Operational meanings of a generalized conditional expectation in quantum metrology

Mankei Tsang[1,2]

[1]Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583

[2]Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551

A unifying formalism of generalized conditional expectations (GCEs) for quantum mechanics has recently emerged, but its physical implications regarding the retrodiction of a quantum observable remain controversial. To address the controversy, here I offer operational meanings for a version of the GCEs in the context of quantum parameter estimation. When a quantum sensor is corrupted by decoherence, the GCE is found to relate the operator-valued optimal estimators before and after the decoherence. Furthermore, the error increase, or regret, caused by the decoherence is shown to be equal to a divergence between the two estimators. The real weak value as a special case of the GCE plays the same role in suboptimal estimation—its divergence from the optimal estimator is precisely the regret for not using the optimal measurement. For an application of the GCE, I show that it enables the use of dynamic programming for designing a controller that minimizes the estimation error. For the frequentist setting, I show that the GCE leads to a quantum Rao-Blackwell theorem, which offers significant implications for quantum metrology and thermal-light sensing in particular. These results give the GCE and the associated divergence a natural, useful, and incontrovertible role in quantum decision and control theory.

1 Introduction

The conditional expectation is an essential concept in classical probability and statistics [1]. Given some observed data in an experiment, the conditional expectation of a hidden random variable is the best approximation of the hidden variable in a least-squares sense and thus plays a central role in Bayesian estimation theory [1, 2]. Another important application is in the Rao-Blackwell theorem [3, 4], which exploits the variance reduction property of the conditional expectation to improve an estimator and has found widespread uses in classical statistics [5, 6, 7].

Many attempts have been made over the past few decades to generalize the concept of conditional expectation for quantum mechanics [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. Umegaki's version for von Neumann algebra may be the earliest [8]. His axiomatic definition is so restrictive, however, that his conditional expectation does not exist in many situations [19, 12]; this existence problem has led Holevo to remark that "conditional expectations play a less important part in quantum than in classical probability" [19]. In quantum estimation theory, Personick [9] and Belavkin and Grishanin [10] proposed an operator-valued estimator that is optimal for Bayesian parameter estimation and can also be regarded as a quantum conditional expectation. On the other hand, Accardi and Cecchini proposed yet another conditional expectation for von Neumann algebra [11], which became instrumental in Petz's work on quantum sufficient channels [12]. Many other investigations of quantum conditional expectations can be found in the literature on weak values [13, 14], quantum filtering [20, 21, 22], quantum retrodiction [23, 24], and quantum smoothing [25, 26, 27, 28, 29, 30, 16, 17, 18]. In recent years, it has been recognized [16, 17, 18] that many of these quantum conditional expectations can be unified under a mathematical formalism of generalized conditional expectations (GCEs) [15]. The GCE formalism can also be rigorously connected to the concepts of quantum states over time and generalized Bayes rules [31, 32], as shown by Parzygnat and Fullwood [33].

Despite the mathematical progress, the GCEs have provoked fierce debates regarding their physical meaning and usefulness, especially when it comes to the weak values [34, 35, 36, 37, 38, 39, 40]. The debates center on two issues: whether it makes any sense to estimate the value of a quantum observable in the past (retrodiction) and whether the GCEs offer any use in quantum metrology, where quantum sensors are used to estimate classical parameters. This work addresses both issues by demonstrating how a certain version of the GCEs—of which the real weak value is a special case—can play fundamental roles in quantum parameter estimation in both Bayesian and frequentist settings.

When a quantum sensor suffers from decoherence,

Mankei Tsang: mankei@nus.edu.sg,
https://blog.nus.edu.sg/mankei/

I show that the GCE relates the two Personick estimators before and after the decoherence. Moreover, the error increase due to the decoherence, henceforth called the regret, is shown to be equal to a divergence measure between the two estimators. By regarding a suboptimal measurement as a decoherence process, I show that the weak value is a special case of the GCE and its divergence from the Personick estimator is precisely the regret due to the measurement suboptimality. For the frequentist setting, I also propose a quantum Rao-Blackwell theorem based on the GCE.

These fundamental results lead to many significant consequences in quantum metrology. To wit, the Markovian nature of the GCE is shown to enable the use of dynamic programming [41] for optimizing a measurement protocol, while Corollaries 1–6 in this work reveal the monotonicity of the Bayesian error, the optimality of von Neumann measurements in Bayesian and frequentist settings, the optimality of symmetric estimators for symmetric states, the optimality of direct-sum estimators for direct-sum states, and the optimality of photon counting for certain thermal states. A key feature of these optimality results is that they are direct statements about the mean-square errors and are valid for both biased and unbiased estimators, unlike many results based on Cramér-Rao-type bounds, which require heavy assumptions about the estimators and the density operators.

This paper is organized as follows. To set the stage and make the paper self-contained, Sec. 2 reviews the concept of GCEs, emphasizing their significance in minimizing a divergence quantity between two operators at different times [18]. Section 3 presents some fundamental properties of the GCEs that are key to their applications to quantum metrology, including a chain rule (Theorem 1) that gives the GCEs a Markovian property for a sequence of channels and a Pythagorean theorem (Theorem 2) that gives the divergence an additive property. Sections 4 and 5 present the core results of this work, namely, the applications of a version of the GCEs to quantum parameter estimation. This GCE follows a particular operator ordering based on the Jordan product and is shown to play a natural role in quantum estimation theory.

Section 4 studies the role of the GCE in Bayesian quantum parameter estimation, a topic that has received renewed interest in recent years [42, 43, 44]. Within Sec. 4, Sec. 4.1 presents the general relations between the Personick estimators for a sensor under decoherence, Sec. 4.2 shows how they enable the use of dynamic programming in quantum sensor measurement design, and Sec. 4.3 discusses the special case of the real weak value.

Section 5 switches to the frequentist setting and presents the quantum Rao-Blackwell theorem, Theorem 3, in Sec. 5.1. Sections 5.2–5.4 present some

significant consequences of the quantum theorem for quantum metrology, while Sec. 5.5 discusses an application of the theorem to thermal-light sensing.

Section 6 is the conclusion, listing some open problems. Appendix A discusses the complementary concept of quantum prediction. Appendix B reviews the classical conditional expectation to give the quantum formalism a more familiar context. Appendix C gives an explicit formula for the GCE for Gaussian systems. Appendix D defines the von Neumann measurement. Appendix E presents the dynamic-programming algorithm. Appendix F justifies the name of Theorem 3 by deriving the classical Rao-Blackwell theorem from it. Appendix G discusses the differences and relations between the Bayesian and frequentist settings. Appendix H compares this work with some prior works. Appendix I offers an alternative derivation of the quantum U-statistics, first introduced by Guţă and Butucea [45], using Theorem 3. Appendix J contains the more technical proofs.

2 Review of generalized conditional expectations

This section follows Ref. [18] and Chap. 6 in Ref. [15]. Let $\mathcal{O}(\mathcal{H})$ be the space of bounded operators on a Hilbert space $\mathcal{H}$ and $\rho \in \mathcal{O}(\mathcal{H})$ be a density operator. Define an inner product between two operators $A, B \in \mathcal{O}(\mathcal{H})$ and a norm as

$$\langle B, A \rangle_\rho \equiv \operatorname{tr} B^\dagger \mathcal{E}_\rho A, \quad \|A\|_\rho \equiv \sqrt{\langle A, A \rangle_\rho}, \quad (2.1)$$

where $\mathcal{E}_\rho : \mathcal{O}(\mathcal{H}) \to \mathcal{O}(\mathcal{H})$ is a linear, self-adjoint, and positive-semidefinite map with respect to the Hilbert-Schmidt inner product

$$\langle B, A \rangle_{\mathrm{HS}} \equiv \operatorname{tr} B^\dagger A. \quad (2.2)$$

The weighted inner product $\langle \cdot, \cdot \rangle_\rho$ is a generalization of the inner product between two random variables in classical probability theory [1]. Some desirable properties of $\mathcal{E}$ are[1]

$$\mathcal{E}_\rho A = \rho A \text{ if } \rho, A \text{ commute}, \quad (2.3)$$

$$\mathcal{E}_\rho(U^\dagger A U) = U^\dagger \left(\mathcal{E}_{U \rho U^\dagger} A \right) U, \quad (2.4)$$

$$\mathcal{E}_{\rho_1 \otimes \rho_2}(A_1 \otimes A_2) = (\mathcal{E}_{\rho_1} A_1) \otimes (\mathcal{E}_{\rho_2} A_2)$$

$$\text{if } \rho_1, A_1 \text{ commute or } \rho_2, A_2 \text{ commute}, \quad (2.5)$$

$$\|A_1 \otimes I_2\|_\rho \leq \|A_1\|_{\operatorname{tr}_2 \rho}, \quad (2.6)$$

where A is any operator on $\mathcal{H}$, U is any unitary operator on $\mathcal{H}$, ρ_j is any density operator on $\mathcal{H}_j$, $\mathcal{H}_j$ is any Hilbert space, A_j is any operator on $\mathcal{H}_j$, I_j is the

---

[1]Equation (2.5) without the commutation condition is proposed in Ref. [15] and repeated in Ref. [18], but it turns out to be false for many operator products, including the Jordan product, as pointed out by Ref. [33]. The results in Ref. [18] remain correct if the commutation condition in Eq. (2.5) is imposed.

identity operator on $\mathcal{H}_j$, ρ in Eq. (2.6) is any density operator on $\mathcal{H}_1 \otimes \mathcal{H}_2$, and tr_j denotes the partial trace with respect to $\mathcal{H}_j$. Examples of $\mathcal{E}$ that satisfy Eqs. (2.3)–(2.6) include

$$\mathcal{E}_\rho A = \frac{1}{2}\left(\rho A + A\rho\right), \qquad (2.7)$$

$$\mathcal{E}_\rho A = \rho A, \qquad (2.8)$$

$$\mathcal{E}_\rho A = \sqrt{\rho} A \sqrt{\rho}. \qquad (2.9)$$

In the following, I fix $\mathcal{E}$ to be a map that satisfies Eqs. (2.3)–(2.6).

Let $L_2(\rho)$ be the completion of $\mathcal{O}(\mathcal{H})$ with respect to the norm $\|\cdot\|_\rho$, such that it becomes a weighted Hilbert space for the operators. Each element of $L_2(\rho)$ is then an equivalence class of operators with zero distance between them. If $\mathcal{H}$ is infinite-dimensional, $\mathcal{O}(\mathcal{H})$ may not be complete and $L_2(\rho)$ may include unbounded operators as well [46]. The infinite-dimensional case is much more complicated to treat with rigor, so I consider only finite-dimensional Hilbert spaces in the following for simplicity, and assume that the results still hold for a couple of the infinite-dimensional problems studied later in Appendix C and Sec. 5.5.

Definition 1. *Let σ be a density operator on $\mathcal{H}_1$ and $\mathcal{F} : \mathcal{O}(\mathcal{H}_1) \to \mathcal{O}(\mathcal{H}_2)$ be a completely positive, trace preserving (CPTP) map that models a quantum channel. Then the divergence between an operator $A \in L_2(\sigma)$ and another operator $B \in L_2(\mathcal{F}\sigma)$ is defined as*

$$D_{\sigma,\mathcal{F}}(A,B) \equiv \|A\|_\sigma^2 - 2\,\mathrm{Re}\,\langle \mathcal{F}^* B, A \rangle_\sigma + \|B\|_{\mathcal{F}\sigma}^2, \qquad (2.10)$$

where Re *denotes the real part and $\mathcal{F}^*$ denotes the Hilbert-Schmidt adjoint of $\mathcal{F}$.*

This divergence, introduced in Ref. [18], can be related to the more usual definition of distance in a larger Hilbert space by considering the Stinespring representation

$$\mathcal{F}\sigma = \mathrm{tr}_{10} U(\sigma \otimes \tau)U^\dagger, \qquad (2.11)$$

where τ is a density operator on $\mathcal{H}_2 \otimes \mathcal{H}_0$, $\mathcal{H}_0$ is some auxiliary Hilbert space, U is a unitary operator on $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_0$ that models the evolution from time t to time $T \geq t$, and tr_{10} is the partial trace over $\mathcal{H}_1$ and $\mathcal{H}_0$. Let $\rho = \sigma \otimes \tau$ and define the Heisenberg pictures of A and B as

$$A_t \equiv A \otimes I_2 \otimes I_0, \quad B_T \equiv U^\dagger (I_1 \otimes B \otimes I_0)U. \quad (2.12)$$

Then it can be shown that

$$D_{\sigma,\mathcal{F}}(A,B) \geq \|A_t - B_T\|_\rho^2, \qquad (2.13)$$

and the divergence is nonnegative. Furthermore, if the $\mathcal{E}$ map obeys the stricter equality condition in Eq. (2.6), then the equality in Eq. (2.13) holds, and D is exactly the squared distance in the larger Hilbert space.

Definition 2. *Given a density operator σ, a CPTP map $\mathcal{F}$, and an $\mathcal{E}$ map that satisfies Eqs. (2.3)–(2.6), the GCE $\mathcal{F}_* : L_2(\sigma) \to L_2(\mathcal{F}\sigma)$ of $A \in L_2(\sigma)$ is defined as*

$$\mathcal{F}_*A \equiv \underset{B \in L_2(\mathcal{F}\sigma)}{\arg\min}\, D_{\sigma,\mathcal{F}}(A,B), \qquad (2.14)$$

which leads to

$$\langle c, \mathcal{F}_*A \rangle_{\mathcal{F}\sigma} = \langle \mathcal{F}^*c, A \rangle_\sigma \quad \forall c \in L_2(\mathcal{F}\sigma). \qquad (2.15)$$

More explicitly, $\mathcal{F}_*A$ is an equivalence class of operators that satisfy

$$\mathcal{E}_{\mathcal{F}\sigma}\mathcal{F}_*A = \mathcal{F}\mathcal{E}_\sigma A. \qquad (2.16)$$

Equation (2.15) can be derived by assuming the ansatz $B = \mathcal{F}_*A + \epsilon c$ with $\epsilon \in \mathbb{R}$, $c \in L_2(\mathcal{F}\sigma)$, and minimizing D with respect to ϵ. Given an A, the existence and uniqueness of $\mathcal{F}_*A$ as an element of $L_2(\mathcal{F}\sigma)$ can be proved by viewing Eq. (2.15) as a linear functional of c and applying the Riesz representation theorem [47]. Equation (2.16) can also be derived independently from a state-over-time formalism [33]. With the GCE, the minimum divergence becomes

$$D_{\sigma,\mathcal{F}}(A,\mathcal{F}_*A) = \min_{B \in L_2(\mathcal{F}\sigma)} D_{\sigma,\mathcal{F}}(A,B)$$
$$= \|A\|_\sigma^2 - \|\mathcal{F}_*A\|_{\mathcal{F}\sigma}^2. \qquad (2.17)$$

Note that the GCE map $\mathcal{F}_*$ depends implicitly on the $\mathcal{E}$ map and the prior state σ; the choice of $\mathcal{E}$ and σ should be clear from the context in the following and, when necessary, σ is stated explicitly in the superscript as $\mathcal{F}_*^\sigma$. Note also that Chap. 6 in Ref. [15] writes $\mathcal{F}_*^\sigma$ as $\mathcal{F}_{\sigma,x}$, where x denotes the $\mathcal{E}$ map being used, while Ref. [18] writes $\mathcal{F}_*^\sigma$ as $\mathcal{F}_\sigma$. Appendix A presents more interesting formulas concerning $\mathcal{F}^*$ and $\mathcal{F}_*$ that justify the new notations, while Appendix B presents a brief and elementary review of the classical conditional expectation to give the quantum formalism a more familiar context.

Some examples are in order. Consider the unitary channel

$$\mathcal{F}\sigma = U\sigma U^\dagger, \qquad (2.18)$$

where U is a unitary operator on $\mathcal{H}_1$. A solution to any GCE is

$$\mathcal{F}_*A = UAU^\dagger, \qquad (2.19)$$

leading to $D_{\sigma,\mathcal{F}}(A,\mathcal{F}_*A) = 0$. Equation (2.19) is called the Heisenberg representation in quantum computing [48], and the GCEs can be regarded as generalizations of the Heisenberg representation for open systems.

With the root product given by Eq. (2.9), the GCE becomes the Accardi-Cecchini GCE [11, 12], and its Hilbert-Schmidt adjoint $(\mathcal{F}_*)^*$ is known as the Petz

recovery map, which is useful in quantum information theory [49].

Appendix C presents another example where σ is a Gaussian state, $\mathcal{F}$ is a Gaussian channel [50], and A is a quadrature operator. Then the GCE in terms of the Jordan product given by Eq. (2.7) and the associated divergence turn out to have the same formulas as the classical conditional expectation and its mean-square error for the usual linear Gaussian model [51].

3 Fundamental properties

With Eqs. (2.15)–(2.17), it is straightforward to prove the following crucial properties of the GCE:

Theorem 1 (Chain rule[2]; see Eq. (6.22) in Ref. [15]). *Let $\mathcal{G} : \mathcal{O}(\mathcal{H}_2) \to \mathcal{O}(\mathcal{H}_3)$ be another CPTP map. Then the GCE of the composite map $\mathcal{G}\mathcal{F}$ is given by*

$$(\mathcal{G}\mathcal{F})_*^\sigma A = \mathcal{G}_*^{\mathcal{F}\sigma} \mathcal{F}_*^\sigma A, \qquad (3.1)$$

which can be abbreviated as

$$(\mathcal{G}\mathcal{F})_* = \mathcal{G}_* \mathcal{F}_*. \qquad (3.2)$$

In other words, the GCE for a chain of CPTP maps is given by a chain of the GCEs associated with the individual CPTP maps.

Theorem 2 (Pythagorean theorem). *Given the two CPTP maps $\mathcal{F}$ and $\mathcal{G}$, the minimum divergences obey*

$$D_{\sigma,\mathcal{G}\mathcal{F}}(A, (\mathcal{G}\mathcal{F})_* A) = D_{\sigma,\mathcal{F}}(A, \mathcal{F}_* A) \\ + D_{\mathcal{F}\sigma,\mathcal{G}}(\mathcal{F}_* A, \mathcal{G}_* \mathcal{F}_* A). \quad (3.3)$$

Proof. Use Eq. (2.17) and Theorem 1. □

Figure 1 offers some diagrams that illustrate the theorems.

Before moving on, I list two more properties of the GCEs—their physical significance for generalizing the Rao-Blackwell theorem [5] will be explained in Sec. 5.

Lemma 1 (Law of total expectation). *For any $A \in L_2(\sigma)$,*

$$\text{tr}\, \sigma A = \langle I_1, A \rangle_\sigma = \langle I_2, \mathcal{F}_* A \rangle_{\mathcal{F}\sigma} = \text{tr}(\mathcal{F}\sigma)(\mathcal{F}_* A). \quad (3.4)$$

Lemma 2. *Let a be any complex number. Then*

$$\|A - aI_1\|_\sigma^2 = \|\mathcal{F}_* A - aI_2\|_{\mathcal{F}\sigma}^2 + D_{\sigma,\mathcal{F}}(A, \mathcal{F}_* A). \quad (3.5)$$

See Appendix J for the proofs of Lemmas 1 and 2.

A map $\mathcal{F}_*$ that satisfies Lemma 1 is also called a coarse graining [12]. Whereas Petz's definition requires a coarse graining to be completely positive, the

[2]I follow Ref. [11] to call this property a chain rule. Note that Ref. [15] calls it associativity, while Refs. [33, 52] call it compositionality.
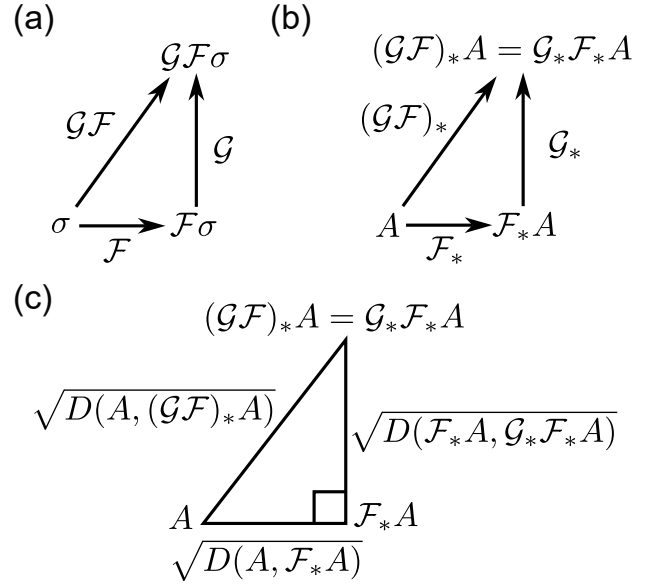

Figure 1: (a) A diagram depicting the map of a density operator σ through the CPTP maps $\mathcal{F}$ and then $\mathcal{G}$. (b) A diagram depicting the map of an observable A through the GCE $(\mathcal{G}\mathcal{F})_*$, or equivalently through the two GCEs $\mathcal{F}_*$ and then $\mathcal{G}_*$, as per Theorem 1. (c) A diagram depicting the root divergences between the operators as lengths of the sides of a right triangle, as per Theorem 2. The subscripts of D are omitted for brevity.

GCEs here need not be. If a in Lemma 2 is set as the mean given by Eq. (3.4), then Lemma 2 says that the generalized variance of $\mathcal{F}_* A$ given by $\|\mathcal{F}_* A - aI_2\|_{\mathcal{F}\sigma}^2$ cannot exceed that of A.

The mathematics of GCEs would be uncontroversial if not for its physical implication: By defining a divergence between two operators at different times, a retrodiction of a hidden quantum observable A can be given a risk measure and therefore a meaning in the spirit of decision theory [2]. In other words, after a channel $\mathcal{F}$ is applied, one can seek an observable B that is the closest to A if the divergence is regarded as a squared distance, and $\mathcal{F}_* A$ is the answer. It remains an open and reasonable question, however, why the divergence between two operators is an important quantity. If A_t at time t does not commute with B_T at a later time in the Heisenberg picture, where A_t and B_T are defined by Eqs. (2.12), then Belavkin's nondemolition principle for their simultaneous measurability is violated [20, 21, 39], no classical observer can access the precise values of both, and the divergence does not seem to have any obvious meaning to the classical world. Notably, Gough claims in Ref. [39] that a retrodiction that violates the nondemolition principle is "misapplying Bayes theorem," "not possible," and "unwarranted." Reference [38], the preprint version of Ref. [39], goes even further in claiming that someone who does not follow the principle may obtain "wholly meaningless" answers and is "in a state of sin." In Ref. [40], James also claims that the principle should

be observed for a quantum conditional expectation to make sense. To show that a retrodiction can make sense beyond the nondemolition principle, the next sections offer natural scenarios in quantum metrology that will give operational meanings to a GCE and the associated divergence.

4 Bayesian quantum parameter estimation

4.1 General results

Consider the typical setup of Bayesian quantum parameter estimation [9] depicted in Fig. 2(a). Let X be a hidden classical random parameter with a countable parameter space $\mathcal{X}$ and a prior probability distribution $P_X : \mathcal{X} \to [0,1]$. A quantum sensor is coupled to X, such that its density operator conditioned on $X = x$ is $\rho_x \in \mathcal{O}(\mathcal{H}_2)$. A classical observer measures the quantum sensor, as modeled by a positive operator-valued measure (POVM) $M : \Sigma_{\mathcal{Y}} \to \mathcal{O}(\mathcal{H}_2)$ on a Borel space $(\mathcal{Y}, \Sigma_{\mathcal{Y}})$, where $\Sigma_{\mathcal{Y}}$ is the Borel sigma-algebra of $\mathcal{Y}$ [46]. The observer uses the outcome $y \in \mathcal{Y}$ to estimate the value of a real random variable $a : \mathcal{X} \to \mathbb{R}$. The problem can be framed in the GCE formalism by writing

$$\sigma = \sum_x P_X(x) \,|x\rangle\,\langle x|, \quad A = \sum_x a(x) \,|x\rangle\,\langle x|, \tag{4.1}$$

$$\mathcal{F}\sigma = \sum_x \rho_x \,\langle x|\, \sigma \,|x\rangle = \sum_x \rho_x P_X(x), \tag{4.2}$$

where $\{|x\rangle : x \in \mathcal{X}\}$ is an orthonormal basis of $\mathcal{H}_1$ and the classical random variable $a(X)$ is framed as the hidden observable A discussed in Secs. 2 and 3. $\mathcal{F}$ here is called a classical-quantum channel and has a natural generalization in the infinite-dimensional case [50].
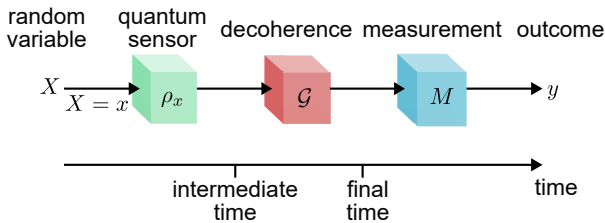

(a)
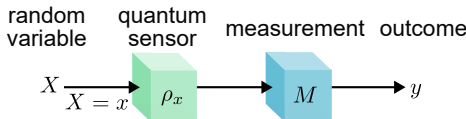
(b)

Figure 2: Some scenarios of Bayesian quantum parameter estimation. See the main text for the definitions of the symbols.

In the following, I consider only Hermitian operators (observables) and assume $\mathcal{E}$ to be the Jordan product given by Eq. (2.7), such that all the operator Hilbert spaces are real, the equalities in Eqs. (2.6) and (2.13) hold, and the GCE is in fact a projection in the larger Hilbert space [15].

Suppose that a von Neumann measurement of an observable B on $\mathcal{H}_2$ is performed, as defined in Appendix D, and the outcome is used as the estimator. I call such a B an operator-valued estimator. The mean-square estimation error averaged over the prior is given by

$$\sum_x P_X(x) \int [b - a(x)]^2 \,\mathrm{tr}\,\Pi(db)\rho_x = D_{\sigma,\mathcal{F}}(A, B), \tag{4.3}$$

where Π is the projection-valued measure of B. Equation (4.3) is precisely the divergence in Definition 1. According to the seminal work of Personick [9], the optimal operator-valued estimator is the GCE $\mathcal{F}_* A$, and the minimum error, hereafter called the Bayesian error, is $D_{\sigma,\mathcal{F}}(A, \mathcal{F}_* A)$. It can also be shown that the von Neumann measurement of $\mathcal{F}_* A$ remains optimal even if POVMs are considered (see Sec. VIII 1(d) in Ref. [53], Appendix A in Ref. [43], or Corollary 2 below).

Now suppose that a complication occurs in the experiment, as depicted by Fig. 2(b): Before the measurement can be performed, the sensor is further corrupted by decoherence, as modeled by another CPTP map $\mathcal{G}$. The error of an operator-valued estimator B' is now

$$\sum_x P_X(x) \int [b - a(x)]^2 \,\mathrm{tr}\,\Pi'(db)\mathcal{G}\rho_x = D_{\sigma,\mathcal{G}\mathcal{F}}(A, B'), \tag{4.4}$$

where Π' is the projection-valued measure of B'. The Personick estimator after $\mathcal{G}$ is then $B' = (\mathcal{G}\mathcal{F})_* A$, and the Bayesian error becomes $D_{\sigma,\mathcal{G}\mathcal{F}}(A, (\mathcal{G}\mathcal{F})_* A)$. A fundamental fact is as follows.

Corollary 1 (Monotonicity of the Bayesian error)**.** *The Bayesian error cannot decrease under decoherence, viz.,*

$$D_{\sigma,\mathcal{G}\mathcal{F}}(A, (\mathcal{G}\mathcal{F})_* A) \geq D_{\sigma,\mathcal{F}}(A, \mathcal{F}_* A) \tag{4.5}$$

for the estimation problem modeled by Eqs. (4.1)–(4.4).

Proof. Use Theorem 2 and the nonnegativity of D. □

The scenario so far is standard and uncontroversial, as A is effectively a classical random variable. Mathematically, A_t and $(\mathcal{F}_* A)_T$ in the Heisenberg picture commute (see Sec. IV F in Ref. [18]) and thus satisfy the nondemolition principle; so do A_t and $[(\mathcal{G}\mathcal{F})_* A]_T$. Physically, the principle implies that another classical

observer can, in theory, access the precise value of A in each trial, the estimates can be compared with the true values by the classical observers after the trials, and D is their expected error. The monotonicity given by Corollary 1 is a noteworthy result, but unsurprising.

More can be said about the error increase, hereafter called the regret (to borrow a term from decision theory [2]). First of all, the chain rule in Theorem 1 gives an operational meaning to the GCE $\mathcal{G}_*$ as the map that relates the intermediate Personick estimator $\mathcal{F}_*A$ to the final $(\mathcal{GF})_*A = \mathcal{G}_*\mathcal{F}_*A$. In other words, the final Personick estimator is equivalent to a retrodiction of the intermediate $\mathcal{F}_*A$, which is a quantum observable. Second, the Pythagorean theorem in Theorem 2 means that the regret caused by the decoherence is precisely the divergence between the intermediate and final estimators:

$$D_{\sigma,\mathcal{GF}}(A,(\mathcal{GF})_*A) - D_{\sigma,\mathcal{F}}(A,\mathcal{F}_*A)$$
$$= D_{\mathcal{F}\sigma,\mathcal{G}}(\mathcal{F}_*A,\mathcal{G}_*\mathcal{F}_*A). \quad (4.6)$$

The two divergences on the left-hand side have a firm decision-theoretic meaning as estimation errors because A is classical. It follows that, even though the divergence on the right-hand side is between two quantum observables, it also has a firm decision-theoretic meaning as the regret—for being unable to perform the optimal measurement and having to suffer from the decoherence. As the regret concerns the performances of the two estimators in separate experiments, the estimators need not obey the nondemolition principle, which is a condition on two observables to be simultaneously measurable in the same experiment.

I stress that the regret is not a contrived concept invented here solely to give an operational meaning to the divergence—its classical version is an established concept in information theory and Bayesian learning [54, 55, 56, 57].

4.2 Dynamic programming

When the decoherence is modeled by a chain of CPTP maps $\mathcal{G} = \mathcal{F}_N \ldots \mathcal{F}_2$, the final error is the sum of all the incremental regrets along the way, viz.,

$$D_{\sigma,\mathcal{G}_N}(A,\mathcal{G}_{N*}A) = \sum_{n=1}^{N} D_n, \quad (4.7)$$

$$\mathcal{G}_n \equiv \mathcal{F}_n \ldots \mathcal{F}_2\mathcal{F}_1, \quad (4.8)$$

$$D_n \equiv D_{\sigma_n,\mathcal{F}_n}(A_n,A_{n+1}), \quad (4.9)$$

$$\sigma_{n+1} \equiv \mathcal{G}_n\sigma = \mathcal{F}_n\sigma_n, \quad \sigma_1 = \sigma, \quad (4.10)$$

$$A_{n+1} \equiv \mathcal{G}_{n*}A = \mathcal{F}_{n*}A_n, \quad A_1 = A, \quad (4.11)$$

where $\mathcal{F}_1 = \mathcal{F}$ for the parameter estimation problem, so even the error at the first step $D_1 = D_{\sigma,\mathcal{F}}(A,\mathcal{F}_*A)$ can be regarded as a regret. Every D_n, bar D_1, is a divergence between a quantum observable A_n and its estimator A_{n+1} that may not commute in the Heisenberg picture.

Suppose that the experimenter can choose the channels $(\mathcal{F}_1,\ldots,\mathcal{F}_N)$ from a set of options and would like to find the optimal choice that minimizes the final error. One example is the use of a programmable photonic circuit [58] to measure light for sensing or imaging. The Markovian nature of Eqs. (4.10)–(4.11) and the additive nature of the final error given by Eq. (4.7)—which originate from Theorems 1 and 2—are precisely the conditions that make this optimal control problem amenable to dynamic programming [41], an algorithm that can reduce the computational complexity substantially [59]. To be specific, let the system state (in the context of control theory) at time n be $s_n \equiv (\sigma_n, A_n)$. Then Eqs. (4.7)–(4.11) imply that the state dynamics and the final error can be expressed as

$$s_{n+1} = f(s_n,\mathcal{F}_n), \quad (4.12)$$

$$D_{\sigma,\mathcal{G}_N} = \sum_{n=1}^{N} g(s_n,\mathcal{F}_n) \quad (4.13)$$

in terms of some functions f and g. Equations (4.12) and (4.13) are now in the form of a Markov decision process that is amenable to dynamic programming for computing the optimal maps $(\mathcal{F}_1,\ldots,\mathcal{F}_N)$ among the set of options to minimize the final error [41]; Appendix E describes the algorithm for the reader's information. As dynamic programming is a cornerstone of control theory, there exist a plethora of exact or approximate methods to implement it, such as neural networks under the guise of reinforcement learning [60].

4.3 Weak value

To elaborate on the operational meaning for the weak value, which is a GCE of a quantum observable given a prior state and a measurement outcome [16, 17, 14, 18], let us return to the scenario depicted by Fig. 2(a). Suppose, for mathematical simplicity, that the outcome space $\mathcal{Y}$ is countable. The measurement can be framed as a $\mathcal{G}$ map given by

$$\mathcal{G}\tau = \sum_y [\operatorname{tr} M(y)\tau] |y\rangle \langle y|, \quad (4.14)$$

where $\{|y\rangle : y \in \mathcal{Y}\}$ is an orthonormal basis of $\mathcal{H}_3$ and M is the POVM of a measurement that may not be optimal. An estimator $b : \mathcal{Y} \to \mathbb{R}$ as a function of the measurement outcome can be framed as the observable

$$B = \sum_y b(y) |y\rangle \langle y|. \quad (4.15)$$

The GCE then leads to the optimal estimator

$$B = \mathcal{G}_* \mathcal{F}_* A, \tag{4.16}$$

$$b(y) = \frac{\operatorname{tr} M(y) \mathcal{E}_{\mathcal{F}\sigma} \mathcal{F}_* A}{\operatorname{tr} M(y) \mathcal{F}\sigma} = \frac{\operatorname{Re} \operatorname{tr} M(y)(\mathcal{F}_* A)(\mathcal{F}\sigma)}{\operatorname{tr} M(y) \mathcal{F}\sigma}. \tag{4.17}$$

The last expression in Eq. (4.17) is precisely the definition of the real weak value of $\mathcal{F}_* A$ given a prior state $\mathcal{F}\sigma$ and a measurement outcome y (see, for example, Eq. (3.13) in Ref. [61], Eq. (10) in Ref. [62], or Eq. (5) in Ref. [14]). Moreover, the divergence between the ideal $\mathcal{F}_* A$ and the B associated with the weak value is precisely the regret caused by the suboptimality of the measurement M, as per Theorem 2. Hence, regardless of how anomalous the weak value may seem, it does have an operational role in parameter estimation, and its divergence from the ideal $\mathcal{F}_* A$ has a concrete decision-theoretic meaning as the regret for not using the optimal measurement.

The preceding discussion also serves as a rough proof of the following corollary, which is proved by different methods in Sec. VIII 1(d) of Ref. [53] and Appendix A of Ref. [43].

Corollary 2. *No POVM can improve upon the Bayesian error $D_{\sigma,\mathcal{F}}(A, \mathcal{F}_* A)$ achieved by a von Neumann measurement of $\mathcal{F}_* A$.*

Corollary 2 may be regarded as a consequence of monotonicity, since any measurement with a countable set of outcomes can be framed as a CPTP $\mathcal{G}$ map given by Eq. (4.14), and by Corollary 1, the error cannot decrease. A POVM with a more general outcome space can still be framed as a quantum-classical channel; see, for example, Theorem 2 in Ref. [63], but it requires a mathematical framework far more complex than what is necessary for this work. An easier proof for general POVMs, to be presented in Appendix J, is to use a later result in Sec. 5.

Note that the optimality of the weak value here does not contradict Ref. [34], which shows that weak-value amplification, a procedure that involves postselection (i.e., discarding some of the outcomes), is suboptimal for metrology. Here, the weak value given by Eq. (4.17) is used directly as an estimator with any measurement outcome, and no postselection is involved. Note also that the optimality is in the specific context of finding the best estimator after a given measurement; it does not mean that any measurement method that is heuristically inspired by the weak-value concept, such as weak-value amplification, can be optimal. In fact, by virtue of Corollary 2, such methods can never outperform the optimal von Neumann measurement.

5 A quantum Rao-Blackwell theorem

5.1 General result

In classical frequentist statistics, the Rao-Blackwell theorem is among the most useful applications of the conditional expectation [5, 6, 7]. Here I outline a quantum generalization. Suppose that the quantum sensor is modeled by a family of density operators $\{\rho_x : x \in \mathcal{X}\} \subset \mathcal{O}(\mathcal{H}_2)$, where the unknown parameter x is now deterministic and there is no longer any need to assume a countable parameter space $\mathcal{X}$. A parameter of interest $a : \mathcal{X} \to \mathbb{R}$ is to be estimated by a Hermitian operator-valued estimator $B \in L_2(\rho_x)$, which need not be unbiased or optimal in any sense. The local mean-square error (MSE) upon a von Neumann measurement of B, as a function of $x \in \mathcal{X}$ and without being averaged over any prior, is given by

$$\mathrm{MSE}_x \equiv \int [b - a(x)]^2 \operatorname{tr} \Pi(db) \rho_x = \|B - a(x) I_2\|_{\rho_x}^2, \tag{5.1}$$

where Π is the projection-valued measure of B and the Jordan product is again assumed for $\mathcal{E}$. Without the unbiased condition $\operatorname{tr} \rho_x B = a(x)$, the quantum Cramér-Rao bounds commonly used in quantum metrology [53, 15] do not apply to MSE_x. There is no longer any simple optimality criterion for an estimator in this general setting, but one can still construct a *partial order* of preference between estimators. Following classical statistics (see p. 48 in Ref. [5]), I say that an operator-valued estimator B' dominates another estimator B if the error MSE_x' of the former never exceeds the error MSE_x of the latter for all $x \in \mathcal{X}$ and can go strictly lower for some x. I call an estimator admissible if it is dominated by none.

In classical frequentist statistics, there can be many admissible estimators for one problem with no clear winner among them, and it may be hard to even prove that a given estimator is admissible. The Rao-Blackwell theorem is then a valuable tool for improving estimators or for proofs regarding admissibility. A quantum version of the theorem can be similarly useful for quantum estimation problems.

To state the quantum theorem, suppose that the quantum sensor goes through a channel modeled by a CPTP map $\mathcal{G} : \mathcal{O}(\mathcal{H}_2) \to \mathcal{O}(\mathcal{H}_3)$ and the GCE $\mathcal{G}_* B$ in terms of ρ_x and the Jordan product is used as an estimator. The error becomes

$$\mathrm{MSE}_x' = \|\mathcal{G}_* B - a(x) I_3\|_{\mathcal{G}\rho_x}^2. \tag{5.2}$$

Lemma 2 can now be used to prove the following.

Theorem 3 (Quantum Rao-Blackwell theorem). *Let $\{\rho_x : x \in \mathcal{X}\}$ be a family of density operators, $a : \mathcal{X} \to \mathbb{R}$ be an unknown parameter, B be a Hermitian operator-valued estimator, and MSE_x be the local error at $x \in \mathcal{X}$. If a channel $\mathcal{G}$ is applied and*

the GCE $\mathcal{G}_*B$ in terms of ρ_x and the Jordan product does not depend on x, then the error MSE'_x of $\mathcal{G}_*B$ as an estimator is lower by the amount

$$\mathrm{MSE}_x - \mathrm{MSE}'_x = D_{\rho_x,\mathcal{G}}(B, \mathcal{G}_*B). \qquad (5.3)$$

Proof. Subtract Eq. (5.2) from Eq. (5.1) and apply Lemma 2. $\qquad\square$

For $\mathcal{G}$ to be realizable and $\mathcal{G}_*B$ to be a valid estimator, both cannot depend on the unknown x. When there are many operator solutions to $\mathcal{G}_*B$ that satisfy Definition 2, any of the solutions can be the estimator in Theorem 3 as long as it does not depend on x.

To demonstrate that Theorem 3 is indeed a quantum generalization of the Rao-Blackwell theorem, Appendix F derives the classical theorem from Theorem 3. As also shown in Appendix F, a parameter-independent conditional expectation in classical statistics can be obtained by conditioning on a sufficient statistic. The conditional expectation can then be used to improve an estimator in a process called Rao-Blackwellization [5]. Roughly speaking, Rao-Blackwellization works by averaging the estimator with respect to unnecessary parts of the data, thereby reducing its variance. A quantum Rao-Blackwellization, enabled by Theorem 3, can be similarly useful for improving a quantum measurement if one can find a channel $\mathcal{G}$ that satisfies the constant GCE condition and gives a large divergence between B and $\mathcal{G}_*B$. As long as $D_{\rho_x,\mathcal{G}}(B, \mathcal{G}_*B) > 0$ for some x, the Rao-Blackwell estimator $\mathcal{G}_*B$ dominates the original estimator B. The improvement stems from two basic facts about the GCE: $\mathcal{G}_*B$ maintains the same bias as that of B by virtue of Lemma 1, while the variance of $\mathcal{G}_*B$ cannot exceed that of B by virtue of Lemma 2. Roughly speaking, the quantum Rao-Blackwell theorem works in the same way as the classical case by averaging the estimator with respect to unnecessary degrees of freedom via the GCE, thereby reducing its variance.

For the confused readers who wonder how a channel increases the error in the Bayesian setting because of monotonicity but reduces the error in the frequentist setting because of the Rao-Blackwell theorem, Appendix G offers a clarification.

It is noteworthy that Sinha also proposed some quantum Rao-Blackwell theorems recently [64], although his versions impose stringent conditions on the commutativity of the operators. Another relevant prior work is Ref. [65] by Łuczak, which studies a concept of sufficiency in von Neumann algebra for minimum-variance unbiased estimation but also makes some stringent assumptions. These prior works, while seminal and mathematically impressive, have questionable relevance to quantum metrology and are discussed in more detail in Appendix H.

Given the close relation between the Rao-Blackwell theorem and the concept of sufficient statistics in the

classical case, it is natural to wonder if a similar relation exists between the quantum Rao-Blackwell theorem here and the concept of sufficient channels defined by Petz [12]. One equivalent condition for a channel $\mathcal{G}$ to be sufficient in Petz's definition is that the Accardi-Cecchini GCE $\mathcal{G}_*$ in terms of the root product given by Eq. (2.9) does not depend on x. The GCE here, on the other hand, is in terms of the Jordan product so that it can be related to the parameter estimation error. The relation between Petz's sufficiency and the constant GCE condition desired here is thus nontrivial.

A trivial example that makes any GCE constant and the channel sufficient in any sense is the unitary channel given by Eqs. (2.18) and (2.19), as long as the unitary operator there does not depend on x. Applying Theorem 3 to the unitary channel gives no error reduction, however. In the following, I offer more useful examples that both satisfy Petz's sufficiency and give the desired constant GCE condition.

5.2 A sufficient channel for tensor-product states

Lemma 3. *Let*

$$\rho_x = \sigma_x \otimes \tau, \qquad \mathcal{G}\rho_x = \mathrm{tr}_0 \rho_x = \sigma_x, \qquad (5.4)$$

where σ_x is a density operator on $\mathcal{H}_1$ and τ is an auxiliary density operator on $\mathcal{H}_0$. A solution to any GCE is

$$\mathcal{G}_*B = \mathrm{tr}_0[(I_1 \otimes \tau)B], \qquad (5.5)$$

which does not depend on x if τ does not.

See Appendix J for the proof.

A sufficient channel may be understood intuitively as a channel that retains all information in the quantum sensor about x. Then it makes sense that the channel in Lemma 3 is sufficient, as it simply amounts to discarding an independent ancilla that carries no information about x. A significant implication of the lemma is a more general version of Corollary 2 for the local error as follows.

Corollary 3. *Given any POVM $M : \Sigma_\mathcal{Y} \to \mathcal{O}(\mathcal{H}_2)$, any estimator $b : \mathcal{Y} \to \mathbb{R}$, and the resulting local error MSE_x, there exists a von Neumann measurement that can perform at least as well for all $x \in \mathcal{X}$.*

Proof. Write the Naimark extension of the POVM as

$$\mathrm{tr}\, M(\cdot)\sigma_x = \mathrm{tr}\, \Pi'(\cdot)(\sigma_x \otimes \tau), \qquad (5.6)$$

where σ_x and τ are defined in Lemma 3 and Π' is a projection-valued measure on $\mathcal{H}_1 \otimes \mathcal{H}_0$. As shown in Appendix D, the measurement and the data processing by b can be framed as a von Neumann measurement of $B = \int b(y)\Pi'(dy)$ on the larger Hilbert space, such that its error MSE_x with respect to $\rho_x = \sigma_x \otimes \tau$

is given by Eq. (5.1). Now assume the channel in Lemma 3. A solution to $\mathcal{G}_*B$ is given by Eq. (5.5), which does not depend on x. It follows from Theorem 3 that the error MSE'_x achieved by a von Neumann measurement of $\mathcal{G}_*B$ is at least as good as MSE_x for all $x \in \mathcal{X}$. $\square$

Note that Corollary 3 is more general than Corollary 2, since the former applies to the local errors for all parameter values, not just the average errors in the Bayesian case. A proof of Corollary 2 using Corollary 3 is presented in Appendix J.

The corollaries imply that, in seeking an admissible estimator for estimating a real scalar parameter under a mean-square-error criterion, it is sufficient to consider only von Neumann measurements, and randomization via an independent ancilla is not helpful in both Bayesian and frequentist settings. For example, optical amplification has been suggested to improve astronomical measurements [66], but since optical amplification must involve an independent ancilla [67, 68], Corollary 3 implies that there always exists a von Neumann measurement that performs at least as well.

The corollaries are reminiscent of a well known result saying that a von Neumann measurement of the so-called symmetric logarithmic derivative (SLD) operator can saturate the quantum Cramér-Rao bound (see Sec. 6.4 in Ref. [15]). Note, however, that the bound assumes unbiased estimators and the differentiability of ρ_x, while the SLD measurement may be a function of the unknown parameter and thus unrealizable. The corollaries here, on the other hand, are much more general and conclusive, as they apply to arbitrary estimators and arbitrary families of density operators, while the von Neumann measurements they offer are all parameter-independent.

Of course, one is often forced to use an ancilla in practice, such as the optical probe in atomic metrology [69] or optomechanics [70]. Then the divergence offers a measure of regret in both Bayesian and frequentist settings through Theorems 2 and 3. Consider atomic metrology for an example [69]. Let ρ_x be the parameter-dependent density operator of the atoms on $\mathcal{H}_1$ and τ be the initial state of an optical probe on $\mathcal{H}_0$. Then the state after the optical probing can be expressed as $U(\rho_x \otimes \tau)U^\dagger$, where U is a unitary operator on $\mathcal{H}_1 \otimes \mathcal{H}_0$ that models the atom-light interaction. If an optical measurement, modeled by a projection-valued measure Π_0 on $\mathcal{H}_0$, is performed and the estimator in terms of the outcome y is $b(y)$, then the B observable in Lemma 3 and Corollary 3 can be expressed as

$$B = \int b(y)\Pi'(dy), \quad \Pi'(\cdot) = U^\dagger \left[I_1 \otimes \Pi_0(\cdot)\right] U, \quad (5.7)$$

and MSE_x is the error of this indirect measurement of the atoms. The GCE $\mathcal{G}_*B$, on the other hand,

is an atomic observable on $\mathcal{H}_1$, and MSE'_x is the error of the direct atomic measurement of $\mathcal{G}_*B$. Then $D_{\rho_x,\mathcal{G}}(B, \mathcal{G}_*B) = \mathrm{MSE}_x - \mathrm{MSE}'_x$ can be regarded as the regret due to the indirectness of the optical measurement. The Bayesian setting can be studied similarly.

5.3 A sufficient channel for symmetric states

Let $\{U_z : z \in \mathcal{Z}\}$ be a set of unitary operators on $\mathcal{H}_2$, and suppose that ρ_x is invariant to all of them, viz.,

$$U_z \rho_x U_z^\dagger = \rho_x \quad \forall z \in \mathcal{Z}. \quad (5.8)$$

Examples include the symmetric states that are invariant to any permutation of a tensor-powered Hilbert space—to be discussed later—and optical states with random phases that are invariant to any phase modulation. ρ_x is also invariant to the random unitary channel

$$\mathcal{G}\rho_x = \int \mu(dz) U_z \rho_x U_z^\dagger = \rho_x \quad (5.9)$$

for any probability measure μ on $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$. $\mathcal{G}$ is then a sufficient channel in Petz's sense, since another equivalent condition for Petz's sufficiency is the existence of an x-independent CPTP map that recovers ρ_x from $\mathcal{G}\rho_x$ [12]. It is straightforward to compute the GCEs.

Lemma 4. *Given Eqs. (5.8) and (5.9), a solution to any GCE is*

$$\mathcal{G}_*B = \int \mu(dz) U_z B U_z^\dagger, \quad (5.10)$$

which does not depend on x if $\{U_z\}$ and μ do not.

See Appendix J for the proof.

Corollary 4. *Given a family of states that are invariant to a set of unitaries $\{U_z\}$, any estimator $B \in L_2(\rho_x)$, and the resulting local error MSE_x, there exists an averaged estimator given by Eq. (5.10) that performs at least as well as B for all $x \in \mathcal{X}$.*

Proof. Use Lemma 4 and Theorem 3. $\square$

If $\mathcal{Z}$ is a group and $\{U_z\}$ is a projective unitary representation of the group that satisfies $U_{z'}U_z = \omega(z', z)U_{z'z}$ for a complex scalar ω with $|\omega| = 1$ [46], then the left Haar measure $\tilde{\mu}$ on the group [1] plays a special role, as the GCE with respect to it, written as

$$\tilde{\mathcal{G}}_*B \equiv \int \tilde{\mu}(dz) U_z B U_z^\dagger, \quad (5.11)$$

is invariant to any subsequent GCE for any random unitary channel, in the sense that

$$\mathcal{G}_*\tilde{\mathcal{G}}_*B = \tilde{\mathcal{G}}_*B \quad (5.12)$$

for any μ. The left Haar measure is thus the ultimate choice that gives the highest error reduction in the context of Corollary 4.

For a concrete example, let $\mathcal{H}_2 = \mathcal{H}_1^{\otimes n}$, $\pi \in S_n$ be a permutation function of $(1, \ldots, n)$, and S_n be the permutation group. Define each unitary by [71]

$$U_\pi(|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle) = |\psi_{\pi^{-1}1}\rangle \otimes \cdots \otimes |\psi_{\pi^{-1}n}\rangle \tag{5.13}$$

for any $\{|\psi_j\rangle \in \mathcal{H}_1 : j = 1, \ldots, n\}$. An operator invariant to all the permutation unitaries is called symmetric. Physically, a symmetric density operator corresponds to n indistinguishable systems. A common example is $\rho_x = \sigma_x^{\otimes n}$, where σ_x is a density operator on $\mathcal{H}_1$. The Haar measure is simply $\tilde{\mu}(\pi) = 1/n!$, and the corresponding GCE is

$$\tilde{\mathcal{G}}_* B = \frac{1}{n!} \sum_\pi U_\pi B U_\pi^\dagger, \tag{5.14}$$

which is a symmetrization. Furthermore, if one assumes

$$B = C \otimes I_1^{\otimes(n-m)}, \qquad C \in \mathcal{O}(\mathcal{H}_1^{\otimes m}), \tag{5.15}$$

then Eq. (5.14) leads to the quantum U-statistics introduced by Guţă and Butucea [45], as shown in Appendix I. The U-statistic is an unbiased estimator of $a(x) = \operatorname{tr} \rho_x B = \operatorname{tr} \rho_x \mathcal{G}_* B$. The simplest example is when $m = 1$ and

$$\tilde{\mathcal{G}}_* B = \frac{1}{n} \sum_{l=1}^n I_1^{\otimes(l-1)} \otimes C \otimes I_1^{\otimes(n-l)}, \tag{5.16}$$

which lowers the variance of B by a factor of n if $\rho_x = \sigma_x^{\otimes n}$.

The derivation of the classical U-statistics by Rao-Blackwellization is well known [72], and Corollary 4 is indeed the appropriate quantum generalization.

5.4 A sufficient channel for direct-sum states

Suppose now that $\{\rho_x : x \in \mathcal{X}\}$ is a family of density operators on a direct sum of Hilbert spaces given by

$$\mathcal{H} = \bigoplus_{n \in \mathcal{N}} \mathcal{H}_n, \tag{5.17}$$

and each ρ_x is given by the direct sum

$$\rho_x = \bigoplus_{n \in \mathcal{N}} \sigma_x^{(n)}, \tag{5.18}$$

where each $\sigma_x^{(n)}$ is a positive-semidefinite operator on $\mathcal{H}_n$. A prominent example in optics is the multimode thermal state, which will be discussed in Sec. 5.5. Let $\Pi_n : \mathcal{H} \to \mathcal{H}_n$ be the projection operator onto $\mathcal{H}_n$. Suppose that the Hilbert-space decomposition given

by Eq. (5.17) is parameter-independent, such that all $\{\Pi_n : n \in \mathcal{N}\}$ do not depend on x. Then the channel

$$\mathcal{G}\rho_x = \bigoplus_n \Pi_n \rho_x \Pi_n = \rho_x \tag{5.19}$$

is sufficient in Petz's sense. To compute the GCEs with respect to Eqs. (5.18) and (5.19), I impose two more properties on the $\mathcal{E}$ map given by

$$\mathcal{E}_{\sigma^{(1)} \oplus \sigma^{(2)}}(A_1 \oplus A_2) = (\mathcal{E}_{\sigma^{(1)}} A_1) \oplus (\mathcal{E}_{\sigma^{(2)}} A_2), \tag{5.20}$$

$$\Pi_1 \left(\mathcal{E}_{\sigma^{(1)} \oplus \sigma^{(2)}} A\right) \Pi_1 = \mathcal{E}_{\sigma^{(1)}}(\Pi_1 A \Pi_1) \tag{5.21}$$

for any $A_j \in \mathcal{O}(\mathcal{H}_j)$, any $A \in \mathcal{O}(\mathcal{H}_1 \oplus \mathcal{H}_2)$, and any density operator on $\mathcal{H}_1 \oplus \mathcal{H}_2$ in the form of $\sigma^{(1)} \oplus \sigma^{(2)}$. These properties are satisfied by the products given by Eqs. (2.7)–(2.9) at least. Then the GCE has the following solution.

Lemma 5. *Given Eqs. (5.18) and (5.19) and assuming a GCE in terms of an $\mathcal{E}$ map that satisfies Eqs. (5.20) and (5.21), a solution to the GCE is*

$$\mathcal{G}_* B = \bigoplus_n \Pi_n B \Pi_n, \tag{5.22}$$

which does not depend on x if the projectors $\{\Pi_n\}$ do not.

See Appendix J for the proof.

The quantum Rao-Blackwell theorem can now be applied to Eqs. (5.18) and (5.19) to prove the following.

Corollary 5. *Assume that the Hilbert space can be decomposed as Eq. (5.17) and the projectors $\{\Pi_n : \mathcal{H} \to \mathcal{H}_n\}$ do not depend on the unknown parameter x. Given a density-operator family in the form of a direct sum as per Eq. (5.18), any estimator $B \in L_2(\rho_x)$, and the resulting local error MSE_x, there exists an estimator $\mathcal{G}_* B$ given by Eq. (5.22), also in the form of a direct sum, that performs at least as well as B for all $x \in \mathcal{X}$.*

Proof. Use Lemma 5 and Theorem 3. $\square$

An example in optics is now in order.

5.5 Thermal-light sensing

A multimode thermal optical state can be expressed as [73]

$$\rho_x = \int d^{2J}\alpha \, \Phi_x(\alpha) |\alpha\rangle \langle\alpha|, \tag{5.23}$$

$$\alpha \equiv \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} \in \mathbb{C}^J, \quad \hat{a}_j |\alpha\rangle = \alpha_j |\alpha\rangle, \tag{5.24}$$

$$\Phi_x(\alpha) = \frac{1}{\det(\pi\Gamma_x)} \exp(-\alpha^\dagger \Gamma_x^{-1} \alpha), \tag{5.25}$$

where $|\alpha\rangle$ is a coherent state, $\hat{a}_j$ is the annihilation operator for the jth mode, $d^{2J}\alpha \equiv \prod_{j=1}^{J} d(\mathrm{Re}\,\alpha_j)d(\mathrm{Im}\,\alpha_j)$, and Γ_x is the positive-definite mutual coherence matrix. In thermal-light sensing and imaging problems [74, 75, 76, 77, 78, 79], Γ_x is assumed to depend on the unknown parameter x.

Let $\mathcal{H}_n$ be the n-photon Hilbert space. Define a pure Fock state with photon numbers $\boldsymbol{m} = (m_1, \ldots, m_J) \in \mathbb{N}_0^J$ as

$$|\boldsymbol{m}\rangle \equiv \left[\prod_j \frac{(\hat{a}_j^\dagger)^{m_j}}{\sqrt{m_j!}}\right]|\boldsymbol{0}\rangle, \qquad (5.26)$$

where $|\boldsymbol{0}\rangle$ denotes the vacuum state. Let $\|\boldsymbol{m}\| \equiv \sum_j m_j$ be the total photon number. Then $\{|\boldsymbol{m}\rangle : \|\boldsymbol{m}\| = n\}$ is an orthonormal basis of $\mathcal{H}_n$. In terms of the Fock basis, each matrix element of ρ_x is given by

$$\langle\boldsymbol{m}|\rho_x|\boldsymbol{l}\rangle = \int d^{2J}\alpha\,\Phi_x(\alpha) \prod_j e^{-|\alpha_j|^2} \frac{\alpha_j^{m_j}(\alpha_j^*)^{l_j}}{\sqrt{m_j!l_j!}}. \qquad (5.27)$$

The Gaussian moment theorem (see Eq. (1.6-33) in Ref. [73]) implies that

$$\langle\boldsymbol{m}|\rho_x|\boldsymbol{l}\rangle = 0 \text{ if } \|\boldsymbol{m}\| \neq \|\boldsymbol{l}\|, \qquad (5.28)$$

meaning that ρ_x can be decomposed in the direct-sum form as

$$\rho = \bigoplus_{n=0}^{\infty} \sigma_x^{(n)}, \qquad (5.29)$$

$$\sigma_x^{(n)} = \sum_{\boldsymbol{m},\boldsymbol{l}:\|\boldsymbol{m}\|=\|\boldsymbol{l}\|=n} \langle\boldsymbol{m}|\rho_x|\boldsymbol{l}\rangle |\boldsymbol{m}\rangle\langle\boldsymbol{l}|, \qquad (5.30)$$

where each $\sigma_x^{(n)}$ is an operator on $\mathcal{H}_n$. Then $\mathrm{tr}\,\sigma_x^{(n)}$ is the probability of having n photons in total and $\sigma_x^{(n)}/\mathrm{tr}\,\sigma_x^{(n)}$ is the conditional n-photon state. The projectors can be written as

$$\Pi_n = \sum_{\boldsymbol{m}:\|\boldsymbol{m}\|=n} |\boldsymbol{m}\rangle\langle\boldsymbol{m}|. \qquad (5.31)$$

Ignoring the mathematical complications due to the infinite-dimensional Hilbert space, Corollary 5 can now be applied to Eq. (5.29).

If an estimator is constructed from a photon-counting measurement with respect to any set of optical modes, it can be expressed in a Fock basis, which commutes with all the projectors $\{\Pi_n\}$. It follows that the estimator is already in the direct-sum form given by Eq. (5.22) and Corollary 5 offers no improvement. On the other hand, notice that Eq. (5.22) must commute with each projector Π_n, viz.,

$$[\mathcal{G}_*B, \Pi_n] = 0 \quad \forall n \in \mathcal{N}. \qquad (5.32)$$

If an estimator does not commute with all $\{\Pi_n\}$, such as one obtained from homodyne detection, then the estimator does not have the direct-sum form and has the potential to be improved by the quantum Rao-Blackwellization.

To introduce a more specific example, diagonalize Γ_x in terms of a diagonal matrix D_x and a unitary matrix V_x as

$$\Gamma_x = V_x D_x V_x^\dagger, \qquad D_{jk,x} = \lambda_{j,x}\delta_{jk}, \qquad (5.33)$$

where δ_{jk} is the Kronecker delta and each $\lambda_{j,x}$ is an eigenvalue of Γ_x. I call $\{\lambda_{j,x} : j = 1, \ldots, J\}$ the spectrum of the thermal state. With the change of variable

$$\beta = V_x^\dagger\alpha, \qquad (5.34)$$

$\Phi_x(\alpha) = \Phi_x(V_x\beta)$ becomes separable in terms of β. Define also a unitary operator $\hat{U}_x$ by

$$\hat{U}_x^\dagger\hat{a}_j\hat{U}_x = \sum_k V_{jk,x}\hat{a}_k \equiv \hat{g}_{j,x}, \qquad (5.35)$$

such that $|\alpha\rangle = |V_x\beta\rangle = \hat{U}_x|\beta\rangle$. ρ_x can then be expressed as

$$\rho_x = \sum_{\boldsymbol{m}} p_x(\boldsymbol{m})|\boldsymbol{m},g\rangle\langle\boldsymbol{m},g|, \qquad (5.36)$$

where

$$p_x(\boldsymbol{m}) = \prod_j \frac{1}{1+\lambda_{j,x}}\left(\frac{\lambda_{j,x}}{1+\lambda_{j,x}}\right)^{m_j} \qquad (5.37)$$

is separable into a product of Bose-Einstein distributions and

$$|\boldsymbol{m},g\rangle \equiv \hat{U}_x|\boldsymbol{m}\rangle = \left[\prod_j \frac{(\hat{g}_{j,x}^\dagger)^{m_j}}{\sqrt{m_j!}}\right]|\boldsymbol{0}\rangle \qquad (5.38)$$

is a Fock state with respect to the optical modes defined by Eq. (5.35). I call these optical modes the eigenmodes of the thermal state. Now suppose that only the spectrum $\{\lambda_{j,x}\}$ depends on the unknown parameter x, while V, U, and thus $\{\hat{g}_j\}$ do not, meaning that the eigenmodes are fixed. This assumption applies to the thermometry problem studied in Ref. [75] but does not apply to the stellar-interferometry problem studied in Ref. [74] or the subdiffraction-imaging problem studied in Refs. [76, 77, 78, 79], because the eigenmodes in the latter two cases vary with x. With fixed eigenmodes, I can define a more fine-grained x-independent projector as

$$\Pi_{\boldsymbol{m}} = |\boldsymbol{m},g\rangle\langle\boldsymbol{m},g|, \qquad (5.39)$$

and apply Corollary 5 to Eq. (5.36). Plugging Eq. (5.39) into Eq. (5.22) leads to the Rao-Blackwell estimator

$$\mathcal{G}_*B = \sum_{\boldsymbol{m}} \langle\boldsymbol{m},g|B|\boldsymbol{m},g\rangle |\boldsymbol{m},g\rangle\langle\boldsymbol{m},g|, \qquad (5.40)$$

which can be implemented by counting the photons in the eigenmodes and using $\langle\boldsymbol{m},g|B|\boldsymbol{m},g\rangle$ as the estimator. Equation (5.40) then implies the following corollary.

Corollary 6. *Suppose that a real scalar parameter $a(x)$ of the spectrum $\{\lambda_{j,x}\}$ of a thermal-state family is to be estimated and the eigenmodes are parameter-independent. Given any measurement, any estimator, and the resulting local error MSE_x, there exists an estimator with eigenmode photon counting that performs at least as well for all $x \in \mathcal{X}$.*

Proof. Corollary 3 means that only von Neumann measurements need to be considered. Any estimator with any von Neumann measurement can be Rao-Blackwellized to become Eq. (5.40), which can be implemented by eigenmode photon counting. Corollary 5 then guarantees that the MSE'_x achieved by Eq. (5.40) can do at least as well for all $x \in \mathcal{X}$. $\qquad\square$

In this example, the family of density operators given by Eq. (5.36) and the $\mathcal{G}_* B$ given by Eq. (5.40) happen to commute with one another, but the original estimator B need not commute with the others, unlike Sinha's assumption in Ref. [64]; see Appendix H for a brief discussion of his theory. Take homodyne detection for example. An estimator constructed from homodyne detection can be framed as

$$B = b(\hat{q}), \qquad (5.41)$$

where $\hat{q}$ is a vectoral quadrature operator that is a linear function of $\{\hat{a}_j\}$. Equation (5.41) does not commute with Eq. (5.39) in general, but Corollary 6 still applies to it.

To demonstrate the possible improvement through an even more specific example, suppose that the spectrum is flat and $a(x) = \lambda_{j,x} = x$, the mean photon number per mode, is the parameter of interest. With homodyne detection, Eq. (5.41) is an unbiased estimator of x if

$$B = \frac{1}{J} \sum_j \hat{q}_j^2 - \frac{1}{2}, \qquad \hat{q}_j = \frac{\hat{g}_j + \hat{g}_j^\dagger}{\sqrt{2}}. \qquad (5.42)$$

The Rao-Blackwell estimator given by Eq. (5.40), on the other hand, can be expressed as

$$\mathcal{G}_* B = \frac{1}{J} \sum_j \hat{g}_j^\dagger \hat{g}_j. \qquad (5.43)$$

With the thermal state, it is straightforward to show that

$$\mathrm{MSE}_x = \frac{2}{J}\left(x + \frac{1}{2}\right)^2, \quad \mathrm{MSE}'_x = \frac{1}{J}\left(x^2 + x\right), \qquad (5.44)$$

which are plotted in Fig. 3, demonstrating the domination of the Rao-Blackwell estimator.

Corollary 6 is reminiscent of the optimality of photon counting for thermometry proved in Ref. [75] in terms of a quantum Cramér-Rao bound. Corollary 6 is more general because it applies directly to the local
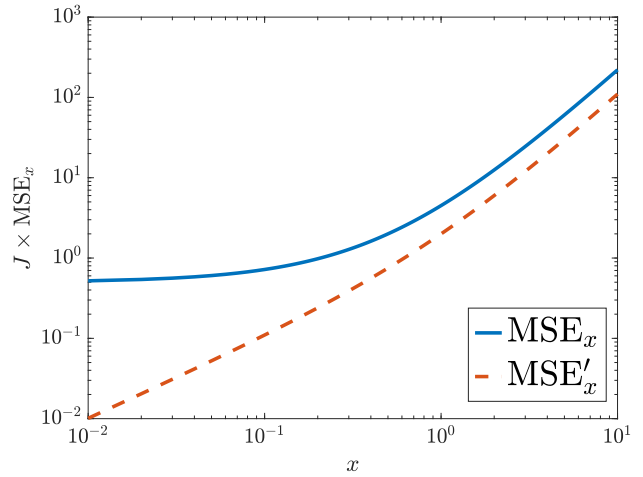

Figure 3: Comparison of the mean-square error obtained by homodyne detection (MSE_x) and that by photon counting (MSE'_x) in estimating the mean photon number per mode x of a thermal state. The plot is in log-log scale, both axes are dimensionless, and the errors are normalized with respect to J, the number of optical modes. The improvement can be regarded as a result of the quantum Rao-Blackwellization.

mean-square error of any biased or unbiased estimator and allows the parametrization of the spectrum $\{\lambda_{j,x}\}$ and the parameter of interest $a(x)$ to be general. The superiority of photon counting over homodyne detection for random displacement models has also been noted in many other contexts [80, 81, 82, 83], although those works, like Ref. [75], rely on the quantum Cramér-Rao bound as well.

If the eigenmodes vary with x, as in the problems of stellar interferometry [74] and subdiffraction imaging [76, 77, 78, 79], then Eq. (5.40) may not be a valid estimator, because x is unknown and the measurement may not be realizable. It is an interesting open question whether the quantum Rao-Blackwell theorem can offer any insight about those problems as well, beyond the optimality of the direct-sum form in Corollary 5. I speculate on two potential directions of future research:

1. Even if the $\mathcal{G}_*$ map may not be constant in general, $\mathcal{G}_* B$ for a particular B may happen to be constant and still a valid estimator.

2. Even if $\mathcal{G}_* B$ varies with x, Eq. (5.3) remains valid and can be used as a lower bound on MSE_x, in which case $\mathrm{MSE}_x \geq \mathrm{MSE}'_x$ is an oracle inequality. An estimator that approximates $\mathcal{G}_* B$, via an adaptive protocol for example [22], may still enjoy an error close to MSE'_x.

6 Conclusion

This work cements the Jordan-product GCE and the associated divergence as essential concepts in quantum metrology. In the Bayesian setting, the GCE

is found to relate the optimal estimators for a sequence of channels. In the frequentist setting, the GCE is found to give a quantum Rao-Blackwell theorem, which can improve a quantum estimator in the same manner as the classical version does and reveal the optimal forms of the estimators in common scenarios. In both settings, the divergence is found to play a significant role in determining the gap between the estimation errors before and after a channel is applied. Given these operational meanings, even the purists [39, 40] can no longer dismiss the GCE and the divergence as pointless concepts. For the more open minds, the concepts have unveiled a new suite of methods for the study of decoherence and the design of better measurements in quantum metrology.

Many open problems remain. First, it should be possible to generalize the theory here rigorously for infinite-dimensional Hilbert spaces. Second, it may be possible to generalize the quantum Rao-Blackwell theorem here for other convex loss functions beyond the square loss, in the same manner as the classical version [5] or Sinha's versions [64]. Third, there may be a deeper relation between Petz's sufficiency and the constant GCE condition desired here, beyond the specific examples in this work. Fourth, there should be no shortage of further interesting examples and applications of the theory here for quantum metrology. Last but not the least, the strategy of using quantum metrology to give operational meanings to GCEs may be generalizable for other versions of GCEs and other metrological tasks, such as multiparameter estimation, thus expanding the fundamental role of GCEs in both quantum metrology and quantum probability theory.

Acknowledgment

A Quantum prediction

In analogy with Eq. (2.14), $\mathcal{F}^*$, the Hilbert-Schmidt adjoint of the CPTP map $\mathcal{F}$ that models a channel, obeys the following interesting formula for any $\mathcal{E}$:

$$\mathcal{F}^*B = \underset{A \in L_2(\sigma)}{\arg\min}\, D_{\sigma,\mathcal{F}}(A, B). \qquad (\text{A.1})$$

In other words, $\mathcal{F}^*$ is the optimal prediction in the same way $\mathcal{F}_*$ is the optimal retrodiction. The notations of $\mathcal{F}^*$ and $\mathcal{F}_*$ also coincide with those of the pullback and the pushforward in differential geometry [84], respectively, and indeed $\mathcal{F}^*$ and $\mathcal{F}_*$ behave like those operations.

With the adjoint relation between $\mathcal{F}^*$ and $\mathcal{F}_*$ given by Eq. (2.15) in terms of the weighted inner product, $\mathcal{F}^*$ can be shown to obey

$$\mathcal{E}_\sigma \mathcal{F}^* B = (\mathcal{F}_*)^* \mathcal{E}_{\mathcal{F}\sigma} B, \qquad (\text{A.2})$$

in analogy with Eq. (2.16), while the minimum divergence in Eq. (A.1) can be expressed as

$$\mathcal{D}_{\sigma,\mathcal{F}}(\mathcal{F}^*B, B) = \min_{A \in L_2(\sigma)} D_{\sigma,\mathcal{F}}(A, B)$$
$$= \|B\|_{\mathcal{F}\sigma}^2 - \|\mathcal{F}^*B\|_\sigma^2, \qquad (\text{A.3})$$

in analogy with Eq. (2.17). The divergence given by Eq. (2.10) can also be rewritten in a time-symmetric form as

$$D_{\sigma,\mathcal{F}}(A, B) = \langle A - \mathcal{F}^*B, A \rangle_\sigma + \langle B - \mathcal{F}_*A, B \rangle_{\mathcal{F}\sigma}. \qquad (\text{A.4})$$

Together with Eqs. (2.14)–(2.17), these formulas complete a satisfying time-symmetric theory of quantum inference.

B Classical conditional expectation

The classical concepts in this Appendix are all special cases of the quantum concepts in Sec. 2 and Appendix A; see Table 1 for the correspondence.

Let X and Y be classical random variables with countable sample spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. Suppose that X is generated at an earlier time than Y. Let $P_X(x)$ be the probability distribution of X and $P_{Y|X}(y|x)$ be the probability distribution of Y conditioned on $X = x$. The unconditional distribution of Y is given by

$$P_Y(y) = \sum_x P_{Y|X}(y|x) P_X(x), \qquad (\text{B.1})$$

which can be expressed as $P_Y = \mathcal{F}P_X$. The divergence between two complex random variables $a(X)$ and $b(Y)$ can be defined as

$$D_{P_X,\mathcal{F}}(a, b) \equiv \sum_{x,y} |a(x) - b(y)|^2 P_{Y|X}(y|x) P_X(x) \qquad (\text{B.2})$$
$$= \|a\|_{P_X}^2 - 2\operatorname{Re}\langle \mathcal{F}^*b, a \rangle_{P_X} + \|b\|_{P_Y}^2, \qquad (\text{B.3})$$

where the inner product and the norm are defined in Table 1 and $\mathcal{F}^*$ is the adjoint of $\mathcal{F}$ with respect to the unweighted inner product in Table 1, such that

$$(\mathcal{F}^*b)(x) = \sum_y b(y) P_{Y|X}(y|x). \qquad (\text{B.4})$$

In other words, $\mathcal{F}^*b$ is the conditional expectation of $b(Y)$ given X.

Let $L_2(P)$ be the Hilbert space of random variables in terms of the inner product $\langle \cdot, \cdot \rangle_P$ defined in Table 1.

Concept	Quantum	Classical
Prior state	Density operator σ	Probability distribution $P_X(x)$
Channel	CPTP map $\mathcal{F}$	$\sum_x P_{Y\mid X}(y\mid x)(\cdot)$
Unweighted inner product	Hilbert-Schmidt $\langle A, B\rangle_{\mathrm{HS}} \equiv$ tr $A^\dagger B$	$\langle a, b\rangle \equiv \sum_\lambda [a(\lambda)]^* b(\lambda)$
Weighted inner product	$\langle A, B\rangle_\rho \equiv$ tr $A^\dagger \mathcal{E}_\rho B$	$\langle a, b\rangle_P \equiv \sum_\lambda [a(\lambda)]^* b(\lambda) P(\lambda)$
Weighted norm	$\|A\|_\rho \equiv \sqrt{\langle A, A\rangle_\rho}$	$\|a\|_P \equiv \sqrt{\langle a, a\rangle_P}$
Divergence between variables	$D_{\sigma, \mathcal{F}}(A, B)$ (Definition 1)	$\sum_{x,y} \|a(x) - b(y)\|^2 P_{XY}(x, y)$
Optimal prediction	$\mathcal{F}^*$ (Hilbert-Schmidt adjoint of $\mathcal{F}$)	$\sum_y (\cdot) P_{Y\mid X}(y\mid x)$
Optimal retrodiction	$\mathcal{F}_*$ (Definition 2)	$\sum_x (\cdot) P_{X\mid Y}(x\mid y)$

Table 1: Concepts in quantum inference and their classical versions.

The conditional expectation $\mathcal{F}^* b$ is the optimal predictor of $b(Y)$ as a function of X, viz.,

$$\mathcal{F}^* b = \underset{a \in L_2(P_X)}{\arg\min} D_{P_X, \mathcal{F}}(a, b). \tag{B.5}$$

Similarly, the conditional expectation of $a(X)$ given Y is the optimal retrodictor of $a(X)$ as a function of Y, viz.,

$$\mathcal{F}_* a = \underset{b \in L_2(P_Y)}{\arg\min} D_{P_X, \mathcal{F}}(a, b), \tag{B.6}$$

$$(\mathcal{F}_* a)(y) = \sum_x a(x) P_{X\mid Y}(x\mid y), \tag{B.7}$$

$$P_{X\mid Y}(x\mid y) = \frac{P_{Y\mid X}(y\mid x) P_X(x)}{P_Y(y)}. \tag{B.8}$$

Equation (B.8) is, of course, the Bayes theorem. It is straightforward to show that the formulas presented thus far are special cases of the quantum formulas in Sec. 2 and Appendix A; this is done by assuming the diagonal forms

$$\sigma = \sum_x P_X(x) |x\rangle \langle x|, \tag{B.9}$$

$$A = \sum_x a(x) |x\rangle \langle x|, \tag{B.10}$$

$$\mathcal{F}\sigma = \sum_{y,x} P_{Y\mid X}(y\mid x) \langle x| \sigma |x\rangle |y\rangle \langle y|, \tag{B.11}$$

$$B = \sum_y b(y) |y\rangle \langle y|, \tag{B.12}$$

where $\{|x\rangle : x \in \mathcal{X}\}$ is an orthonormal basis of $\mathcal{H}_1$ and $\{|y\rangle : y \in \mathcal{Y}\}$ is an orthonormal basis of $\mathcal{H}_2$.

Since the joint distribution $P_{XY}(x, y) = P_{Y\mid X}(y\mid x) P_X(x)$ is another probability distribution, Eq. (B.2) can be written as the squared norm

$$D_{P_X, \mathcal{F}}(a, b) = \|a - b\|_{P_{XY}}^2, \tag{B.13}$$

and since $L_2(P_X)$ and $L_2(P_Y)$ are subspaces of $L_2(P_{XY})$, the Hilbert projection theorem implies that the conditional expectation is a projection, viz.,

$$\mathcal{F}^* b = \underset{a \in L_2(P_X)}{\arg\min} \|a - b\|_{P_{XY}}^2 = \Pi[b|L_2(P_X)], \tag{B.14}$$

$$\mathcal{F}_* a = \underset{b \in L_2(P_Y)}{\arg\min} \|a - b\|_{P_{XY}}^2 = \Pi[a|L_2(P_Y)], \tag{B.15}$$

where $\Pi[u|V]$ is the projection of a Hilbert-space element u into a subspace V. In the quantum case, $\mathcal{F}^*$ and $\mathcal{F}_*$ are also projections into appropriate Hilbert spaces if the equality condition in Eq. (2.13) holds.

Because X is assumed to be generated earlier than Y and experiments follow an arrow of time, it is more straightforward to obtain $P_{Y\mid X}$ by experiments than $P_{X\mid Y}$. This is the reason why one commonly assumes that $P_{Y\mid X}$ rather than $P_{X\mid Y}$ is known in an inference problem, and the prediction formula in terms of $P_{Y\mid X}$ looks simpler than the retrodiction formula in terms of the Bayes theorem. Probability theory in itself is agnostic to the arrow of time, and if $P_{X\mid Y}$ is given instead, the retrodiction formula would look simpler and the prediction formula would require the Bayes theorem. The same arrow of time is commonly assumed in quantum theory, where the $\mathcal{F}$ map models the evolution of a quantum system forward in time. It is therefore unsurprising that the prediction map $\mathcal{F}^*$ in terms of $\mathcal{F}$ also looks simpler than the retrodiction map $\mathcal{F}_*$ in terms of Eq. (2.16), which is a generalization of the Bayes theorem. If the map $(\mathcal{F}_*)^*$ is given instead, then the retrodiction map $\mathcal{F}_*$ is simply its Hilbert-Schmidt adjoint, while the prediction map $\mathcal{F}^*$ is given by the complicated Eq. (A.2).

C GCE for Gaussian systems

I first briefly review the theory of quantum Gaussian systems, following Chap. 12 in Ref. [50]. Let $\mathcal{H}_1$ be the Hilbert space for s bosonic modes. On $\mathcal{H}_1$, define the canonical observables as

$$Q \equiv \begin{pmatrix} q_1 & p_1 & \ldots & q_s & p_s \end{pmatrix}^\top, \quad [q_j, p_k] = i\delta_{jk}, \tag{C.1}$$

and the Weyl operator as

$$W(z) \equiv \exp(iQ^\top z), \tag{C.2}$$

$$z \equiv \begin{pmatrix} x_1 & y_1 & \ldots & x_s & y_s \end{pmatrix}^\top \in \mathbb{R}^{2s}, \tag{C.3}$$

where $\top$ denotes the transpose. If σ is a Gaussian state, its characteristic function can be expressed as

$$\phi(z) \equiv \operatorname{tr} \sigma W(z) = \exp\left(im^\top z - \frac{1}{2} z^\top \Sigma z\right), \tag{C.4}$$

where $m \in \mathbb{R}^{2s}$ is the mean vector and $\Sigma \in \mathbb{R}^{2s \times 2s}$ is the covariance matrix of the Gaussian state. Σ is symmetric, positive-semidefinite, and must observe an uncertainty relation that need not concern us here.

Similar to the preceding definitions, let $\mathcal{H}_2$ be the Hilbert space for t bosonic modes and define $\tilde{Q}$ and $\tilde{W}(\zeta)$ as the canonical observables and the Weyl operator on $\mathcal{H}_2$, respectively. If $\mathcal{F} : \mathcal{O}(\mathcal{H}_1) \to \mathcal{O}(\mathcal{H}_2)$ is a CPTP map that models a Gaussian channel, it can be defined by

$$\mathcal{F}^* \tilde{W}(\zeta) = f(\zeta) W(F^\top \zeta), \tag{C.5}$$

$$f(\zeta) = \exp\left(il^\top \zeta - \frac{1}{2}\zeta^\top R\zeta\right), \tag{C.6}$$

where $F \in \mathbb{R}^{2t \times 2s}$ is a transition matrix, $l \in \mathbb{R}^{2t}$ is the mean displacement introduced by the channel, and $R \in \mathbb{R}^{2s \times 2s}$ is the channel covariance matrix. F and R must obey a certain matrix inequality for the map to be CPTP, but again the inequality need not concern us here. With the Gaussian input state and the Gaussian channel, the output state remains Gaussian and its characteristic function is given by

$$\tilde{\phi}(\zeta) \equiv \operatorname{tr}(\mathcal{F}\sigma)\tilde{W}(\zeta) = f(\zeta)\phi(F^\top \zeta) \tag{C.7}$$

$$= \exp\left(i\tilde{m}^\top \zeta - \frac{1}{2}\zeta^\top \tilde{\Sigma}\zeta\right), \tag{C.8}$$

$$\tilde{m} = Fm + l, \tag{C.9}$$

$$\tilde{\Sigma} = F\Sigma F^\top + R. \tag{C.10}$$

An explicit GCE formula can now be presented.

Proposition 1. *Assume a Gaussian state defined by Eq. (C.4), a Gaussian channel defined by Eqs. (C.5) and (C.6), a quadrature operator given by*

$$A = u^\top Q, \quad u = \begin{pmatrix} u_1 & \dots & u_{2s} \end{pmatrix}^\top \in \mathbb{R}^{2s}, \quad \text{(C.11)}$$

and the $\mathcal{E}$ map given by the Jordan product in Eq. (2.7). A solution to the GCE is

$$\mathcal{F}_* A = u^\top \left[m + K\left(\tilde{Q} - Fm - l\right)\right], \tag{C.12}$$

$$K \equiv \Sigma F^\top \left(F\Sigma F^\top + R\right)^{-1}, \tag{C.13}$$

while the divergence is

$$D_{\sigma,\mathcal{F}}(A, \mathcal{F}_* A) = u^\top \left(\Sigma - KF\Sigma\right) u. \tag{C.14}$$

See Appendix J for the proof.

It is interesting to note that Eqs. (C.12)–(C.14) are identical to the formulas for the classical conditional expectation $\mathbb{E}(A|Y)$ and its mean-square error $\mathbb{E}\{[\mathbb{E}(A|Y) - A]^2\}$ when $A = u^\top X$, $Y = FX + Z$, and $X \sim N(m, \Sigma)$ and $Z \sim N(l, R)$ are independent normal random variables [51]. Here, the canonical observables Q and $\tilde{Q}$ play the roles of X and Y, respectively. When $\mathcal{F}$ is a measurement map, similar formulas have been derived in Refs. [25, 26, 27, 18] and may be useful for studying waveform estimation [42] beyond the stationary assumption.

D von Neumann measurement

Let B be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Then B admits the spectral representation [85, 46]

$$B = \int b\Pi(db) \tag{D.1}$$

in terms of a projection-valued measure Π on the Borel space $(\mathbb{R}, \Sigma_\mathbb{R})$. A projection-valued measure, also called an orthogonal resolution of identity, is a POVM on $(\mathcal{Y}, \Sigma_\mathcal{Y})$ with the additional properties that $\Pi(C)$ for any $C \in \Sigma_\mathcal{Y}$ is a projection operator and $\Pi(C_1)\Pi(C_2) = 0$ if $C_1 \cap C_2 = \varnothing$ [46]. The projection-valued measure in Eq. (D.1) is unique to B in the sense that no other projection-valued measure can satisfy Eq. (D.1) for a given B. A von Neumann measurement of B is defined as a measurement with the projection-valued measure Π of B as the POVM.

If $\mathcal{H}$ is finite-dimensional, B has a finite number of eigenvalues, and the spectral representation is much simpler and can be written as

$$B = \sum_{b \in \Lambda_B} b\Pi(b), \tag{D.2}$$

where $\Lambda_B \subset \mathbb{R}$ is the set of B's eigenvalues and $\{\Pi(b) : b \in \Lambda_B\}$ are projection operators that obey $\Pi(b)\Pi(c) = \delta_{bc}\Pi(b)$ and $\sum_b \Pi(b) = I$. The set of outcomes can then be restricted to Λ_B.

Consider now a converse situation, where the POVM is given by a projection-valued measure Π' on $(\mathcal{Y}, \Sigma_\mathcal{Y})$ and each outcome $y \in \mathcal{Y}$ is processed by a function $b : \mathcal{Y} \to \mathbb{R}$ to produce a final outcome $b(y)$. The measurement Π' together with the data processing by $b(y)$ can be interpreted as a von Neumann measurement of

$$B = \int b(y)\Pi'(dy). \tag{D.3}$$

This is because Eq. (D.3) can be expressed in the spectral representation given by Eq. (D.1), with

$$\Pi(C) = \Pi'[b^{-1}(C)], \quad b^{-1}(C) \equiv \{y : b(y) \in C\}, \tag{D.4}$$

and the probability $\operatorname{tr}\Pi'[b^{-1}(C)]\rho$ of each event $C \in \Sigma_\mathbb{R}$ generated by (Π', b) coincides with the probability $\operatorname{tr}\Pi(C)\rho$ from a von Neumann measurement of B. If $\mathcal{Y}$ is countable, then $B = \sum_y b(y)\Pi'(y)$ can be expressed as Eq. (D.2), with

$$\Lambda_B = \{b(y) : y \in \mathcal{Y}\}, \quad \Pi(c) = \sum_{y : b(y) = c} \Pi'(y). \tag{D.5}$$

E The dynamic-programming algorithm

The dynamic-programming algorithm is based on the so-called principle of optimality for Markov decision

processes. For the process given by Eqs. (4.12) and (4.13), the principle states that, if $(\tilde{\mathcal{F}}_1, \ldots, \tilde{\mathcal{F}}_N)$ are the optimal "controls" that minimize the total "cost" given by Eq. (4.13), then $(\tilde{\mathcal{F}}_k, \ldots, \tilde{\mathcal{F}}_N)$ must also be the optimal controls that minimize the "cost-to-go" at time k, defined as

$$J_k \equiv \sum_{n=k}^{N} g(s_n, \mathcal{F}_n). \tag{E.1}$$

Based on this principle, the algorithm starts by choosing the final control $\mathcal{F}_N$ that minimizes the cost-to-go $J_N = g(s_N, \mathcal{F}_N)$. Let this minimum be

$$\tilde{J}_N(s_N) \equiv \min_{\mathcal{F}_N} J_N = \min_{\mathcal{F}_N} g(s_N, \mathcal{F}_N). \tag{E.2}$$

The optimal $\mathcal{F}_N$ obtained this way is a function $\tilde{\mathcal{F}}_N(s_N)$ of the state s_N. Next, write $\tilde{J}_N(s_N) = \tilde{J}_N[f(s_{N-1}, \mathcal{F}_{N-1})]$ using Eq. (4.12) and find the control $\mathcal{F}_{N-1}$ that minimizes the cost-to-go at time $k = N - 1$, given by

$$\tilde{J}_{N-1}(s_{N-1}) \equiv \min_{(\mathcal{F}_{N-1}, \mathcal{F}_N)} J_{N-1} \tag{E.3}$$

$$= \min_{\mathcal{F}_{N-1}} \Big\{ g(s_{N-1}, \mathcal{F}_{N-1}) + $$

$$\tilde{J}_N \left[f(s_{N-1}, \mathcal{F}_{N-1}) \right] \Big\}. \tag{E.4}$$

The optimal $\mathcal{F}_{N-1}$ is again a function $\tilde{\mathcal{F}}_{N-1}(s_{N-1})$ of the state s_{N-1}. This procedure continues with

$$\tilde{J}_k(s_k) \equiv \min_{(\mathcal{F}_k, \ldots, \mathcal{F}_N)} J_k \tag{E.5}$$

$$= \min_{\mathcal{F}_k} \left\{ g(s_k, \mathcal{F}_k) + \tilde{J}_{k+1} \left[f(s_k, \mathcal{F}_k) \right] \right\}, \tag{E.6}$$

$$\tilde{\mathcal{F}}_k(s_k) = \arg\min_{\mathcal{F}_k} \{\ldots\} \tag{E.7}$$

for $k = N - 2, N - 3, \ldots$ until $k = 1$, when the complete optimal control law $(\tilde{F}_1(s_1), \ldots, \tilde{F}_N(s_N))$ has been found and $\tilde{J}_1(s_1)$ is the minimum total cost.

F Classical Rao-Blackwell theorem

To derive the classical Rao-Blackwell theorem from the quantum theorem given by Theorem 3, assume the diagonal forms

$$\rho_x = \sum_y P_{Y|X}(y|x) |y\rangle \langle y|, \tag{F.1}$$

$$B = \sum_y b(y) |y\rangle \langle y|, \tag{F.2}$$

$$\mathcal{G}\rho_x = \sum_{z,y} P_{Z|Y}(z|y) \langle y| \rho_x |y\rangle |z\rangle \langle z| \tag{F.3}$$

$$= \sum_z P_{Z|X}(z|x) |z\rangle \langle z|, \tag{F.4}$$

$$P_{Z|X}(z|x) = \sum_y P_{Z|Y}(z|y) P_{Y|X}(y|x), \tag{F.5}$$

where $\{|y\rangle : y \in \mathcal{Y}\}$ is an orthonormal basis of $\mathcal{H}_2$, $\{|z\rangle : z \in \mathcal{Z}\}$ is an orthonormal basis of $\mathcal{H}_3$, X, Y, and Z are classical variables, and $P_{O|O'}$ is the probability distribution of O conditioned on a value of O'. For the estimation problem, X is the hidden parameter fixed at $X = x$, Y is the observation, and Z is a statistic generated from Y without knowing X, such that $P_{Z|Y,X}(z|y,x) = P_{Z|Y}(z|y)$. The parameter of interest $a(x)$ is assumed to be real for simplicity. The local error given by Eq. (5.1) becomes

$$\mathrm{MSE}_x = \sum_y [b(y) - a(x)]^2 P_{Y|X}(y|x), \tag{F.6}$$

which agrees with the classical definition. Equation (2.16) for the GCE $\mathcal{G}_* B$ becomes

$$\mathcal{E}_{\mathcal{G}\rho_x} \mathcal{G}_* B = \mathcal{G}\mathcal{E}_{\rho_x} B, \tag{F.7}$$

a solution of which is

$$\mathcal{G}_* B = \sum_z c(z,x) |z\rangle \langle z|, \tag{F.8}$$

$$c(z,x) = \sum_y P_{Y|Z,X}(y|z,x) b(y), \tag{F.9}$$

$$P_{Y|Z,X}(y|z,x) = \frac{P_{Z|Y}(z|y) P_{Y|X}(y|x)}{P_{Z|X}(z|x)}. \tag{F.10}$$

$c(z,x)$ is the expectation of $b(Y)$ conditioned on $Z = z$ and $X = x$. If $c(z,x) = c(z)$ does not depend on x, then it is a valid estimator as a function of the statistic Z, and its error as per Eq. (5.2) becomes

$$\mathrm{MSE}'_x = \sum_z [c(z) - a(x)]^2 P_{Z|X}(z|x). \tag{F.11}$$

It follows from Theorem 3 and Eq. (2.17) that the difference between Eqs. (F.6) and (F.11) is

$$\mathrm{MSE}_x - \mathrm{MSE}'_x$$
$$= \sum_y [b(y)]^2 P_{Y|X}(y|x) - \sum_z [c(z)]^2 P_{Z|X}(z|x), \tag{F.12}$$

which is nonnegative and coincides with a form of the Rao-Blackwell theorem (see, for example, Problem 1.7.9 on p. 73 in Ref. [5]).

Z is called a sufficient statistic for the estimation problem if the distribution $P_{Y|Z,X}(y|z,x)$ given by Eq. (F.10) does not depend on x [5]. Then Eq. (F.9) also does not depend on x for any $b(y)$ and is always a valid estimator.

G Comparison of the Bayesian and frequentist settings

Both the monotonicity of the Bayesian error given by Corollary 1 and the error reduction due to the quantum Rao-Blackwell theorem in Theorem 3 are unsurprising results given their classical origins, but they

may be confusing in that they seem to say opposite things about the effect of a channel. I offer a clarification here.

First, note that the Bayesian setting concerns the "global" error $D_{\sigma,\mathcal{F}}(A, B)$ only, whereas the frequentist setting concerns the local error MSE_x as a function of the unknown parameter x. The global error is a cruder measure because it is only an average of the local error given by

$$D_{\sigma,\mathcal{F}}(A, B) = \sum_x P_X(x)\,\mathrm{MSE}_x, \qquad (\mathrm{G.1})$$

assuming Eqs. (4.3) and (5.1).

Second, the Bayesian results in Sec. 4, and Corollary 1 in particular, concern only the estimators $\mathcal{F}_*A$ and $\mathcal{G}_*\mathcal{F}_*A$ that are optimal with respect to the global error. Theorem 3 in the frequentist setting, on the other hand, is about the local errors of an estimator B and its Rao-Blackwellization $\mathcal{G}_*B$, with no special assumptions about the original estimator B. The theorem also says nothing about whether the Rao-Blackwell estimator is optimal in the global sense, only that it is at least as good as the original.

Third, note that the Personick estimators considered in the Bayesian setting do not depend on the unknown parameter and are naturally realizable, and Corollary 1 applies to any channel. In the frequentist setting, the channel and the Rao-Blackwell estimator must be parameter-independent for the measurement to be realizable, so there is a stringent requirement on the $\mathcal{G}$ channel for the improvement to be realizable, let alone significant.

In practice, the classical Rao-Blackwellization is typically used to improve an initial estimator design that is not expected to be optimal or even good in any sense; the derivation of the U-statistics [72] is a representative example. If the initial estimator is already optimal in the Bayesian sense, then the Rao-Blackwellization cannot offer any improvement almost everywhere with respect to the prior P_X.

Corollary 7. *Assume the Bayesian problem specified by Eqs. (4.1) and (4.2) and let $B = \mathcal{F}_*^\sigma A$ be the Personick estimator. If another CPTP map $\mathcal{G}$ is applied and both $\mathcal{G}$ and the Rao-Blackwell estimator $\mathcal{G}_*^{\rho_x}B$ in Theorem 3 do not depend on x, then $\mathcal{G}_*^{\rho_x}B$ is a solution to the final Personick estimator $\mathcal{G}_*^{\mathcal{F}\sigma}B$. Moreover, both the Bayesian error and the local error remain the same after the $\mathcal{G}$ channel, in the sense of*

$$D_{\sigma,\mathcal{F}}(A, \mathcal{F}_*^\sigma A) = \sum_x P_X(x)\,\mathrm{MSE}_x = \sum_x P_X(x)\,\mathrm{MSE}_x'$$
$$= D_{\sigma,\mathcal{GF}}(A, \mathcal{G}_*^{\mathcal{F}\sigma}\mathcal{F}_*^\sigma A), \qquad (\mathrm{G.2})$$
$$\mathrm{MSE}_x = \mathrm{MSE}_x' \quad \textit{almost everywhere } P_X. \tag{G.3}$$

See Appendix J for the proof.

H   Comparison with some prior works

Although Sinha's formalism in Ref. [64] is applicable to infinite dimensions and any convex loss function, he makes heavy assumptions about the commutativity of all the involved operators. To explain, consider his definition of a sufficient-statistic observable $S \in \mathcal{O}(\mathcal{H}_2)$, which is key to his quantum Rao-Blackwell theorems. Assuming that the spectrum of S, denoted by $\Lambda_S \subseteq \mathbb{R}$, is countable for simplicity, S can be expressed in the spectral representation

$$S = \sum_{s \in \Lambda_S} s\Pi_s, \qquad (\mathrm{H.1})$$

where Π is a projection-valued measure. Consider any observable B that commutes with S. By the spectral theorem (see Chap. VII in Ref. [85]), B commutes with all $\{\Pi_s\}$, meaning that B can be expressed as

$$B = \sum_s \Pi_s B \Pi_s = \bigoplus_s B_s, \qquad (\mathrm{H.2})$$

where $B_s \in \mathcal{O}(\Pi_s\mathcal{H}_2)$ is equal to $\Pi_s B \Pi_s$ on the subspace $\Pi_s\mathcal{H}_2$. Given a family of density operators $\{\rho_x : x \in \mathcal{X}\}$, Sinha defines a sufficient S by the existence of a positive function $\varphi : \Lambda_S \times \mathcal{X} \to \mathbb{R}_+$ and a parameter-independent positive-semidefinite operator $\sigma_s \in \mathcal{O}(\Pi_s\mathcal{H}_2)$ such that, for any B that commutes with S,

$$\mathrm{tr}\,\rho_x B = \sum_s \varphi(s, x)\,\mathrm{tr}\,\sigma_s B_s. \qquad (\mathrm{H.3})$$

Since Sinha considers only POVMs that commute with S in his Rao-Blackwell theorems (see Theorems 4.4 and 5.2 in Ref. [64]), Eq. (H.3) implies that each ρ_x can be factorized as

$$\rho_x = \bigoplus_s \varphi(s, x)\sigma_s, \qquad (\mathrm{H.4})$$

meaning that $\{\rho_x\}$ all commute with S (see Remark 3.3 in Ref. [64]). In fact, Eq. (H.4) also implies that $\{\rho_x\}$ commute with one another. Such assumptions are extremely restrictive, as noncommutativity is precisely what distinguishes quantum probability theory from the classical version, and Sinha's restrictions to it are simply unprecedented in quantum metrology [53, 46, 15]. This work, on the other hand, does not impose any commutativity requirements on the operators. The key advance here is the use of the GCE formalism in Secs. 2 and 3 that generalizes classical probability theory from the ground level for noncommuting operators, so that Sinha's commutativity assumptions are never necessary.

Sinha also avoids any explicit use of quantum conditional expectations (see Remark 5.3(i) in Ref. [64]) or even CPTP maps. The use of a GCE in this work, on the other hand, makes Theorem 3 a more natural generalization of the classical theorem. As the

conditional expectation is a standard and crucial step in the classical Rao-Blackwellization, the GCE can be similarly instrumental for the quantum case, as demonstrated by the corollaries and examples in this paper.

It is straightforward to show that a measurement of Sinha's sufficient-statistic observable is a special case of a sufficient channel.

Proposition 2. *Given Eqs. (H.2) and (H.4) and assuming the measurement map*

$$\mathcal{G}\rho_x = \sum_s (\operatorname{tr} \Pi_s \rho_x \Pi_s)\Pi_s, \tag{H.5}$$

which represents the state of the classical random variable obtained by a measurement of S, a solution to any GCE is

$$\mathcal{G}_* B = \sum_s \frac{\operatorname{tr} \sigma_s B_s}{\operatorname{tr} \sigma_s} \Pi_s, \tag{H.6}$$

which is independent of x.

See Appendix J for the proof.

Equation (H.6) is used implicitly in Sinha's Rao-Blackwell theorems and Theorem 3 here also applies to it. Physically, Proposition 2 means that, when the density operators are given by Eq. (H.4) and the original estimator is given by Eq. (H.2), one can measure the sufficient-statistic observable S given by Eq. (H.1), and the Rao-Blackwell estimator as a function of the outcome s is $(\operatorname{tr} \sigma_s B_s)/\operatorname{tr} \sigma_s$.

Another relevant prior work is Ref. [65] by Łuczak, which studies a concept of sufficiency in von Neumann algebra for minimum-variance unbiased estimation in Sec. 5 of Ref. [65]. His Theorem 5.1 states that a subalgebra with a special property called completeness is sufficient for the estimation if and only if there exists a constant GCE in terms of the Jordan product that projects onto the subalgebra. He makes no commutativity assumptions like Sinha's, but the completeness assumption is unfortunately rather restrictive, as is well known in classical statistics [5] and recognized by Łuczak himself [65]. Even in classical statistics, completeness is difficult to check, and not many models are known to satisfy it. It is unclear what quantum models beyond the known classical cases can satisfy the property. Theorem 3 here, on the other hand, does not require the unbiasedness and completeness assumptions.

Lastly, it is worth mentioning that Refs. [86, 87] by Shmaya and Chefles concern a quantum generalization of another Blackwell theorem, which, to my knowledge, has no relation to the Rao-Blackwell theorem, apart from Blackwell's name being attached to both.

I Quantum U-statistics

The goal here is to compute the GCE given by Eq. (5.14) for an operator in the form of Eq. (5.15).

A few definitions are necessary before I can proceed. Let

$$\{e(u) : u \in \mathcal{U}\} \tag{I.1}$$

be an orthonormal basis of $\mathcal{O}(\mathcal{H}_1)$ and

$$\{E(\boldsymbol{u}) \equiv e(u_1) \otimes \cdots \otimes e(u_n) : \boldsymbol{u} \in \mathcal{U}^n\} \tag{I.2}$$

be an orthonormal basis of $\mathcal{O}(\mathcal{H}_1^{\otimes n})$, where $\boldsymbol{u}$ is a column vector and the orthonormality relations are

$$\langle e(u), e(v)\rangle_{\mathrm{HS}} = \delta_{uv}, \quad \langle E(\boldsymbol{u}), E(\boldsymbol{v})\rangle_{\mathrm{HS}} = \delta_{\boldsymbol{uv}}. \tag{I.3}$$

For example, one can assume the matrix units $e(u) = |u'\rangle\langle u''|$ with $u = (u', u'')$. Any $B \in \mathcal{O}(\mathcal{H}_1^{\otimes n})$ can be expressed as

$$B = \sum_{\boldsymbol{u}} B(\boldsymbol{u})E(\boldsymbol{u}), \quad B(\boldsymbol{u}) = \langle E(\boldsymbol{u}), B\rangle_{\mathrm{HS}}. \tag{I.4}$$

Define the permutation matrix $\hat{\pi}$ on a column vector as

$$\hat{\pi}_{jk} \equiv \delta_{j\pi(k)}, \qquad (\hat{\pi}\boldsymbol{u})_j = u_{\pi^{-1}j}. \tag{I.5}$$

Then

$$U_\pi E(\boldsymbol{u})U_\pi^\dagger = E(\hat{\pi}\boldsymbol{u}), \tag{I.6}$$

and the symmetrization map given by Eq. (5.11) becomes

$$\frac{1}{n!}\sum_\pi U_\pi B U_\pi^\dagger = \frac{1}{n!}\sum_\pi \sum_{\boldsymbol{u}} B(\boldsymbol{u})E(\hat{\pi}\boldsymbol{u}) \tag{I.7}$$

$$= \sum_{\boldsymbol{u}} \tilde{B}(\boldsymbol{u})E(\boldsymbol{u}), \tag{I.8}$$

$$\tilde{B}(\boldsymbol{u}) = \frac{1}{n!}\sum_\pi B(\hat{\pi}\boldsymbol{u}), \tag{I.9}$$

which boils down to a symmetrization of $B(\boldsymbol{u})$. In general, a symmetric operator on $\mathcal{H}_1^{\otimes m}$ is defined by

$$U_{\pi_m} C U_{\pi_m}^\dagger = C, \quad C(\hat{\pi}_m\boldsymbol{v}) = C(\boldsymbol{v}) \quad \forall \pi_m \in S_m. \tag{I.10}$$

Given any operator on $\mathcal{H}_1^{\otimes m}$, a symmetric version can be obtained by applying the symmetrization map.

Define a projection matrix $\Pi_{\boldsymbol{j}} : \mathcal{U}^n \to \mathcal{U}^{\dim \boldsymbol{j}}$ by

$$\Pi_{\boldsymbol{j}}\boldsymbol{u} = \begin{pmatrix} u_{j_1} \\ \vdots \\ u_{j_m} \end{pmatrix}, \tag{I.11}$$

where $\boldsymbol{j} = (j_1, \ldots, j_m) \in \mathcal{J}_m$ is a vector of indices with $1 \leq m \leq n$ and $\mathcal{J}_m$ is the set of m-permutations of $\{1, \ldots, n\}$ (ordered sampling without replacement). Define also $\{\boldsymbol{j}\}$ for a $\boldsymbol{j} \in \mathcal{J}_m$ as the vector of indices sorted in ascending order and define the set of all such vectors as

$$\mathcal{K}_m \equiv \{\boldsymbol{k} \in \mathcal{J}_m : \boldsymbol{k} = \{\boldsymbol{k}\}\}, \tag{I.12}$$

which is equivalent to the set of m-combinations of $\{1, \ldots, n\}$ (unordered sampling without replacement).

A formula for the symmetrization can now be presented.

Proposition 3. *Suppose that $B \in \mathcal{O}(\mathcal{H}_1^{\otimes n})$ can be decomposed as*

$$B = C \otimes C', \tag{I.13}$$

where $C \in \mathcal{O}(\mathcal{H}_1^{\otimes m})$ applies to the first m Hilbert subspaces in $\mathcal{H}_1^{\otimes n}$ and $C' \in \mathcal{O}(\mathcal{H}_1^{\otimes (n-m)})$ applies to the rest. Assume that both C and C' are symmetric. Then the symmetrized B is given by

$$\frac{1}{n!}\sum_\pi U_\pi B U_\pi^\dagger = \binom{n}{m}^{-1} \sum_{\boldsymbol{k} \in \mathcal{K}_m} (C \otimes C')_{\boldsymbol{k}},$$

$$(C \otimes C')_{\boldsymbol{k}} \equiv \sum_{\boldsymbol{u}} C(\Pi_{\boldsymbol{k}} \boldsymbol{u}) C'(\Pi_{\boldsymbol{k}'} \boldsymbol{u}) E(\boldsymbol{u}), \quad (I.14)$$

where $\{E(\boldsymbol{u})\}$ is an orthonormal basis of $\mathcal{O}(\mathcal{H}_1^{\otimes n})$ given by Eq. (I.2), $C(\boldsymbol{v})$ and $C'(\boldsymbol{w})$ are the components of C and C' with respect to the same basis, the projection matrix Π is defined by Eq. (I.11), $\mathcal{K}_m$ is the m-combinations of $\{1, \ldots, n\}$, and for each $\boldsymbol{k}$, $\boldsymbol{k}'$ is defined as the rest of the indices in $\{1, \ldots, n\}$.

See Appendix J for the proof.

Each $(C \otimes C')_{\boldsymbol{k}}$ in Eqs. (I.14) is an application of C on the m Hilbert subspaces in $\mathcal{H}_1^{\otimes n}$ indexed by $\boldsymbol{k} = (k_1, \ldots, k_m)$ and an application of C' on the other $n - m$ Hilbert subspaces. If C is not symmetric, it can be symmetrized first before Proposition 3 is used. This is because the left Haar measure is also the right Haar measure for the permutation group, making the symmetrization map invariant to any prior permutation as well. One is therefore free to symmetrize C in $C \otimes C'$ first before the total symmetrization in Proposition 3. The same goes for C'. If B is in the general form of $\otimes_n C_n$, Proposition 3 can be applied recursively to produce a generalized multinomial form of Eqs. (I.14).

Proposition 3 gives the quantum U-statistics in Ref. [45] if Eq. (5.15), a special case of Eq. (I.13) with $C' = I_1^{\otimes (n-m)}$, is assumed. The classical U-statistics [5, 72] are obtained by assuming $\{e(u) = |u\rangle \langle u|\}$, where $\{|u\rangle\}$ is an orthonormal basis of $\mathcal{H}_1$, and $C'(\Pi_{\boldsymbol{k}'} \boldsymbol{u}) = 1$, such that the estimator in terms of the classical variable $\boldsymbol{u}$ becomes

$$\tilde{B}(\boldsymbol{u}) = \binom{n}{m}^{-1} \sum_{\boldsymbol{k} \in \mathcal{K}_m} C(\Pi_{\boldsymbol{k}} \boldsymbol{u}). \tag{I.15}$$

J Proofs

Proof of Lemma 1. To prove the first and last equalities in Eq. (3.4), write

$$\langle I_1, A \rangle_\sigma = \langle I_1, \mathcal{E}_\sigma A \rangle_{\mathrm{HS}} = \langle \mathcal{E}_\sigma I_1, A \rangle_{\mathrm{HS}}$$
$$= \langle \sigma, A \rangle_{\mathrm{HS}} = \mathrm{tr}\, \sigma A, \tag{J.1}$$

where the self-adjoint property of $\mathcal{E}_\sigma$ and Eq. (2.3) have been used. To prove the second equality in Eq. (3.4), plug $c = I_2$ into Eq. (2.15) and use the unital property $\mathcal{F}^* I_2 = I_1$. □

Proof of Lemma 2. Write

$$\|A - a I_1\|_\sigma^2 = \|A\|_\sigma^2 - 2\,\mathrm{Re}\left[a^* \langle I_1, A \rangle_\sigma\right] + |a|^2, \tag{J.2}$$

$$\|\mathcal{F}_* A - a I_2\|_{\mathcal{F}_\sigma}^2 = \|\mathcal{F}_* A\|_{\mathcal{F}_\sigma}^2 - 2\,\mathrm{Re}\left[a^* \langle I_2, \mathcal{F}_* A \rangle_{\mathcal{F}_\sigma}\right] + |a|^2. \tag{J.3}$$

Lemma 1 gives $\langle I_1, A \rangle_\sigma = \langle I_2, \mathcal{F}_* A \rangle_{\mathcal{F}_\sigma}$. Then Eq. (J.2) minus Eq. (J.3) gives Eq. (3.5) via Eq. (2.17). □

Proof of Lemma 3. Equation (2.15) gives, for any $c \in L_2(\sigma_x)$,

$$\langle c, \mathcal{G}_* B \rangle_{\sigma_x} = \langle \mathcal{G}^* c, B \rangle_{\rho_x} \tag{J.4}$$

$$= \mathrm{tr}\left[c^\dagger \mathrm{tr}_0\left(\mathcal{E}_{\sigma_x \otimes \tau} B\right)\right] \tag{J.5}$$

$$= \mathrm{tr}\left[(c \otimes I_0)^\dagger \mathcal{E}_{\sigma_x \otimes \tau} B\right] \tag{J.6}$$

$$= \langle c \otimes I_0, \mathcal{E}_{\sigma_x \otimes \tau} B \rangle_{\mathrm{HS}} \tag{J.7}$$

$$= \langle \mathcal{E}_{\sigma_x \otimes \tau}(c \otimes I_0), B \rangle_{\mathrm{HS}} \tag{J.8}$$

$$= \langle (\mathcal{E}_{\sigma_x} c) \otimes (\mathcal{E}_\tau I_0), B \rangle_{\mathrm{HS}} \tag{J.9}$$

$$= \langle (\mathcal{E}_{\sigma_x} c) \otimes \tau, B \rangle_{\mathrm{HS}} \tag{J.10}$$

$$= \mathrm{tr}\left\{[(\mathcal{E}_{\sigma_x} c)^\dagger \otimes \tau] B\right\} \tag{J.11}$$

$$= \mathrm{tr}\left\{(\mathcal{E}_{\sigma_x} c)^\dagger \mathrm{tr}_0[(I_1 \otimes \tau) B]\right\} \tag{J.12}$$

$$= \langle \mathcal{E}_{\sigma_x} c, \mathrm{tr}_0[(I_1 \otimes \tau) B] \rangle_{\mathrm{HS}} \tag{J.13}$$

$$= \langle c, \mathrm{tr}_0[(I_1 \otimes \tau) B] \rangle_{\sigma_x}, \tag{J.14}$$

where the self-adjoint property of $\mathcal{E}$ and Eqs. (2.3) and (2.5) have been used at various steps. Equation (J.14) means that $\mathrm{tr}_0[(I_1 \otimes \tau) B]$ is a solution to the GCE $\mathcal{G}_* B$. □

Proof of Corollary 2. Corollary 3 states that, given the local error MSE_x for any POVM M and any estimator b, there exists an operator-valued estimator on $\mathcal{H}_2$ with an error MSE'_x that satisfies $\mathrm{MSE}_x \geq \mathrm{MSE}'_x$ for all $x \in \mathcal{X}$. The average error of (M, b) is then also bounded as

$$\sum_x P_X(x) \mathrm{MSE}_x \geq \sum_x P_X(x) \mathrm{MSE}'_x \geq D_{\sigma, \mathcal{F}}(A, \mathcal{F}_* A), \tag{J.15}$$

where the last inequality follows from the optimality of the Personick estimator $\mathcal{F}_* A$ among all observables on $\mathcal{H}_2$, as per Definition 2.

□

Proof of Lemma 4.

$$\mathcal{G} \mathcal{E}_{\rho_x} B = \int \mu(dz) U_z \left(\mathcal{E}_{\rho_x} B\right) U_z^\dagger \tag{J.16}$$

$$= \int \mu(dz) \mathcal{E}_{U_z \rho_x U_z^\dagger}(U_z B U_z^\dagger) \tag{J.17}$$

$$= \mathcal{E}_{\rho_x} \int \mu(dz) U_z B U_z^\dagger, \tag{J.18}$$

where Eqs. (2.4) and (5.8) have been used. The interchange of $\mathcal{E}_{\rho_x}$ and the Bochner integral is valid because $\mathcal{E}_{\rho_x}$ is a linear map on a finite-dimensional operator space (more assumptions would be needed for infinite-dimensional operator spaces; see Corollary 2 on p. 134 in Ref. [88]). By Eq. (2.16), Eq. (J.18) is equal to $\mathcal{E}_{\mathcal{G}\rho_x}\mathcal{G}_*B = \mathcal{E}_{\rho_x}\mathcal{G}_*B$, resulting in a solution to the GCE given by Eq. (5.10). $\qquad\square$

Proof of Lemma 5. Assuming Eq. (5.22) and using Eq. (5.20), one obtains

$$\mathcal{E}_{\rho_x}\mathcal{G}_*B = \bigoplus_n \mathcal{E}_{\sigma_x^{(n)}}\left(\Pi_n B \Pi_n\right), \qquad \text{(J.19)}$$

which is equal to

$$\mathcal{G}\mathcal{E}_{\rho_x}B = \bigoplus_n \Pi_n\left(\mathcal{E}_{\rho_x}B\right)\Pi_n = \bigoplus_n \mathcal{E}_{\sigma_x^{(n)}}\left(\Pi_n B \Pi_n\right), \qquad \text{(J.20)}$$

by virtue of Eq. (5.21). It follows that Eq. (5.22) is a solution to the GCE, as per Eq. (2.16). $\qquad\square$

Note that Lemmas 1–5 apply to classes of GCEs and not just the Jordan version. Note also that the GCEs for any sequence of the channels can be computed by chaining the individual GCEs in a manner reminiscent of calculus.

Proof of Proposition 1. The GCE defined by Eq. (2.16) can be solved by the operator Fourier transform

$$\operatorname{tr}\left(\mathcal{E}_{\mathcal{F}\sigma}\mathcal{F}_*A\right)\tilde{W}(\zeta) = \operatorname{tr}\left(\mathcal{F}\mathcal{E}_\sigma A\right)\tilde{W}(\zeta). \qquad \text{(J.21)}$$

The right-hand side can be expressed as

$$\operatorname{tr}\left(\mathcal{F}\mathcal{E}_\sigma A\right)\tilde{W}(\zeta) = \operatorname{tr}(\mathcal{E}_\sigma A)\mathcal{F}^*\tilde{W}(\zeta) \qquad \text{(J.22)}$$

$$= f(\zeta)\operatorname{tr}(\mathcal{E}_\sigma A)W(F^\top\zeta) \qquad \text{(J.23)}$$

$$= f(\zeta)\left[-iu^\top\nabla\phi(z)\right]_{z=F^\top\zeta} \qquad \text{(J.24)}$$

$$= u^\top\left(m + i\Sigma F^\top\zeta\right)\tilde{\phi}(\zeta), \qquad \text{(J.25)}$$

where Eq. (J.23) has used the Gaussian-channel definition given by Eq. (C.5), Eq. (J.24) has used Eq. (5.4.43) in Ref. [46] with $\nabla \equiv \begin{pmatrix} \partial/\partial x_1 & \partial/\partial y_1 & \dots & \partial/\partial x_s & \partial/\partial y_s \end{pmatrix}^\top$, and Eq. (J.25) has used the Gaussian $\phi(z)$ given by Eq. (C.4) and the output $\tilde{\phi}(\zeta)$ given by Eq. (C.7). With similar steps and the ansatz

$$\mathcal{F}_*A = v^\top\tilde{Q} + c, \qquad v \in \mathbb{R}^{2t}, \qquad c \in \mathbb{R}, \qquad \text{(J.26)}$$

the left-hand side of Eq. (J.21) can be expressed as

$$\operatorname{tr}\left(\mathcal{E}_{\mathcal{F}\sigma}\mathcal{F}_*A\right)\tilde{W}(\zeta) = \left[v^\top(\tilde{m} + i\tilde{\Sigma}\zeta) + c\right]\tilde{\phi}(\zeta), \qquad \text{(J.27)}$$

where Eqs. (C.7) and (C.8) are assumed. Equating Eq. (J.25) with Eq. (J.27) leads to

$$v^\top = u^\top\Sigma F^\top\tilde{\Sigma}^{-1}, \qquad \text{(J.28)}$$

$$c = u^\top m - v^\top\tilde{m}. \qquad \text{(J.29)}$$

Equations (J.28) and (J.29) can then be substituted into Eq. (J.26) to give Eqs. (C.12) and (C.13) via Eqs. (C.9) and (C.10).

To derive Eq. (C.14), use Lemmas 1 and 2 to write

$$a = \operatorname{tr}\sigma A = \operatorname{tr}(\mathcal{F}\sigma)(\mathcal{F}_*A), \qquad \text{(J.30)}$$

$$D_{\sigma,\mathcal{F}}(A, \mathcal{F}_*A) = \|A - aI_1\|_\sigma^2 - \|\mathcal{F}_*A - aI_2\|_{\mathcal{F}\sigma}^2 \qquad \text{(J.31)}$$

$$= u^\top\Sigma u - v^\top\tilde{\Sigma}v, \qquad \text{(J.32)}$$

where the last step has used the fact that A and $\mathcal{F}_*A$ are both quadrature operators and their variances are determined by the covariance matrices of the Gaussian states. Substituting Eqs. (J.28) and (C.10) into Eq. (J.32) leads to Eq. (C.14). $\qquad\square$

Proof of Corollary 7. Let c be any operator on $\mathcal{H}_3$. By the definition of $\mathcal{G}_*^{\rho_x}B$ given by Eq. (2.15),

$$\langle c, \mathcal{G}_*^{\rho_x}B\rangle_{\mathcal{G}\rho_x} = \langle\mathcal{G}^*c, B\rangle_{\rho_x} \qquad \forall x \in \mathcal{X}. \qquad \text{(J.33)}$$

Taking the expectation of this equation with respect to $P_X(x)$, one obtains

$$\sum_x P_X(x)\langle c, \mathcal{G}_*^{\rho_x}B\rangle_{\mathcal{G}\rho_x} = \sum_x P_X(x)\langle\mathcal{G}^*c, B\rangle_{\rho_x}, \qquad \text{(J.34)}$$

$$\langle c, \mathcal{G}_*^{\rho_x}B\rangle_{\mathcal{G}\mathcal{F}\sigma} = \langle\mathcal{G}^*c, B\rangle_{\mathcal{F}\sigma}, \qquad \text{(J.35)}$$

where Eq. (J.35) has used the facts that c, B, $\mathcal{G}$, and $\mathcal{G}_*^{\rho_x}B$ all do not depend on x, the trace and $\mathcal{G}$ are linear, the Jordan product is bilinear, and $\sum_x P_X(x)\rho_x = \mathcal{F}\sigma$. Equation (J.35) means that $\mathcal{G}_*^{\rho_x}B$ satisfies the definition of the final Personick estimator $\mathcal{G}_*^{\mathcal{F}\sigma}B$ as per Eq. (2.15).

Equation (G.2) can be proved by combining the monotonicity of the Bayesian error (Corollary 1) and the quantum Rao-Blackwell theorem (Theorem 3).

Equation (G.3) can be proved by contradiction: assume that there exists a $x \in \mathcal{X}$ with $P_X(x) > 0$ such that $\mathrm{MSE}_x > \mathrm{MSE}'_x$. Since $\mathrm{MSE}_x \geq \mathrm{MSE}'_x$ by Theorem 3, the assumption would imply $\sum_x P_X(x)\mathrm{MSE}_x > \sum_x P_X(x)\mathrm{MSE}'_x$, which contradicts Eq. (G.2). It follows that the assumption cannot hold and one must have Eq. (G.3). $\qquad\square$

Proof of Proposition 2. Given Eq. (H.5), a solution to any GCE is

$$\mathcal{G}_*B = \sum_s \frac{\operatorname{tr}\Pi_s\mathcal{E}_{\rho_x}B\Pi_s}{\operatorname{tr}\Pi_s\rho_x\Pi_s}\Pi_s. \qquad \text{(J.36)}$$

The numerator can be expressed as

$$\operatorname{tr}\Pi_s\mathcal{E}_{\rho_x}B\Pi_s = \langle\Pi_s, \mathcal{E}_{\rho_x}B\rangle_{\mathrm{HS}} = \langle\mathcal{E}_{\rho_x}\Pi_s, B\rangle_{\mathrm{HS}} \qquad \text{(J.37)}$$

$$= \langle\rho_x\Pi_s, B\rangle_{\mathrm{HS}} = \varphi(s, x)\operatorname{tr}\sigma_s B_s, \qquad \text{(J.38)}$$

where the self-adjoint property of $\mathcal{E}_{\rho_x}$, the commutativity between ρ_x and Π_s, and Eqs. (2.3), (H.2), and

(H.4) have been used at various steps. Similarly, the denominator in Eq. (J.36) can be expressed as

$$\operatorname{tr} \Pi_s \rho_x \Pi_s = \varphi(s, x) \operatorname{tr} \sigma_s. \tag{J.39}$$

Equation (H.6) then follows. $\qquad\square$

Proof of Proposition 3. With Eq. (I.13), $B(\boldsymbol{u})$ becomes

$$B(\boldsymbol{u}) = C(\Pi_{[1,m]}\boldsymbol{u})C'(\Pi_{[m+1,n]}\boldsymbol{u}), \tag{J.40}$$
$$[1, m] \equiv (1, \ldots, m), \tag{J.41}$$
$$[m + 1, n] \equiv (m + 1, \ldots, n). \tag{J.42}$$

With the identity

$$\Pi_{\boldsymbol{j}}\hat{\pi}\boldsymbol{u} = \Pi_{\pi^{-1}\boldsymbol{j}}\boldsymbol{u}, \tag{J.43}$$

$B(\hat{\pi}\boldsymbol{u})$ in Eq. (I.9) becomes

$$B(\hat{\pi}\boldsymbol{u}) = C(\Pi_{[1,m]}\hat{\pi}\boldsymbol{u})C'(\Pi_{[m+1,n]}\hat{\pi}\boldsymbol{u}) \tag{J.44}$$
$$= C(\Pi_{\pi^{-1}[1,m]}\boldsymbol{u})C'(\Pi_{\pi^{-1}[m+1,n]}\boldsymbol{u}). \tag{J.45}$$

The symmetry of C and C' implies

$$C(\Pi_{\boldsymbol{j}}\boldsymbol{u}) = C(\Pi_{\{\boldsymbol{j}\}}\boldsymbol{u}), \quad \forall \boldsymbol{j} \in \mathcal{J}_m, \tag{J.46}$$
$$C'(\Pi_{\boldsymbol{j}}\boldsymbol{u}) = C'(\Pi_{\{\boldsymbol{j}\}}\boldsymbol{u}), \quad \forall \boldsymbol{j} \in \mathcal{J}_{n-m}, \tag{J.47}$$
$$B(\hat{\pi}\boldsymbol{u}) = C(\Pi_{\{\pi^{-1}[1,m]\}}\boldsymbol{u})C'(\Pi_{\{\pi^{-1}[m+1,n]\}}\boldsymbol{u}). \tag{J.48}$$

The $n!$ summands in Eq. (I.9) with respect to π can now be divided into subsets indexed by Eq. (I.12). Each subset, indexed by a $\boldsymbol{k} \in \mathcal{K}_m$, contains $m!(n - m)!$ terms all equal to $C(\Pi_{\boldsymbol{k}}\boldsymbol{u})C'(\Pi_{\boldsymbol{k}'}\boldsymbol{u})$ with

$$\boldsymbol{k} = \left\{\pi^{-1}[1, m]\right\}, \quad \boldsymbol{k}' = \left\{\pi^{-1}[m + 1, n]\right\}. \tag{J.49}$$

The sum in Eq. (I.9) becomes

$$\frac{1}{n!}\sum_\pi B(\hat{\pi}\boldsymbol{u}) = \binom{n}{m}^{-1}\sum_{\boldsymbol{k} \in \mathcal{K}_m} C(\Pi_{\boldsymbol{k}}\boldsymbol{u})C'(\Pi_{\boldsymbol{k}'}\boldsymbol{u}), \tag{J.50}$$

where $\binom{n}{m} \equiv n!/m!(n - m)! = |\mathcal{K}_m|$ is the binomial coefficient. The proposition hence follows. $\qquad\square$

References

[1] K. R. Parthasarathy. *Introduction to Probability and Measure.* Hindustan Book Agency, New Delhi, 2005. ISBN 9789386279279. DOI: 10.1007/978-93-86279-27-9. URL https://www.springer.com/gp/book/9789386279279.

[2] James O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York, 1985. DOI: 10.1007/978-1-4757-4286-2.

[3] C. Radhakrishna Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

[4] David Blackwell. Conditional expectation and unbiased sequential estimation. *Annals of Mathematical Statistics*, 18(1):105–110, March 1947. ISSN 0003-4851. DOI: 10.1214/aoms/1177730497.

[5] E. L. Lehmann and George Casella. *Theory of Point Estimation.* Springer, New York, second edition, 1998. ISBN 978-0-387-22728-3. DOI: 10.1007/b98854.

[6] Steven K. Thompson. *Sampling.* John Wiley & Sons, Ltd., Chichester, England, UK, February 2012. ISBN 978-0-47040231-3. DOI: 10.1002/9781118162934.

[7] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods.* Springer, New York, NY, USA, 2004. ISBN 978-1-4757-4145-2. DOI: 10.1007/978-1-4757-4145-2.

[8] Hisaharu Umegaki. Conditional expectation in an operator algebra. *Tohoku Mathematical Journal, Second Series*, 6(2-3):177–181, 1954. ISSN 0040-8735. DOI: 10.2748/tmj/1178245177.

[9] S. Personick. Application of quantum estimation theory to analog communication over quantum channels. *IEEE Transactions on Information Theory*, 17(3):240–246, May 1971. ISSN 0018-9448. DOI: 10.1109/TIT.1971.1054643.

[10] V. P. Belavkin and B. A. Grishanin. Optimum estimation in quantum channels by the generalized Heisenberg inequality method. *Probl. Peredachi Inf.*, 9:44–52, 1973. URL http://mi.mathnet.ru/eng/ppi907.

[11] Luigi Accardi and Carlo Cecchini. Conditional expectations in von Neumann algebras and a theorem of Takesaki. *Journal of Functional Analysis*, 45(2):245–273, February 1982. ISSN 0022-1236. DOI: 10.1016/0022-1236(82)90022-2. URL https://www.sciencedirect.com/science/article/pii/0022123682900222.

[12] Dénes Petz. *Quantum Information Theory and Quantum Statistics.* Springer, Berlin, Germany, 2010. ISBN 978-3-540-74636-2. DOI: 10.1007/978-3-540-74636-2. URL https://link.springer.com/book/10.1007/978-3-540-74636-2.

[13] Yakir Aharonov, David Z. Albert, and Lev Vaidman. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Physical Review Letters*, 60:1351–1354, Apr 1988. DOI: 10.1103/PhysRevLett.60.1351. URL http://link.aps.org/doi/10.1103/PhysRevLett.60.1351.

[14] Justin Dressel. Weak values as interference phenomena. *Physical Review A*, 91(3):032116, March 2015. DOI: 10.1103/PhysRevA.91.032116. URL https://link.aps.org/doi/10.1103/PhysRevA.91.032116.

[15] Masahito Hayashi. *Quantum Information Theory: Mathematical Foundation.* Springer, Berlin,

Accepted in ⟨ ⟩uantum 2023-10-26, click title to verify. Published under CC-BY 4.0.

21

second edition, 2017. DOI: 10.1007/978-3-662-49725-8.

[16] Kentaro Ohki. A smoothing theory for open quantum systems: The least mean square approach. In *2015 54th IEEE Conference on Decision and Control (CDC)*, page 4350–4355, Osaka, Japan, December 2015. IEEE. DOI: 10.1109/CDC.2015.7402898.

[17] Kentaro Ohki. An invitation to quantum filtering and smoothing theory based on two inner products. In *2018 Building Foundations for Quantum Statistical Modeling*, volume 2018, page 18–44. Kyoto University, Kyoto, 2018. URL http://hdl.handle.net/2433/231710.

[18] Mankei Tsang. Generalized conditional expectations for quantum retrodiction and smoothing. *Physical Review A*, 105(4):042213, April 2022. ISSN 2469-9934. DOI: 10.1103/PhysRevA.105.042213.

[19] Alexander S. Holevo. *Statistical Structure of Quantum Theory*. Springer-Verlag, Berlin, 2001. DOI: 10.1007/3-540-44998-1.

[20] Viacheslav P. Belavkin. Nondemolition principle of quantum measurement theory. *Foundations of Physics*, 24(5):685–714, 1994. ISSN 0015-9018. DOI: 10.1007/BF02054669. URL http://dx.doi.org/10.1007/BF02054669.

[21] L. Bouten, R. Van Handel, and M. James. An introduction to quantum filtering. *SIAM Journal on Control and Optimization*, 46(6):2199–2241, 2007. DOI: 10.1137/060651239. URL http://epubs.siam.org/doi/abs/10.1137/060651239.

[22] Howard M. Wiseman and Gerard J. Milburn. *Quantum Measurement and Control*. Cambridge University Press, Cambridge, 2010. DOI: 10.1017/CBO9780511813948.

[23] Stephen M. Barnett, David T. Pegg, and John Jeffers. Bayes' theorem and quantum retrodiction. *Journal of Modern Optics*, 47 (11):1779–1789, September 2000. ISSN 0950-0340. DOI: 10.1080/09500340008232431. URL https://www.tandfonline.com/doi/abs/10.1080/09500340008232431.

[24] Stephen M. Barnett, John Jeffers, and David T. Pegg. Quantum Retrodiction: Foundations and Controversies. *Symmetry*, 13(4):586, April 2021. DOI: 10.3390/sym13040586. URL https://www.mdpi.com/2073-8994/13/4/586.

[25] Masahiro Yanagisawa. Quantum smoothing. *arXiv:0711.3885 [quant-ph]*, December 2007. DOI: 10.48550/arXiv.0711.3885. URL http://arxiv.org/abs/0711.3885.

[26] Mankei Tsang. Time-symmetric quantum theory of smoothing. *Physical Review Letters*, 102:250403, Jun 2009. DOI: 10.1103/PhysRevLett.102.250403. URL http://link.aps.org/doi/10.1103/PhysRevLett.102.250403.

[27] Mankei Tsang. Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. *Physical Review A*, 80:033840, Sep 2009. DOI: 10.1103/PhysRevA.80.033840. URL http://link.aps.org/doi/10.1103/PhysRevA.80.033840.

[28] Mankei Tsang. Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. II. Applications to atomic magnetometry and Hardy's paradox. *Physical Review A*, 81:013824, Jan 2010. DOI: 10.1103/PhysRevA.81.013824. URL http://link.aps.org/doi/10.1103/PhysRevA.81.013824.

[29] Ivonne Guevara and Howard Wiseman. Quantum state smoothing. *Physical Review Letters*, 115:180407, Oct 2015. DOI: 10.1103/PhysRevLett.115.180407. URL http://link.aps.org/doi/10.1103/PhysRevLett.115.180407.

[30] Areeya Chantasri, Ivonne Guevara, Kiarn T. Laverick, and Howard M. Wiseman. Unifying theory of quantum state estimation using past and future information. *Physics Reports*, 930:1–40, October 2021. ISSN 0370-1573. DOI: 10.1016/j.physrep.2021.07.003. URL https://www.sciencedirect.com/science/article/pii/S0370157321002787.

[31] M. S. Leifer and Robert W. Spekkens. Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference. *Physical Review A*, 88(5):052130, November 2013. DOI: 10.1103/PhysRevA.88.052130. URL https://link.aps.org/doi/10.1103/PhysRevA.88.052130.

[32] Dominic Horsman, Chris Heunen, Matthew F. Pusey, Jonathan Barrett, and Robert W. Spekkens. Can a quantum state over time resemble a quantum state at a single time? *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2205): 20170395, 2017. DOI: 10.1098/rspa.2017.0395. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2017.0395.

[33] Arthur J. Parzygnat and James Fullwood. From time-reversal symmetry to quantum bayes' rules. *PRX Quantum*, 4(2):020334, June 2023. ISSN 2691-3399. DOI: 10.1103/PRXQuantum.4.020334.

[34] Christopher Ferrie and Joshua Combes. Weak value amplification is suboptimal for estimation and detection. *Physical Review Letters*, 112:040406, Jan 2014. DOI: 10.1103/PhysRevLett.112.040406. URL http://link.aps.org/doi/10.1103/PhysRevLett.112.040406.

[35] Christopher Ferrie and Joshua Combes. How the result of a single coin toss can turn out to be 100 heads. *Physical Review Letters*, 113:120404, Sep 2014. DOI: 10.1103/Phys-

RevLett.113.120404. URL http://link.aps.org/doi/10.1103/PhysRevLett.113.120404.

[36] Lev Vaidman. Comment on "how the result of a single coin toss can turn out to be 100 heads". *arXiv:1409.5396*, September 2014. DOI: 10.48550/arXiv.1409.5386.

[37] Andrew N. Jordan, Julián Martínez-Rincón, and John C. Howell. Technical advantages for weak-value amplification: When less is more. *Physical Review X*, 4(1):011031, March 2014. ISSN 2160-3308. DOI: 10.1103/PhysRevX.4.011031.

[38] John E. Gough. How to estimate past measurement interventions on a quantum system undergoing continuous monitoring. *arXiv:1904.06364v1 [quant-ph]*, April 2019. DOI: 10.48550/arXiv.1904.06364. URL http://arxiv.org/abs/1904.06364v1.

[39] John E. Gough. How to Estimate Past Quantum Measurement Interventions After Continuous Monitoring. *Russian Journal of Mathematical Physics*, 27(2):218–227, April 2020. ISSN 1555-6638. DOI: 10.1134/S1061920820020089. URL https://doi.org/10.1134/S1061920820020089.

[40] M. R. James. Optimal quantum control theory. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):343–367, May 2021. ISSN 2573-5144. DOI: 10.1146/annurev-control-061520-010444.

[41] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*, volume I. Athena Scientific, Belmont, 3rd edition, 2005.

[42] Mankei Tsang, Howard M. Wiseman, and Carlton M. Caves. Fundamental Quantum Limit to Waveform Estimation. *Physical Review Letters*, 106(9):090401, March 2011. DOI: 10.1103/PhysRevLett.106.090401. URL https://link.aps.org/doi/10.1103/PhysRevLett.106.090401.

[43] Katarzyna Macieszczak, Martin Fraas, and Rafał Demkowicz-Dobrzański. Bayesian quantum frequency estimation in presence of collective dephasing. *New Journal of Physics*, 16(11):113002, October 2014. ISSN 1367-2630. DOI: 10.1088/1367-2630/16/11/113002. URL https://doi.org/10.1088/1367-2630/16/11/113002.

[44] Jesús Rubio and Jacob Dunningham. Quantum metrology in the presence of limited data. *New Journal of Physics*, 21(4):043037, April 2019. ISSN 1367-2630. DOI: 10.1088/1367-2630/ab098b. URL https://doi.org/10.1088%2F1367-2630%2Fab098b.

[45] Mădălin Guţă and Cristina Butucea. Quantum u-statistics. *Journal of Mathematical Physics*, 51(10), October 2010. ISSN 0022-2488. DOI: 10.1063/1.3476776.

[46] Alexander S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. Scuola Normale Su-periore Pisa, Pisa, Italy, 2011. DOI: 10.1007/978-88-7642-378-9.

[47] Lokenath Debnath and Piotr Mikusiński. *Introduction to Hilbert Spaces with Applications*. Elsevier, Amsterdam, 2005.

[48] Daniel Gottesman. The Heisenberg representation of quantum computers. In S. P. Corney, R. Delbourgo, and P. D. Jarvis, editors, *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, page 32–43, Cambridge, MA, 1999. International Press. URL https://arxiv.org/quant-ph/9807006.

[49] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, Cambridge, England, UK, February 2017. ISBN 978-1-10717616-4. DOI: 10.1017/9781316809976.

[50] Alexander S. Holevo. *Quantum Systems, Channels, Information*. De Gruyter, Berlin, 2nd edition, July 2019. ISBN 978-3-11064249-0. DOI: 10.1515/9783110642490.

[51] Brian D. O. Anderson and John B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, 1979.

[52] Arthur J. Parzygnat and Francesco Buscemi. Axioms for retrodiction: achieving time-reversal symmetry with a prior. *Quantum*, 7:1013, May 2023. DOI: 10.22331/q-2023-05-23-1013.

[53] Carl W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, New York, 1976. URL http://www.sciencedirect.com/science/bookseries/00765392/123.

[54] Tsachy Weissman. The relationship between causal and noncausal mismatched estimation in continuous-time AWGN channels. *IEEE Transactions on Information Theory*, 56(9):4256–4273, August 2010. ISSN 1557-9654. DOI: 10.1109/TIT.2010.2054430.

[55] Yihong Wu and Sergio Verdu. Functional properties of minimum mean-square error and mutual information. *IEEE Transactions on Information Theory*, 58(3):1289–1301, November 2011. ISSN 1557-9654. DOI: 10.1109/TIT.2011.2174959.

[56] Mankei Tsang. Mismatched quantum filtering and entropic information. In *2014 IEEE International Symposium on Information Theory (ISIT)*, page 321–325, June 2014. DOI: 10.1109/ISIT.2014.6874847.

[57] Aolin Xu and Maxim Raginsky. Minimum excess risk in Bayesian learning. *IEEE Transactions on Information Theory*, 68(12):7935–7955, May 2022. ISSN 1557-9654. DOI: 10.1109/TIT.2022.3176056.

[58] Wim Bogaerts, Daniel Pérez, José Capmany, David A. B. Miller, Joyce Poon, Dirk Englund, Francesco Morichetti, and Andrea Melloni. Programmable photonic circuits. *Nature*, 586(7828):207–216, October 2020. ISSN 1476-4687. DOI:

10.1038/s41586-020-2764-0. URL https://www.nature.com/articles/s41586-020-2764-0.

[59] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, Cambridge, 2009.

[60] Dimitri P. Bertsekas. *Reinforcement Learning and Optimal Control*. Athena Scientific, Belmont, 2019.

[61] H. M. Wiseman. Weak values, quantum trajectories, and the cavity-QED experiment on wave-particle correlation. *Physical Review A*, 65(3): 032111, February 2002. DOI: 10.1103/PhysRevA.65.032111. URL https://link.aps.org/doi/10.1103/PhysRevA.65.032111.

[62] Michael J. W. Hall. Prior information: How to circumvent the standard joint-measurement uncertainty relation. *Physical Review A*, 69:052113, May 2004. DOI: 10.1103/PhysRevA.69.052113. URL http://link.aps.org/doi/10.1103/PhysRevA.69.052113.

[63] Alberto Barchielli and Giancarlo Lupieri. Instruments and channels in quantum information. *Banach Center Publications*, 73:65–80, 2006. DOI: 10.4064/bc73-0-4. https://arxiv.org/abs/quant-ph/0412116.

[64] Kalyan B. Sinha. Sufficient statistic and Rao–Blackwell theorem in quantum probability. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 25:2240005, October 2022. ISSN 0219-0257. DOI: 10.1142/S0219025722400057.

[65] Andrzej Łuczak. On a general concept of sufficiency in von Neumann algebras. *Probability and Mathematical Statistics*, 35(2):313–324, 2015. URL https://www.math.uni.wroc.pl/~pms/files/35.2/Article/35.2.9.pdf.

[66] Aglaé N. Kellerer and Erez N. Ribak. Beyond the diffraction limit via optical amplification. *Optics Letters*, 41(14):3181–3184, July 2016. ISSN 1539-4794. DOI: 10.1364/OL.41.003181. URL http://www.osapublishing.org/ol/abstract.cfm?uri=ol-41-14-3181.

[67] Hermann A. Haus. *Electromagnetic Noise and Quantum Optical Measurements*. Springer-Verlag, Berlin, 2000. DOI: 10.1007/978-3-662-04190-1.

[68] Carlton M. Caves. Quantum limits on noise in linear amplifiers. *Physical Review D*, 26(8): 1817–1839, October 1982. DOI: 10.1103/PhysRevD.26.1817. URL https://link.aps.org/doi/10.1103/PhysRevD.26.1817.

[69] W. M. Itano, J. C. Bergquist, J. J. Bollinger, J. M. Gilligan, D. J. Heinzen, F. L. Moore, M. G. Raizen, and D. J. Wineland. Quantum projection noise: Population fluctuations in two-level systems. *Physical Review A*, 47(5):3554–3570, May 1993. ISSN 2469-9934. DOI: 10.1103/PhysRevA.47.3554.

[70] Vladimir B. Braginsky and Farid Ya. Khalili. *Quantum Measurement*. Cambridge University Press, Cambridge, 1992.

[71] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, Cambridge, 2018. ISBN 9781107180567. DOI: 10.1017/9781316848142. URL https://www.cambridge.org/core/books/theory-of-quantum-information/AE4AA5638F808D2CFEB070C55431D897.

[72] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK, 1998. DOI: 10.1017/CBO9780511802256.

[73] Leonard Mandel and Emil Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, Cambridge, 1995. DOI: 10.1017/CBO9781139644105.

[74] Mankei Tsang. Quantum nonlocality in weak-thermal-light interferometry. *Physical Review Letters*, 107:270402, Dec 2011. DOI: 10.1103/PhysRevLett.107.270402. URL http://link.aps.org/doi/10.1103/PhysRevLett.107.270402.

[75] Ranjith Nair and Mankei Tsang. Quantum optimality of photon counting for temperature measurement of thermal astronomical sources. *The Astrophysical Journal*, 808(2):125, 2015. DOI: 10.1088/0004-637X/808/2/125. URL http://stacks.iop.org/0004-637X/808/i=2/a=125.

[76] Mankei Tsang, Ranjith Nair, and Xiao-Ming Lu. Quantum theory of superresolution for two incoherent optical point sources. *Physical Review X*, 6:031033, Aug 2016. DOI: 10.1103/PhysRevX.6.031033. URL http://link.aps.org/doi/10.1103/PhysRevX.6.031033.

[77] Mankei Tsang. Resolving starlight: a quantum perspective. *Contemporary Physics*, 60(4): 279–298, October 2019. ISSN 0010-7514. DOI: 10.1080/00107514.2020.1736375. URL https://doi.org/10.1080/00107514.2020.1736375.

[78] Ranjith Nair and Mankei Tsang. Far-Field Superresolution of Thermal Electromagnetic Sources at the Quantum Limit. *Physical Review Letters*, 117(19):190801, November 2016. DOI: 10.1103/PhysRevLett.117.190801. URL http://link.aps.org/doi/10.1103/PhysRevLett.117.190801.

[79] Cosmo Lupo and Stefano Pirandola. Ultimate Precision Bound of Quantum and Subwavelength Imaging. *Physical Review Letters*, 117(19): 190802, November 2016. DOI: 10.1103/PhysRevLett.117.190802. URL http://link.aps.org/doi/10.1103/PhysRevLett.117.190802.

[80] Shilin Ng, Shan Zheng Ang, Trevor A. Wheatley, Hidehiro Yonezawa, Akira Furusawa, Elanor H. Huntington, and Mankei Tsang. Spectrum analysis with quantum dynamical systems. *Physical*

Accepted in 〈 〉uantum 2023-10-26, click title to verify. Published under CC-BY 4.0.

24

Review A, 93(4):042121, April 2016. ISSN 2469-9934. DOI: 10.1103/PhysRevA.93.042121.

[81] Wojciech Górecki, Alberto Riccardi, and Lorenzo Maccone. Quantum metrology of noisy spreading channels. *Physical Review Letters*, 129(24): 240503, December 2022. ISSN 1079-7114. DOI: 10.1103/PhysRevLett.129.240503.

[82] Haowei Shi and Quntao Zhuang. Ultimate precision limit of noise sensing and dark matter search. *npj Quantum Information*, 9(27): 1–10, March 2023. ISSN 2056-6387. DOI: 10.1038/s41534-023-00693-w.

[83] Mankei Tsang. Quantum noise spectroscopy as an incoherent imaging problem. *Physical Review A*, 107(1):012611, January 2023. ISSN 2469-9934. DOI: 10.1103/PhysRevA.107.012611.

[84] John M. Lee. *Introduction to Smooth Manifolds*. Springer-Verlag, New York, 2003. ISBN 9780387217529. DOI: 10.1007/978-0-387-21752-9. URL https://www.springer.com/gp/book/9780387217529.

[85] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics. I: Functional Analysis*. Academic Press, San Diego, 1980.

[86] Eran Shmaya. Comparison of information structures and completely positive maps. *Journal of Physics A: Mathematical and General*, 38(44): 9717, October 2005. ISSN 0305-4470. DOI: 10.1088/0305-4470/38/44/008.

[87] Anthony Chefles. The quantum Blackwell theorem and minimum error state discrimination. *arXiv:0907.0866*, July 2009. DOI: 10.48550/arXiv.0907.0866.

[88] Kôsaku Yosida. *Functional Analysis*. Springer, Berlin, 6th edition, 1995. ISBN 978-3-642-61859-8. DOI: 10.1007/978-3-642-61859-8.