal adaptive measurements for quantum interferometry," Phys. Rev. Lett. **85**, 5098 (2000).

[3] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, "Optimal quantum tomography of states, measurements, and transformations," Phys. Rev. Lett. **102**, 010404 (2009).

[4] Yu. I. Bogdanov, G. Brida, M. Genovese, S. P. Kulik, E. V. Moreva, and A. P. Shurupov, "Statistical estimation of the efficiency of quantum state tomography protocols," Phys. Rev. Lett. **105**, 010404 (2010).

[5] W. G. Cochran, "Experiments for nonlinear functions," Journal of the American Statistical Association, vol. 68, 771 (1973).

[6] A. Fujiwara and H. Nagaoka, "Quantum Fisher metric and estimation for pure state models," Phys. Lett. A 201, 119 (1995).

[7] A. Fujiwara and H. Nagaoka, "An estimation theoretical characterization of coherent states," J. Math. Phys., vol. 40, 4227 (1999).

[8] A. Fujiwara, "Geometry of quantum information systems," in *Geometry in Present Day Sciences*, ed. O. E. Barndorff-Nielsen and E. B. V. Jensen (World Scientific, Singapore, 1999) p. 35.

[9] A. Fujiwara, "Strong consistency and asymptotic efficiency for adaptive quantum estimation problems," J. Phys. A: Math. Gen., **39**, 12489 (2006); "Corrigendum," J. Phys. A: Math. Theor, **44** 079501 (2011).

[10] A. Fujiwara and H. Imai, "A fibre bundle over manifolds of quantum channels and its application to quantum statistics," J. Phys. A: Math. Theor., **41**, 255304 (2008).

[11] R. D. Gill and M. Guţă, "On asymptotic quantum statistical inference," *IMS Collections From Probability to Statistics and Back: High-Dimensional Models and Processes* Vol. 9, 105 (2012).

[12] R. D. Gill and B. Y. Levit, "Applications of the Van Trees inequality: A Bayesian Cramér-Rao bound," Bernoulli **1**, 59 (1995).

[13] R. D. Gill and S. Massar, "State estimation for large ensembles," *Phys. Rev. A*, **61**, 042312 (2000).

[14] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum-enhanced measurements: beating the standard quantum limit," Science **306**, 13330 (2004).

[15] M. Guţă and J. Kahn, "Local asymptotic normality for qubit states," Phys. Rev. A, 73:052108 (2006).

[16] M. Guţă and A. Jençová, "Local asymptotic normality in quantum statistics," Commun. Math. Phys., 276, 341 (2007).

[17] M. Hayashi, "A linear programming approach to attainable Cramér-Rao type bounds," in *Quantum Communication, Computing, and Measurement*, ed. by Hirota et al. (Plenum, NY, 1997), p. 99.

[18] M. Hayashi, *Quantum Information: An Introduction* (Springer Verlag, Berlin-Heidelberg, 2006).

[19] M. Hayashi and K. Matsumoto, "Asymptotic performance of optimal state estimation in qubit system," J. Math. Phys. **49**, 102101 (2008).

[20] C. W. Helstrom, "Minimum mean-square error of estimates in quantum statistics," Phys. Lett. 25A, 101 (1967).

[21] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[22] A. Hentschel and B. C. Sanders, "Efficient algorithm for optimizing adaptive quantum metrology processes," Phys. Rev. Lett. **107**, 233601 (2011).

[23] R. V. Hogg, J. W. McKean, A. T. Craig, *Introduction to Mathematical Statistics* (Pearson Prentice Hall, Upper Saddle River, NJ, 2005).

[24] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[25] C. K. Hong and L. Mandel, "Experimental realisation of a localized one-photon state," Phys. Rev. Lett. **56**, 58 (1986).

[26] F. Huszár and N. M. T. Houlsby, "Adaptive Bayesian quantum tomography," Phys. Rev. A **85**, 052120 (2012).

[27] H. Imai and A. Fujiwara, "Geometry of optimal estimation scheme for $SU(D)$ channels," J. Phys. A: Math. Theor., **40**, 4391 (2007).

[28] V. Jaksić, Y. Pautrat, and C.-A. Pillet, "A quantum central limit theorem for sums of independent identically distributed random variables," J. Math. Phys. **51**, 015208 (2010).

[29] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, "On the measurement of qubits," Phys. Rev. A **64**, 052312 (2001).

[30] J. A. Jones, S. D. Karlen, J. Fitzsimons, A. Ardavan, S. C. Benjamin, G. A. D. Briggs and J. J. L. Morton, "Magnetic field sensing beyond the standard quantum limit using 10-spin NOON states," Science **324** 1166 (2009).

[31] J. Kahn and M. Guţă , "Local asymptotic normality for finite dimensional quantum systems," Commun. Math. Phys. 289, 597 (2009).

[32] G. Kuperberg, "A tracial quantum central limit theorem," Trans. Am. Math. Soc. **357**, 459 (2005).

[33] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*, (Springer Verlag, New York, 1986).

[34] J. Manuceau, M. Sirugue, D. Testard, and A. Verbeure, "The smallest C-algebra for canonical commutations relations," Commun. Math. Phys. **32**, 231 (1973).

[35] K. Matsumoto, "A new approach to the Cramer-Rao type bound of the pure state model," J. Phys. A **35**, 3111 (2002).

[36] H. Nagaoka, "An asymptotically efficient estimator for a one-dimensional parametric model of quantum statistical operators," in Proc. Int. Symp. on Inform. Theory, p. 198 (1988); H. Nagaoka, "On the parameter estimation problem for quantum statistical models," in Proc. 12th Symp. on Inform. Theory and its Appl., pp. 577 (1989), reprinted in *Asymptotic Theory of Quantum Statistical Inference*, ed. M. Hayashi (World Scientific, Singapore, 2005), p. 125.

[37] H. Nagaoka, "A generalization of the simultaneous diagonalization of Hermitian matrices and its relation to quantum estimation theory," *Transactions of the Japan Society for Industrial and Applied Mathematics*, **1**, 305 (1991) (in Japanese).

[38] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, S. Takeuchi, "Beating the standard quantum limit with four entangled photons," Science **316**, 726 (2007).

[39] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[40] M. Ohya and D. Petz, *Quantum Entropy and its Use.* (Springer Verlag, Berlin-Heidelberg, 2004).

[41] R. Okamoto, H. F. Hofmann, T. Nagata, J. L. O'Brien, K. Sasaki, S. Takeuchi, "Beating the standard quantum limit: phase super-sensitivity of $N-$photon interferometers," New J. Phys. **10**, 073033 (2008).

[42] M. G. A. Paris, "Quantum estimation for quantum technology," Int. J. Quant. Inform. **7**, 125 (2009).

[43] D. Petz, *Quantum Information Theory and Quantum Statistics.* (Springer Verlag, Berlin-Heidelberg, 2010).

[44] D. Petz, *An Invitation to the Algebra of Canonical Commutation Relations*, Leuven Notes in Mathematical and Theoretical Physics. Series A: Mathematical Physics Vol. 2 (Leuven University Press, Leuven, 1990).

[45] D. Petz, "Monotone metrics on matrix spaces," *Linear Algebra and its Applications*, **244**, 81 (1996).

[46] B. C. Sanders and G. J. Milburn, "Optimal quantum measurements for phase estimation," Phys. Rev. Lett. **75**, 2944 (1995).

[47] A. W. van der Vaart, *Asymptotic Statistics.* (Cambridge University Press, Cambridge, 1998).

[48] G. Y. Xiang, B. L. Higgins, D. W. Berry, H. M. Wiseman and G. J. Pryde, "Entanglement-enhanced measurement of a completely unknown optical phase," Nature Photonics **5**, 43 (2011)

[49] K. Yamagata, "Efficiency of quantum state tomography for qubits," Int. J. Quant. Inform., **9** 1167 (2011).

[50] H. P. Yuen and M. Lax, "Multiple-parameter quantum estimation and measurement of non-selfadjoint observables," IEEE Trans. Inform. Theor., 19, 740 (1973).