Generalized conditional expectations for quantum retrodiction and smoothing

Mankei Tsang [*]

Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583 and Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551

(Received 20 December 2021; revised 26 February 2022; accepted 6 April 2022; published 21 April 2022)

The inference of a hidden variable's historical value, based on observations before and after the fact, is a controversial subject in quantum mechanics. Here I address the controversy by proposing a formalism that unifies and generalizes some of the previous proposals for the task, including the quantum minimum-mean-square-error estimators proposed by Ohki, the generalized conditional expectation proposed by Accardi and Cecchini, the quantum smoothing theory proposed by Tsang, the optimal observables for parameter estimation proposed by Personick, Belavkin, and Grishanin, and the weak values proposed by Aharonov, Albert, and Vaidman. The formalism is based on Ohki's suggestion of a distance between two observables in the Heisenberg picture, which remains well defined for incompatible observables and serves as a more general foundation for quantum inference than Belavkin's nondemolition principle.

DOI: 10.1103/PhysRevA.105.042213

I. INTRODUCTION

The inference of a hidden variable's historical value, based on observations before and after the fact, is a fascinating yet controversial subject in quantum mechanics; see, for example, Refs. [1–8], and references therein. This problem, called retrodiction or smoothing in the engineering literature [9,10], is well defined in classical statistics and, indeed, a common and uncontroversial endeavor in human activities, but its definition in quantum mechanics is less settled. The core of the issue is the compatibility between the inferred observable and the measured observable in the Heisenberg picture. If the future value of an observable ahead of the measurement is to be inferred, as in the prediction and filtering problem, the compatibility holds, as per the seminal work of Belavkin [11]. The compatibility implies that the observables can, in principle, be measured jointly, and the laws of classical probability apply to their joint statistics. This so-called nondemolition principle (NDP) underlies the quantum prediction and filtering theory pioneered by Belavkin [11,12]. In the retrodiction and smoothing problem, on the other hand, the observables may not commute, and although many have proposed quantum formulations of the task [2,6,13–15], such as the weak values [2], others have argued that the inference does not make sense if it violates the NDP [3].

To clarify the murky state of affairs, here I propose a formalism that unifies some of the existing attempts at the quantum retrodiction and smoothing theory. The formalism is based on a principle of quantum inference that supersedes the NDP. I start with Ohki's proposal of quantum minimum-mean-square-error estimators based on certain inner products [15]. The definition of an error, and an estimator to minimize it, gives a concrete decision-theoretic meaning to the inference problem, even if the involved observables do not commute. I then convert the Heisenberg-picture approach of Ohki to a formalism based on open quantum system theory in the Schrödinger picture. Remarkably, the optimal estimators then coincide with some of the generalized conditional expectations (GCEs) that have been studied in mathematical physics; see, for example, Refs. [16–20], and references therein. In the latter context, the concept has a long history and has been found to be useful as an intermediate mathematical tool, but there does not seem to be any attempt at applying it to a physical quantum inference problem or relating it to quantum retrodiction and smoothing. The main goal of this paper is to forge connections between the different areas. Many other prior works on quantum inference [1,2,5,13,14,21–24] emerge as special cases of the formalism here.

This paper is structured as follows: To set the stage, Sec. II reviews the established concepts of classical conditional expectation and the NDP. Section III presents the formalism of GCEs. Section IV studies the application of the GCEs to problems that obey the NDP, including the classical and hybrid retrodiction and smoothing problems 1,5,13–15,21,22] and some quantum estimation problems [23,24]. Section V presents some examples that may violate the principle, namely, the weak values [2] and the application of retrodiction and smoothing to linear Gaussian systems [13,14]. Section VI is the conclusion.

II. REVIEW OF ESTABLISHED CONCEPTS

A. Classical conditional expectation

Before discussing quantum generalizations of the conditional expectation, I first review the concept in classical probability theory [25,26]. Let (Ω, Σ, P) be a probability space, where Ω, the sample space, is the set of all possible outcomes of an experiment, Σ is a sigma-algebra that consists of

[*]mankei@nus.edu.sg; https://blog.nus.edu.sg/mankei/

subsets of Ω, and $P : \Sigma \to [0, 1]$ is a probability measure. If $A : \Omega \to \mathbb{R}$ is a Borel map that models a real-valued random variable and Σ_1 is a subsigma-algebra of Σ, then the expectation of A conditioned on Σ_1, denoted by $E(A|\Sigma_1) : \Omega \to \mathbb{R}$, is a Σ_1-measurable function defined by

$$\int_{S_1} A(\omega)P(d\omega) = \int_{S_1} E(A|\Sigma_1)_\omega P(d\omega) \; \forall \, S_1 \in \Sigma_1. \tag{2.1}$$

The conditional probability of an event $S \in \Sigma$ is then given by $E(1_S|\Sigma_1)$, where 1_S is the indicator function [$1_S(\omega) = 1$ if $\omega \in S$ and 0 otherwise]. To be more concrete, suppose that Σ_1 is generated by an observed random variable $Y : \Omega \to \mathcal{Y}$, in the sense of

$$\Sigma_1 = \{Y^{-1}(S) : S \in \mathcal{B}_\mathcal{Y}\}, \tag{2.2}$$

$$Y^{-1}(S) \equiv \{\omega : Y(\omega) \in S\}, \tag{2.3}$$

where $\mathcal{Y}$ is a topological space and $\mathcal{B}_\mathcal{Y}$ is the Borel sigma-algebra with respect to $\mathcal{Y}$. Then any Σ_1-measurable function is simply a function that can be expressed in terms of another Borel map $c : \mathcal{Y} \to \mathbb{R}$ as

$$C(\omega) = c[Y(\omega)] \equiv (Y^* c)(\omega), \tag{2.4}$$

where Y^* denotes the pullback. Henceforth, I denote the conditional expectation as $E(A|Y)$ if Σ_1 is generated by Y.

If we restrict our attention to random variables with finite variance, the conditional expectation can be defined in terms of Hilbert-space theory [25]. Define an inner product between two real-valued random variables as

$$\langle B, A\rangle_P \equiv \int B(\omega)A(\omega)P(d\omega), \tag{2.5}$$

the associated norm as

$$\|A\|_P \equiv \sqrt{\langle A, A\rangle_P}, \tag{2.6}$$

and the distance between two random variables as

$$d_P(A, B) \equiv \|A - B\|_P. \tag{2.7}$$

A Hilbert space $L_2(P)$ of random variables can then be constructed, with each element corresponding to an equivalence class of random variables with zero distance between them, while a subspace $L_2^Y(P)$ can be constructed from the Σ_1-measurable functions. $E(A|Y)$ can be defined as the $L_2^Y(P)$ element that satisfies

$$\langle C, A\rangle_P = \langle C, E(A|Y)\rangle_P \; \forall \, C \in L_2^Y(P), \tag{2.8}$$

which implies that $E(A|Y)$ is the projection of A into $L_2^Y(P)$. It follows from basic Hilbert-space theory that $E(A|Y)$ is the $L_2^Y(P)$ element closest to A.

If A is hidden and B is the estimator given Y in a Bayesian inference problem, then the distance given by Eq. (2.7) is the root-mean-square error, and $E(A|Y)$ is the minimum-mean-square-error estimator of A given Y [25].

Since $E(A|Y)_\omega$ is a Σ_1-measurable function, it can be expressed in terms of a Borel map $\breve{a} : \mathcal{Y} \to \mathbb{R}$ as

$$E(A|Y)_\omega = \breve{a}[Y(\omega)]. \tag{2.9}$$

Let P_Y be the coarse-grained measure induced by Y, defined as

$$P_Y(S) \equiv P[Y^{-1}(S)], \quad S \in \mathcal{B}_\mathcal{Y}. \tag{2.10}$$

Then the right-hand side of Eq. (2.8) can be expressed as

$$\langle C, E(A|Y)\rangle_P = \int c(y)\breve{a}(y)P_Y(dy) = \langle c, \breve{a}\rangle_{P_Y}, \tag{2.11}$$

and Eq. (2.8) becomes

$$\langle Y^* c, A\rangle_P = \langle c, \breve{a}\rangle_{P_Y} \; \forall \, c \in L_2(P_Y). \tag{2.12}$$

This equation for the conditional expectation turns out to be the most convenient one for quantum generalizations.

To make a connection with more elementary probability theory, suppose that A can be expressed as a Borel map $a : \mathcal{X} \to \mathbb{R}$ of another random variable $X : \Omega \to \mathcal{X}$, viz.,

$$A(\omega) = a[X(\omega)], \tag{2.13}$$

and assume that the ranges $\mathcal{X}$ and $\mathcal{Y}$ of $X(\omega)$ and $Y(\omega)$ are finite sets. Let

$$P_{XY}(x, y) \equiv P(\{\omega : X(\omega) = x \text{ and } Y(\omega) = y\}) \tag{2.14}$$

be the joint probability distribution of X and Y. Then Eq. (2.12) becomes

$$\sum_{x,y} c(y)a(x)P_{XY}(x, y) = \sum_y c(y)\breve{a}(y)P_Y(y) \forall \, c \in L_2(P_Y). \tag{2.15}$$

The solution for $\breve{a}(y)$ with $P_Y(y) > 0$ is

$$\breve{a}(y) = \sum_x a(x)\frac{P_{XY}(x, y)}{P_Y(y)}, \tag{2.16}$$

which is the Bayes theorem. Plugging Kronecker deltas in the place of $a(x)$ leads to the posterior probability distribution of X given $Y = y$.

B. Nondemolition principle

I now review Belavkin's NDP for quantum inference. To focus on the physics and avoid cumbersome mathematical technicalities, assume that all the Hilbert spaces considered hereafter (until Sec. V B) are finite-dimensional, so that the principle becomes especially simple [12]. Let $\mathcal{O}(\mathcal{H})$ be the set of operators on a Hilbert space $\mathcal{H}$ and ρ a density operator on $\mathcal{H}$. Let $A \in \mathcal{O}(\mathcal{H})$ be the hidden observable to be estimated and $B \in \mathcal{O}(\mathcal{H})$ an operator-valued estimator, both in the Heisenberg picture. Assume that both are Hermitian. Typically, B is restricted to come from a subspace of $\mathcal{O}(\mathcal{H})$. The NDP demands that A and B commute; that is, $[A, B] \equiv AB - BA = 0$. Then there exists a common orthonormal basis $\{|\omega\rangle \in \mathcal{H} : \omega \in \Omega\}$ such that A and B can be simultaneously diagonalized as

$$A = \sum_\omega a(\omega)\,|\omega\rangle\,\langle\omega|\,, \quad B = \sum_\omega b(\omega)\,|\omega\rangle\,\langle\omega| \tag{2.17}$$

in terms of some functions $a, b : \Omega \to \mathbb{R}$. The physical meaning of the compatibility is that A and B can be jointly measured by external classical observers in the same experiment, and the observers can compare the outcomes to evaluate the quality of the estimator. More generally, one may use a set of commuting

observables that are measured to construct B. With the common basis, all the outcomes observe the probability measure

$$P(\omega) = \langle\omega|\,\rho|\omega\rangle. \tag{2.18}$$

Classical probability theory, as well as the classical conditional expectation presented in Sec. II A, can then be applied to the inference problem.

One key application of the NDP is Belavkin's approach to quantum filtering, in which B is a function of the observables of a field probing a system and A is an observable of the system ahead of the field measurements. The compatibility among the observables in the Heisenberg picture can be proved for a general class of Markovian models [11]; see also Sec. 3.2.2 in Ref. [27]. The formula for the conditional expectation of an arbitrary A can then be used to derive the stochastic master equation that governs the posterior quantum state [11,12].

While the NDP has been successful in producing a mathematically satisfying theory of quantum filtering, it is, at its physical core, nothing but the orthodox quantum measurement theory that goes back to von Neumann. The use of only standard concepts is in fact a key virtue of Belavkin's work, as it clarifies that no extension to standard quantum mechanics is needed to solve his problem [11]. On the other hand, Belavkin's writings do not seem to express any strong opinion about what one should or should not do with incompatible observables. Some researchers following his work have nonetheless adopted a harder line [3,28], arguing that quantum inference should be done *only* if the NDP is observed, and retrodictive questions do not make sense, such as the question of which slit a photon goes through in a two-slit interference experiment [3].

There is no doubt that the NDP agrees with standard quantum measurement theory and quantum inference methods based on the NDP agree with classical probability theory. It is debatable, however, whether inference methods that violate the NDP should be strictly forbidden. Despite the strong view of some, the physics community has continued to demonstrate significant interest in such "forbidden" problems [2,29].

III. GENERALIZED CONDITIONAL EXPECTATIONS

A. Inner products and Hilbert spaces for operators

The mathematical physics literature has recognized other ways of generalizing the conditional expectation in quantum mechanics [16–20]. One route is to generalize the Hilbert-space treatment of random variables in Sec. II A for operators. Given two operators $A, B \in \mathcal{O}(\mathcal{H})$, define an inner product and a norm as

$$\langle B, A\rangle_\rho \equiv \mathrm{tr} B^\dagger \mathcal{E}_\rho A, \tag{3.1}$$

$$\|A\|_\rho \equiv \sqrt{\langle A, A\rangle_\rho}, \tag{3.2}$$

where † denotes the adjoint, tr denotes the trace, and $\mathcal{E}_\rho : \mathcal{O}(\mathcal{H}) \to \mathcal{O}(\mathcal{H})$ is a linear map that depends on ρ. Assume that $\mathcal{E}_\rho$ is self-adjoint and positive-semidefinite with respect to the Hilbert-Schmidt inner product

$$\langle B, A\rangle_{\mathrm{HS}} \equiv \mathrm{tr} B^\dagger A, \tag{3.3}$$

such that Eq. (3.1) also qualifies as an inner product. Assume further that $\mathcal{E}$ satisfies the following properties:

$$\mathcal{E}_\rho A = \rho A \text{ if } \rho \text{ and } A \text{ commute}, \tag{3.4}$$

$$\mathcal{E}_\rho(U^\dagger A U) = U^\dagger(\mathcal{E}_{U\rho U^\dagger} A)U, \tag{3.5}$$

$$\mathcal{E}_{\rho\otimes\rho'}(A \otimes A') = (\mathcal{E}_\rho A) \otimes (\mathcal{E}_{\rho'} A'), \tag{3.6}$$

where U is a unitary operator, ρ' is another density operator on Hilbert space $\mathcal{H}'$, and $A' \in \mathcal{O}(\mathcal{H}')$. Equation (3.4) ensures that the inner product coincides with the classical version for commuting operators. Equations (3.5) and (3.6) are desirable properties in dealing with dynamics and composite Hilbert spaces [20]. In what follows, I further require that

$$\|A \otimes I\|_\rho = \|A\|_{\mathrm{tr}'\rho}, \tag{3.7}$$

where I is the identity operator, ρ is a density operator on $\mathcal{H} \otimes \mathcal{H}'$, and tr′ is the partial trace over $\mathcal{H}'$. Equation (3.7) is a reasonable requirement for the definition of a quantum variance. Given these properties, $\mathcal{E}$ is not unique. Some prominent examples that satisfy Eqs. (3.4)–(3.6) include the left product

$$\mathcal{E}_\rho A = \rho A, \tag{3.8}$$

the Jordan product

$$\mathcal{E}_\rho A = \tfrac{1}{2}(\rho A + A\rho), \tag{3.9}$$

and the root product

$$\mathcal{E}_\rho A = \sqrt{\rho} A \sqrt{\rho}. \tag{3.10}$$

More generally, a class of $\mathcal{E}$ can be constructed from convex combinations of $\mathcal{E}_\rho A = \rho^\lambda A \rho^{1-\lambda}$, $0 \leqslant \lambda \leqslant 1$ [20,30]. With the left product, Eq. (3.1) becomes the inner product in the Gelfand-Naimark-Segal construction [31]. With the Jordan product, Eq. (3.1) becomes an inner product proposed by Holevo that is useful in quantum statistics [32,33]. The other products also have their uses in mathematical physics and statistical mechanics [18,19,30]. The left product and the Jordan product further satisfy Eq. (3.7), although the root product and many others do not [20].

With an inner product and the associated norm at hand, an operator Hilbert space $L_2(\rho)$ can be constructed from $\mathcal{O}(\mathcal{H})$, generalizing the classical $L_2(P)$ space described in Sec. II A. Because $\mathcal{H}$ is finite-dimensional, there is no need to complete the space with unbounded operators [33].

A distance between two operators can be defined as

$$d_\rho(A, B) \equiv \|A - B\|_\rho. \tag{3.11}$$

In the context of quantum inference, Eq. (3.11) can serve as a generalization of the classical root-mean-square error given by Eq. (2.7) as a performance criterion [15,34]. The choice of an estimator to minimize Eq. (3.11) may be called the minimum-error principle. Note that the principle imposes no requirement on the compatibility between A and B. The whole point of inference is that A is hidden and can only be inferred, so the NDP's requirement that A be jointly measurable with B by external classical observers may seem too restrictive if A is never measured in reality. The minimum-error principle, on the other hand, avoids the stringent requirement and serves as a more general principle for quantum inference beyond

an exact correspondence with probability theory. To quote Belavkin himself on the limitation of quantum probability theory [11]:

> It is nonsense to consider seriously a complete observation in the closed universe; there is no universal quantum observation, no universal reduction and spontaneous localization for the wave function of the world. Nobody can prepare an *a priori* state compatible with a complete world observation and reduce the *a posteriori* state, except God. But acceptance of God as an external subject of the physical world is at variance with the closeness assumption of the universe. Thus, the world state-vector has no statistical interpretation, and the humanitarian validity of these interpretations would, in any case, be zero. The probabilistic interpretation of the state-vector is relevant to only the induced states of the quantum open objects being prepared by experimentalists in an appropriate compound system for the nondemolition observation to produce the reduced states after the registration.

Unless we impose an artificial classical-quantum boundary or force quantum mechanics to serve our classical intuition, nothing in quantum mechanics mandates that observables should commute.

B. Application of open quantum system theory

To treat time evolution and open quantum systems, suppose that

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \tag{3.12}$$

consists of three Hilbert subspaces $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$,

$$\rho = \sigma \otimes \tau \tag{3.13}$$

is the density operator at, say, time t with σ being a density operator on $\mathcal{H}_1$ and τ a density operator on $\mathcal{H}_2 \otimes \mathcal{H}_3$,

$$A_t \equiv A \otimes I \otimes I \tag{3.14}$$

is the hidden observable with $A \in \mathcal{O}(\mathcal{H}_1)$,

$$B_T \equiv U^\dagger (I \otimes B \otimes I) U \tag{3.15}$$

is the Heisenberg picture of an operator-valued estimator $B \in \mathcal{O}(\mathcal{H}_2)$ at time $T \geqslant t$, and U is a unitary operator on $\mathcal{H}$ that models the time evolution from t to T. The mean-square error of B_T in inferring A_t becomes

$$d_\rho^2(A_t, B_T) = \|A\|_\sigma^2 - 2\mathrm{Re}\langle \mathcal{F}^\dagger B, A\rangle_\sigma + \|I \otimes B \otimes I\|_{U\rho U^\dagger}^2, \tag{3.16}$$

where

$$\mathcal{F}\sigma \equiv \mathrm{tr}_{13} U(\sigma \otimes \tau) U^\dagger \tag{3.17}$$

is a trace-preserving completely positive (TPCP) map $\mathcal{F}: \mathcal{O}(\mathcal{H}_1) \to \mathcal{O}(\mathcal{H}_2)$ in the Stinespring representation [20], $\mathcal{F}^\dagger$ is its adjoint with respect to the Hilbert-Schmidt inner product, tr_{13} denotes the partial trace over the Hilbert spaces numbered by the subscript ($\mathcal{H}_1 \otimes \mathcal{H}_3$ here), and the properties given by Eqs. (3.5) and (3.6) have been used. If Eq. (3.7) is also used, then

$$d_\rho^2(A_t, B_T) = D_{\sigma,\mathcal{F}}(A, B), \tag{3.18}$$

$$D_{\sigma,\mathcal{F}}(A, B) \equiv \|A\|_\sigma^2 - 2\mathrm{Re}\langle \mathcal{F}^\dagger B, A\rangle_\sigma + \|B\|_{\mathcal{F}\sigma}^2. \tag{3.19}$$

Let $B = \mathcal{F}_\sigma A$ be an estimator of A that minimizes Eq. (3.19). By substituting $B = \mathcal{F}_\sigma A + \epsilon c$ into Eq. (3.19), differentiating with respect to ϵ, and assuming that c is an arbitrary operator, it is straightforward to show that $\mathcal{F}_\sigma A$ obeys

$$\langle \mathcal{F}^\dagger c, A\rangle_\sigma = \langle c, \mathcal{F}_\sigma A\rangle_{\mathcal{F}\sigma} \;\forall\, c \in L_2(\mathcal{F}\sigma). \tag{3.20}$$

Notice that this equation is a generalization of Eq. (2.12), with σ generalizing the probability measure P, $\mathcal{F}\sigma$ generalizing the coarse-grained measure P_Y, $\mathcal{F}^\dagger c$ generalizing the pullback $Y^* c$, and $\mathcal{F}_\sigma A$ generalizing the conditional expectation $\check{a}$. The existence and uniqueness of $\mathcal{F}_\sigma A$ as an element in $L_2(\mathcal{F}\sigma)$ is guaranteed by the Riesz representation theorem [25], if the left-hand side of Eq. (3.20) is regarded as a linear functional of c, necessarily bounded in the finite-dimensional case considered here.

Equation (3.20) can be expressed in terms of the Hilbert-Schmidt inner product as

$$\langle \mathcal{F}^\dagger c, \mathcal{E}_\sigma A\rangle_{\mathrm{HS}} = \langle c, \mathcal{F}\mathcal{E}_\sigma A\rangle_{\mathrm{HS}} = \langle c, \mathcal{E}_{\mathcal{F}\sigma}\mathcal{F}_\sigma A\rangle_{\mathrm{HS}}. \tag{3.21}$$

Since c is arbitrary, the equation is reduced to

$$\mathcal{E}_{\mathcal{F}\sigma}\mathcal{F}_\sigma A = \mathcal{F}\mathcal{E}_\sigma A. \tag{3.22}$$

Equation (3.22) coincides with the general definition of a generalized conditional expectation (GCE) given by Eq. (6.21) in Ref. [20]; see also Ref. [18]. If $\mathcal{E}_{\mathcal{F}\sigma}$ is invertible, $\mathcal{F}_\sigma: \mathcal{O}(\mathcal{H}_1) \to \mathcal{O}(\mathcal{H}_2)$ as a map is explicitly given by

$$\mathcal{F}_\sigma = \mathcal{E}_{\mathcal{F}\sigma}^{-1}\mathcal{F}\mathcal{E}_\sigma. \tag{3.23}$$

Any $\mathcal{F}_\sigma A$ that satisfies Eq. (3.22) leads to

$$\min_{B \in L_2(\mathcal{F}\sigma)} D_{\sigma,\mathcal{F}}(A, B) = \|A\|_\sigma^2 - \|\mathcal{F}_\sigma A\|_{\mathcal{F}\sigma}^2. \tag{3.24}$$

Equation (3.22) can be used to define a GCE even if the $\mathcal{E}$ map does not satisfy Eq. (3.7). For example, the Accardi-Cecchini GCE [16] given by

$$\mathcal{F}_\sigma A = (\mathcal{F}\sigma)^{-1/2}\mathcal{F}(\sqrt{\sigma}A\sqrt{\sigma})(\mathcal{F}\sigma)^{-1/2} \tag{3.25}$$

results from Eq. (3.23) if $\mathcal{E}$ is the root product given by Eq. (3.10) [18,19]. Then the GCE still has the meaning of an operator that minimizes the D in Eq. (3.19). As long as $\mathcal{E}$ leads to $\|I \otimes B \otimes I\|_{U\rho U^\dagger} \leqslant \|B\|_{\mathcal{F}\sigma}$ and thus $d_\rho^2(A_t, B_T) \leqslant D_{\sigma,\mathcal{F}}(A, B)$, Eq. (3.24) and therefore Eq. (3.19) remain nonnegative [20].

It is straightforward to show that many versions of the $\mathcal{E}$ map satisfy the property of mapping Hermitian operators to Hermitian operators. An example is

$$\mathcal{E}_\rho A = \tfrac{1}{2}(\rho^\lambda A \rho^{1-\lambda} + \rho^{1-\lambda} A \rho^\lambda), \quad 0 \leqslant \lambda \leqslant 1, \tag{3.26}$$

which includes the Jordan product and the root product. More generally, any convex combination of $\mathcal{E}$'s satisfying the property also satisfies it. The inner product then becomes real if the operators are restricted to be Hermitian. Since any completely positive map $\mathcal{F}$ also satisfies such a property (as easily proved using its Kraus representation), the GCE given by Eq. (3.23) associated with such an $\mathcal{E}$ map also maps Hermitian operators to Hermitian operators, a property that some may find desirable.

At this juncture, the works by Leifer and Spekkens [7] and Horsman and coworkers [8] on two-time quantum states deserve a mention. They diverged from the formalism here by expressing the TPCP map in its Choi form, although they also recognized the usefulness of the root product or the Jordan product in their formalism. It is outside the scope of this

paper to investigate the connections of these works to those considered here.

C. Discussion

Among the choices of $\mathcal{E}$ considered here, the Jordan product seems to stand out as the most reasonable, as it satisfies all the desirable properties given by Eqs. (3.4)–(3.7) and gives a Hermitian GCE for a Hermitian hidden observable. With the Jordan product and Hermitian A_t and B_T, the error given by Eq. (3.11) becomes

$$d_\rho^2(A_t, B_T) = \mathrm{tr}\rho(A_t - B_T)^2. \tag{3.27}$$

$A_t - B_T$ is another Hermitian observable that can be measured in principle, and the resulting variance coincides with the error, so the error does have a probabilistic interpretation. Although a measurement of $A_t - B_T$ seems difficult in practice, this interpretation does not appear to be much less reasonable than the interpretation of the NDP in terms of the joint measurement of A_t and B_T, which is also quite impractical.

Along the same line, for parameter estimation problems, where A models a classical real parameter, the Jordan product has the desirable feature of making the error agree with the classical estimation error upon the measurement of a Hermitian B [23,24].

In the more mathematical context of quantum information theory, the Jordan product leads to the smallest—and thus the most useful—quantum version of the Fisher information for scalar parameter estimation [20,35]. The GCE given by Eq. (3.22) determines the relation between the scores (logarithmic derivative operators) of a quantum parametric model before and after the $\mathcal{F}$ map, generalizing the classical case (see, for example, Sec. 25.5 in Ref. [36]), and the fact that $\mathcal{F}_\sigma$ is a projection can be used to prove the monotonicity of the quantum Fisher information [20,37].

Other problems may require a different product or a different GCE, however. For example, the left product leads to a quantum Fisher information matrix that may be more useful for vectoral parameter estimation [33,38], while the Accardi-Cecchini GCE and its Hilbert-Schmidt adjoint, the Petz recovery map, are central to the study of quantum channel sufficiency [17–19]. There is no single best definition of a GCE, nor does there need to be, just as there is no single best quasiprobability representation for every problem.

IV. EXAMPLES THAT OBEY THE NONDEMOLITION PRINCIPLE

A. Classical

The classical case is unequivocal. Let $P_X(x)$ be the prior probability distribution of a hidden random variable $X \in \mathcal{X}$ and $P_{Y|X}(y|x)$ be the probability distribution of the observed $Y \in \mathcal{Y}$ conditioned on $X = x$. For simplicity, assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite sets and all the probabilities are positive. Then $P_Y(y) = \sum_x P_{Y|X}(y|x)P_X(x) > 0$. Let

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

$$A = \sum_x a(x)\,|x\rangle\,\langle x|\,, \tag{4.2}$$

$$\mathcal{F}\sigma = \sum_{y,x} P_{Y|X}(y|x)\,\langle x|\,\sigma\,|x\rangle\,|y\rangle\,\langle y|\,, \tag{4.3}$$

$$\mathcal{F}_\sigma A = \sum_y \check{a}(y)\,|y\rangle\,\langle y|\,, \tag{4.4}$$

where $\{|x\rangle : x \in \mathcal{X}\}$ is an orthonormal basis of $\mathcal{H}_1$, $\{|y\rangle : y \in \mathcal{Y}\}$ is an orthonormal basis of $\mathcal{H}_2$, $a : \mathcal{X} \to \mathbb{C}$ is a classical random variable, and $\check{a} : \mathcal{Y} \to \mathbb{C}$ is an estimator given the observation. The GCE given by Eq. (3.23), regardless of the choice of the $\mathcal{E}$ map, becomes

$$\check{a}(y) = \sum_x \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}a(x), \tag{4.5}$$

which is, of course, the Bayes theorem.

B. Retrodiction

For an example with quantum ingredients, consider the retrodiction problem studied by Watanabe [1] and Barnett, Pegg, and Jeffers [5]. Assume the classical model given by Eqs. (4.1)–(4.5). In addition, assume that the observation distribution $P_{Y|X}(y|x)$ arises from Born's rule, viz.,

$$P_{Y|X}(y|x) = \mathrm{tr}M(y)\mathcal{G}\rho_x = \langle M(y), \mathcal{G}\rho_x\rangle_{\mathrm{HS}} \tag{4.6}$$

$$= \left\langle \mathcal{G}^\dagger M(y), \rho_x \right\rangle_{\mathrm{HS}}, \tag{4.7}$$

where ρ_x is the density operator of a quantum system conditioned on $X = x$ at an initial time, M is the positive operator-valued measure (POVM) that models a measurement of the system at a final time, and $\mathcal{G}$ is a TPCP map that models the time evolution in-between. As recognized by Refs. [1,5], one can choose to evolve ρ_x forward in time by $\mathcal{G}$ or evolve $M(y)$ backward in time by $\mathcal{G}^\dagger$. Either way, the GCE for this retrodiction problem is still the classical Bayes theorem given by Eq. (4.5) and can be written as

$$\check{a}(y) = \sum_x \frac{\mathrm{tr}M(y)\mathcal{G}\rho(x)}{\sum_{x'}\mathrm{tr}M(y)\mathcal{G}\rho(x')}a(x), \tag{4.8}$$

$$\rho(x) \equiv \rho_x P_X(x), \tag{4.9}$$

where $\rho(x)$ is the hybrid density operator [39]. In this context, there is some freedom in how one normalizes $M(y)$ and $\rho(x)$ [5].

Although Barnett and coworkers called this problem quantum retrodiction, here it may be more appropriate to specify it as hybrid retrodiction, since the hidden observable is classical, while the observation arises from Born's rule in quantum mechanics.

C. Hybrid smoothing

A generalization of the retrodiction problem is the so-called smoothing problem [9,10], where the value of a time-varying waveform at a certain intermediate time t is to be estimated using observations both before and after t. Consider in particular a classical waveform, such as a gravitational wave, perturbing a quantum system, such as an optomechanical sensor. Sequential or continuous measurements are made on the quantum system, and the outcomes are used to infer the waveform. To model this hybrid smoothing problem, assume

again the classical model given by Eqs. (4.1)–(4.5). Let the classical waveform value at an intermediate time t be X and the measurement model be

$$P_{Y|X}(y|x) = \mathrm{tr}M(y|x)\rho_x, \tag{4.10}$$

where ρ_x is now the density operator of the quantum system at time t conditioned on $X = x$, the POVM $M(y|x)$ is also conditioned on $X = x$, and Y represents all the "future" observations after t. All quantities, including $P_X(x)$, are implicitly assumed to be conditioned on the "past" observations before t. The GCE becomes

$$\check{a}(y) = \sum_x \frac{\mathrm{tr}M(y|x)\rho(x)}{\sum_{x'}\mathrm{tr}M(y|x')\rho(x')}a(x), \tag{4.11}$$

which is the central formula employed by Tsang in his hybrid smoothing theory [13,14] (with the obvious generalization of $P_X(x)$ to a density and $\sum_x$ to an integral).

Hybrid filtering refers to the estimation of X with observations up to time t only and can be accomplished by computing the $\rho(x)$ conditioned on the past observations [39]. As is well known in engineering [10], smoothing is more accurate than filtering if the waveform is stochastic, since the future observations contain information about the waveform value that is absent in the past observations. These considerations arguably make the hybrid smoothing theory the most useful offshoot of the quantum retrodiction and smoothing formalism.

If a continuous measurement is performed on the quantum system, $M(y|x)$ and $\rho(x)$ can be solved via a time-symmetric pair of stochastic master equations, as proposed by Tsang [13,14]. $\rho(x)$ is to be solved using a forward-time stochastic master equation for hybrid filtering that goes from an initial time to the intermediate time t, while $M(y|x)$ is to be solved using an adjoint equation that goes backward from a final time to t, generalizing the retrodiction formalism in Sec. IV B.

Equation (4.11) may be solved more efficiently by considering the quasiprobability representations of $M(y|x)$ and $\rho(x)$, which can admit more succinct forms in special cases. For example, for linear Gaussian systems, the Wigner representations of $M(y|x)$ and $\rho(x)$ are both Gaussian, and the smoother coincides with the optimal linear smoother in the classical setting [10,13,14]. The "Gaussian theory of hindsight" proposed earlier by Petersen and Mølmer for atomic magnetometry [40] can then be viewed as a special case. Tsang, Wiseman, and Caves showed that the smoothing technique is needed to achieve the fundamental quantum limit to waveform estimation in optomechanical force sensing [41]. References [42] report experimental demonstrations of the smoothing technique for quantum optical systems.

The more recent proposal of "past quantum state" by Gammelmark, Julsgaard, and Mølmer [21] is nothing but a special case of the hybrid smoothing theory. Let $\{\kappa_x : x \in \mathcal{X}\}$ be a set of Kraus operators that model a measurement at time t and ρ' be the quantum state before the measurement. Their theory can be reproduced from Eq. (4.11) by assuming

$$\rho_x = \frac{\kappa_x\rho'\kappa_x^\dagger}{\mathrm{tr}\kappa_x\rho'\kappa_x^\dagger}, \quad P_X(x) = \mathrm{tr}\kappa_x\rho'\kappa_x^\dagger, \tag{4.12}$$

$$\rho(x) = \rho_x P_X(x) = \kappa_x\rho'\kappa_x^\dagger. \tag{4.13}$$

In other words, the measurement outcome X that is assumed to be hidden in their setup can simply be treated as a classical random variable in the hybrid theory; see also Ref. [3]. Mølmer and coworkers have since produced a series of papers on the subject [6], such as Ref. [43], which rediscovers the hybrid smoothing theory.

The quantum state smoothing theory proposed by Guevara and Wiseman [6,22], which concerns the estimation of a time-dependent density matrix with partial observations, may also be considered as a special case of hybrid smoothing. In their scenario, a partial observer named Alice has only partial access to a sequence of observations of a quantum system, and her goal is to estimate the density matrix possessed by an omniscient observer named Bob, who has complete access to the observations and updates his density matrix continuously with them. The key is to view the problem as a generalization of quantum state tomography, where the density matrix is a *classical* matrix-valued parameter [44]. Under this view, Bob's density matrix is a classical stochastic process, whose equation of motion happens to be the stochastic master equation driven by the observations as an effective system noise. The hybrid approach should therefore be applicable to the state estimation problem, although the technical details remain to be worked out. It is an interesting open question whether the use of other cost functions in Ref. [6] may be applied to the hybrid theory.

D. Optimal quantum estimation of a classical parameter

A generalization of the previous examples is to allow the experimenter to pick any quantum measurement that minimizes the error. Keep the classical σ and A given by Eqs. (4.1) and (4.2), but assume that the output state

$$\mathcal{F}\sigma = \sum_x \rho_x \langle x|\,\sigma\,|x\rangle \tag{4.14}$$

is quantum. Then the GCE given by Eq. (3.22) obeys

$$\mathcal{E}_{\sum_x \rho(x)}\mathcal{F}_\sigma A = \sum_x \rho(x)a(x). \tag{4.15}$$

If a is real and $\mathcal{E}$ is the Jordan product, any solution for $\mathcal{F}_\sigma A$ is an optimal observable to be measured for estimating $a(x)$, as discovered by Personick [23]; see Ref. [45] for a more recent related work.

E. Optimal quantum estimation

The quantum estimation problem considered by Belavkin and Grishanin [24] is a further variation of the previous examples. Let the larger Hilbert space be $\mathcal{H}_1 \otimes \mathcal{H}_2$ and the state on this Hilbert space be σ. The parameter to be estimated is now a quantum observable A on $\mathcal{H}_1$, while the estimator is another quantum observable B on $\mathcal{H}_2$. Assume that both are Hermitian. Their TPCP map is simply the partial trace $\mathcal{F}\sigma = \mathrm{tr}_2\sigma$. The resulting optimal observable obtained by Belavkin and Grishanin coincides with the GCE here in terms of the Jordan product.

F. Compliance with the nondemolition principle

All the previous examples obey the NDP. To check this explicitly for the examples in Secs. IV A–IV D, notice that the TPCP maps there can all be expressed in the Stinespring representation given by Eq. (3.17) if one assumes

$$\tau = |\phi\rangle\langle\phi|, \tag{4.16}$$

$$U = \sum_x |x\rangle\langle x| \otimes V_x, \tag{4.17}$$

where $|\phi\rangle$ is a pure state in $\mathcal{H}_2 \otimes \mathcal{H}_3$, $\{|x\rangle : x \in \mathcal{X}\}$ is the orthonormal basis of $\mathcal{H}_1$ assumed in Secs. IV A–IV D, and $V_x \in \mathcal{O}(\mathcal{H}_2 \otimes \mathcal{H}_3)$ is a unitary operator controlled by x. The TPCP map becomes

$$\mathcal{F}\sigma = \sum_x \langle x|\sigma|x\rangle \operatorname{tr}_3 V_x |\phi\rangle\langle\phi| V_x^\dagger. \tag{4.18}$$

The classical model in Sec. IV A, the retrodiction model in Sec. IV B, and the hybrid smoothing model in Sec. IV C are obtained if

$$\operatorname{tr}_3 V_x |\phi\rangle\langle\phi| V_x^\dagger = \sum_y P_{Y|X}(y|x)|y\rangle\langle y|, \tag{4.19}$$

while Personick's model in Sec. IV D is obtained if

$$\operatorname{tr}_3 V_x |\phi\rangle\langle\phi| V_x^\dagger = \rho_x. \tag{4.20}$$

Given the density operator on the right-hand side of Eq. (4.19) or Eq. (4.20), standard open quantum system theory [20] assures that there always exist a purification of the density operator and a unitary V_x that maps $|\phi\rangle$ to the purification. The fact that $P_{Y|X}(y|x)$ in the hybrid models arises from Born's rule turns out to be irrelevant to the compatibility between the hidden observable and the estimator.

In the larger Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$, the hidden observable A_t and the estimator observable B_T in the Heisenberg picture are given by Eqs. (3.14) and (3.15), respectively. The classical A given by Eq. (4.2) becomes

$$A_t = \sum_x a(x)|x\rangle\langle x| \otimes I \otimes I, \tag{4.21}$$

while applying the U given by Eq. (4.17) to Eq. (3.15) leads to

$$B_T = \sum_x |x\rangle\langle x| \otimes V_x^\dagger (B \otimes I) V_x. \tag{4.22}$$

The A_t and B_T here commute, meaning that the NDP is observed.

To treat infinite-dimensional problems, the obvious generalization would be to replace $|x\rangle\langle x|$ by a projection-valued measure and $\sum_x$ by an integral, although a rigorous analysis of the infinite-dimensional case is outside the scope of this paper.

For the problem considered by Belavkin and Grishanin and described in Sec. IV E, the observables in the larger Hilbert space are simply $A \otimes I$ and $I \otimes B$, so their compatibility is obvious.

V. EXAMPLES THAT MAY VIOLATE THE NONDEMOLITION PRINCIPLE

A. Weak values

For an example that may violate the NDP, let σ and A be quantum, A be Hermitian, $\mathcal{F}$ be the measurement map given by

$$\mathcal{F}\sigma = \sum_y [\operatorname{tr} M(y)\sigma]|y\rangle\langle y|, \tag{5.1}$$

and the estimator be the classical form given by Eq. (4.4). Then Eq. (3.23) becomes

$$\check{a}(y) = \frac{\operatorname{tr} M(y)\mathcal{E}_\sigma A}{\operatorname{tr} M(y)\sigma}. \tag{5.2}$$

This GCE coincides with the weak value if $\mathcal{E}$ is the left product and the real part of the weak value if $\mathcal{E}$ is the Jordan product [15]. The weak values may be generalized by considering other operator products for $\mathcal{E}$, while a larger class of quasiprobability distributions that are conditioned on both past and future observations may be generated by plugging quantum generalizations of the Kronecker deltas in the place of A, such as projectors [34] or phase-point operators [46].

As the weak value is well known to create paradoxes that defy classical logic [2], one should not expect it to obey the NDP in general. To check, notice that the measurement map given by Eq. (5.1) can be expressed in the Stinespring representation given by Eq. (3.17) if one assumes

$$M(y) = \operatorname{tr}_3 \Pi_{13}(y)(I \otimes \tau_3), \tag{5.3}$$

$$\tau = |\phi\rangle\langle\phi| \otimes \tau_3, \tag{5.4}$$

$$U = \sum_y \Pi_{13}(y) \otimes V_y, \tag{5.5}$$

where $\Pi_{13} \in \mathcal{O}(\mathcal{H}_1 \otimes \mathcal{H}_3)$ is the projection-valued measure and $\tau_3 \in \mathcal{O}(\mathcal{H}_3)$ is the ancilla state that arise from the Naimark extension of $M(y)$ [20], while $V_y \in \mathcal{O}(\mathcal{H}_2)$ is a unitary that gives $V_y|\phi\rangle = |y\rangle$. Note that Eq. (5.5) is expressed in the order of $\mathcal{O}(\mathcal{H}_1 \otimes \mathcal{H}_3) \otimes \mathcal{O}(\mathcal{H}_2)$. Following the same order, the operator-valued estimator given by Eq. (3.15) can be expressed as

$$B_T = \sum_y \Pi_{13}(y) \otimes V_y^\dagger B V_y, \tag{5.6}$$

which may not commute with the hidden observable given by Eq. (3.14).

Despite the possible violation of the NDP, the weak values are not necessarily paradoxical. The literature on the weak values tends to focus on the paradoxes, but it is arguably more important to demonstrate that they make sense for large classes of problems, so that they can serve as a discerning test of nonclassicality. The next section demonstrates that, for a certain class of systems called linear Gaussian systems, a classical probability model can indeed be constructed even for incompatible observables, thus guaranteeing the inference to conform with classical logic.

B. Linear Gaussian systems

A large class of quantum optics experiments, such as optomechanics, can be modeled as linear Gaussian systems [39,47,48]. They are defined by the following conditions:

(1) The Wigner representations of all the density operators involved are Gaussian.

(2) The observables of interest are restricted to quadrature operators, defined as real linear combinations of canonical position and momentum operators.

(3) All the unitary operators involved are generated by Hamiltonians that are quadratic with respect to the quadrature operators, such that the equations of motion for the quadratures are linear.

(4) The measurements are restricted to spectral resolutions of the quadrature operators, such as homodyne detection in optics.

These conditions can also be applied to TPCP maps and POVMs via their Stinespring or Naimark representations. It is well established that, for linear Gaussian systems, the Wigner representation offers a classical probability model for the quadratures, despite the incompatibility among them [39,47,48].

Assume a quantum system with N bosonic modes with density operator σ at time t. Let Q be a column vector of the $2N$ phase-space (position and momentum) operators. Some of the phase-space operators can be used to model classical continuous variables in a hybrid problem [49]. Suppose that the measurement outcomes from the system are used to estimate a quadrature operator A, which can be written as

$$A = b^\top Q, \tag{5.7}$$

where b is a column vector of $2N$ constants and $\top$ denotes the transpose. Let $Y \in \mathcal{Y} = \mathbb{R}^L$ be the noisy observation of L quadratures after t and M be the corresponding POVM. Assume that M can be expressed as

$$M(S) = \int_S d^L y \gamma(y), \tag{5.8}$$

where $\gamma(y)$ is an operator-valued density of M. The probability density of Y is then

$$f_Y(y) = \mathrm{tr}\gamma(y)\sigma. \tag{5.9}$$

For example, if ideal homodyne detection is performed, $\gamma(y) = |y\rangle\langle y|$ in terms of the Dirac eigenkets $\{|y\rangle : y \in \mathbb{R}^L, \langle y|y'\rangle = \delta^L(y - y')\}$ for the L quadratures [48]. Of course, $\gamma(y)$ can also incorporate the effects of more general dynamics and measurements after t.

Assume the Jordan product for the $\mathcal{E}$ map hereafter. The $L_2(\sigma)$ space remains well defined for this infinite-dimensional problem if one completes the space with limit points of the space of bounded operators [33], although I ignore the mathematical complications that may arise when generalizing the earlier finite-dimensional results in the following, as per standard practice in quantum optics [48]. The appropriate generalization of the $L_2(\mathcal{F}\sigma)$ space in this case is the classical $L_2(f_Y)$ space with the inner product $\langle b, a\rangle_{f_Y} \equiv \int d^L y b(y)a(y)f_Y(y)$, while the appropriate generalization of Eq. (5.2) is

$$\check{a}(y) = \frac{\mathrm{tr}\gamma(y)\mathcal{E}_\sigma A}{\mathrm{tr}\gamma(y)\sigma}. \tag{5.10}$$

Let $W_\sigma(q)$ and $W_\gamma(y|q)$ be the Wigner representations of σ and $\gamma(y)$, respectively, where q is a column vector of the $2N$ phase-space coordinates. For the smoothing problem, σ and W_σ are implicitly assumed to be conditioned on the past observations before t. It is straightforward to show that the GCE given by Eq. (5.10) becomes [14]

$$\check{a}(y) = \int d^{2N} q W(q|y) b^\top q, \tag{5.11}$$

$$W(q|y) = \frac{W_\gamma(y|q)W_\sigma(q)}{\int d^{2N}q' W_\gamma(y|q')W_\sigma(q')}, \tag{5.12}$$

$$d^{2N}q \equiv dq_1 \cdots dq_{2N}. \tag{5.13}$$

Equation (5.12) has the form of the Bayes theorem, with W_σ and W_γ playing the roles of P_X and $P_{Y|X}$ in the classical model in Sec. IV A.

For a linear Gaussian system, the Gaussianity of W_σ and W_γ can be proved [47,48]. Assume

$$W_\sigma(q) \propto \exp\left[-\tfrac{1}{2}(q - \check{q}_\sigma)^\top K_\sigma^{-1}(q - \check{q}_\sigma)\right], \tag{5.14}$$

$$W_\gamma(y|q) \propto \exp\left[-\tfrac{1}{2}(y - hq)^\top R^{-1}(y - hq)\right], \tag{5.15}$$

where $\check{q}_\sigma$ and K_σ are the mean vector and covariance matrix of W_σ, y is an L-dimensional column vector, h is an $L \times 2N$ matrix, and R is an $L \times L$ covariance matrix. Since both W_σ and W_γ are positive here, the GCE has all the nice properties of a classical conditional expectation, even if the NDP is violated.

Assume that $h^\top R^{-1} h$ is positive-definite, such that its inverse is defined. Let

$$W_\gamma(y|q) \propto \exp\left[-\tfrac{1}{2}(q - \check{q}_\gamma)^\top K_\gamma^{-1}(q - \check{q}_\gamma)\right], \tag{5.16}$$

$$\check{q}_\gamma = K_\gamma h^\top R^{-1} y, \tag{5.17}$$

$$K_\gamma = (h^\top R^{-1} h)^{-1}. \tag{5.18}$$

Then the posterior distribution given by Eq. (5.12) becomes

$$W(q|y) \propto \exp\left[-\tfrac{1}{2}(q - \check{q})^\top K^{-1}(q - \check{q})\right], \tag{5.19}$$

$$\check{q} = K\left(K_\sigma^{-1}\check{q}_\sigma + K_\gamma^{-1}\check{q}_\gamma\right), \tag{5.20}$$

$$K = \left(K_\sigma^{-1} + K_\gamma^{-1}\right)^{-1}, \tag{5.21}$$

and the GCE of A becomes

$$\check{a} = b^\top \check{q}. \tag{5.22}$$

The mean-square error is given by Eqs. (3.18) and (3.24), and one can use the correspondence between the operator inner product $\langle\cdot,\cdot\rangle_\sigma$ for quadratures and the classical inner product with respect to the Wigner function $\langle\cdot,\cdot\rangle_{W_\sigma}$ [39] to obtain

$$d_\rho^2 = b^\top K b. \tag{5.23}$$

Note that $\check{q}$ is the vectoral GCE of the quadrature operator vector Q and may violate the NDP as an estimator, since it is a function of the retrodictive estimator $\check{q}_\gamma$ and thus the future

observations. Despite the controversial status of $\check{q}$, the hidden observable of interest A may still obey the NDP with respect to the measurement, in a hybrid smoothing problem for example. Then no one can deny the usefulness of the retrodictive estimator, as an intermediate tool at least, in determining the final conditional expectation of A. Compared with the filtering estimator and its error based on σ alone, given by

$$\check{a} = b^\top \check{q}_\sigma, \quad d_\rho^2 = b^\top K_\sigma b, \tag{5.24}$$

the advantage of smoothing may be substantial.

As noticed first by Refs. [14,50], the posterior covariance matrix K given by Eq. (5.21), unlike K_σ or K_γ, may violate the Heisenberg uncertainty relation, but it is not a paradox from the perspective of the effective classical model with the positive Wigner functions.

VI. CONCLUSION

I have presented a unifying theory of generalized conditional expectations (GCEs) for quantum retrodiction and smoothing. It is fair to say that I have drawn on many prior works in establishing the theory and constructing the examples. Rather than proposing yet another approach to the problem, the key contribution of this paper is to make hitherto unappreciated connections among the existing works. As these works come from diverse fields in physics and engineering and have different motivations, it is worth pointing out that they share a common thread and offer complementary perspectives on the quantum retrodiction and smoothing problem.

On one hand, the GCEs are shown to be natural consequences of generalizing the Hilbert-space treatment of the classical conditional expectation. On the other, they are shown to coincide with many quantum estimation methods that have philosophical as well as engineering implications. These considerations establish the GCEs as a principled and sensible concept for quantum inference, enabling one to address questions previously forbidden by the nondemolition principle (NDP). Providing answers to such questions, even if they seem paradoxical in special cases, may serve as a mental aid for researchers and students to develop insights and intuition about quantum systems. On a more practical level, the hybrid smoother, discussed in Secs. IV C and V B, illustrates how an inference about the past of a quantum sensor can improve the estimation of a classical waveform. Regardless of one's position on the NDP, there is no denying that the GCEs can, at the very least, serve as useful intermediate tools in quantum sensing and information-processing applications.

Given the fundamental importance of the conditional expectation in classical probability and statistics, there should be plenty of room for the concept of GCEs to grow even further in the quantum arena.

ACKNOWLEDGMENTS

This research is supported by the National Research Foundation (NRF) Singapore, under its Quantum Engineering Programme (Grant No. QEP-P7).

[1] S. Watanabe, Symmetry of physical laws. Part III. Prediction and retrodiction, Rev. Mod. Phys. **27**, 179 (1955).

[2] Y. Aharonov, D. Z. Albert, and L. 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, Phys. Rev. Lett. **60**, 1351 (1988); Y. Aharonov and D. Rohrlich, *Quantum Paradoxes: Quantum Theory for the Perplexed* (Wiley, New York, 2005); J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Colloquium: Understanding quantum weak values: basics and applications, Rev. Mod. Phys. **86**, 307 (2014).

[3] J. Gough, How to estimate past quantum measurement interventions after continuous monitoring, Russ. J. Math. Phys. **27**, 218 (2020).

[4] J. E. Gough, The collapse before a quantum jump transition, Rep. Math. Phys. **85**, 69 (2020).

[5] S. M. Barnett, J. Jeffers, and D. T. Pegg, Quantum retrodiction: Foundations and controversies, Symmetry **13**, 586 (2021); S. M. Barnett, D. T. Pegg, and J. Jeffers, Bayes' theorem and quantum retrodiction, J. Mod. Opt. **47**, 1779 (2000); D. T. Pegg, S. M. Barnett, and J. Jeffers, Quantum retrodiction in open systems, Phys. Rev. A **66**, 022106 (2002).

[6] A. Chantasri, I. Guevara, K. T. Laverick, and H. M. Wiseman, Unifying theory of quantum state estimation using past and future information, Phys. Rep. **930**, 1 (2021).

[7] M. S. Leifer and R. W. Spekkens, Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference, Phys. Rev. A **88**, 052130 (2013).

[8] D. Horsman, C. Heunen, M. F. Pusey, J. Barrett, and R. W. Spekkens, Can a quantum state over time resemble a quantum state at a single time? Proc. R. Soc. London, Ser. A **473**, 20170395 (2017).

[9] R. E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. **82**, 35 (1960).

[10] S. Särkkä, *Bayesian Filtering and Smoothing* (Cambridge University Press, Cambridge, 2013).

[11] V. P. Belavkin, Nondemolition principle of quantum measurement theory, Found. Phys. **24**, 685 (1994).

[12] L. Bouten, R. Van Handel, and M. James, An introduction to quantum filtering, SIAM J. Control Optim. **46**, 2199 (2007).

[13] M. Tsang, Time-Symmetric Quantum Theory of Smoothing, Phys. Rev. Lett. **102**, 250403 (2009).

[14] M. Tsang, Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing, Phys. Rev. A **80**, 033840 (2009); Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing. II. Applications to atomic magnetometry and Hardy's paradox, **81**, 013824 (2010); A Bayesian quasi-probability approach to inferring the past of quantum observables, arXiv:1403.3353.

[15] K. Ohki, A smoothing theory for open quantum systems: The least mean square approach, in *2015 54th IEEE Conference on Decision and Control (CDC)* (IEEE, Osaka, Japan, 2015) pp. 4350–4355; An invitation to quantum filtering and smoothing theory based on two inner products, in *2015 54th IEEE Conference on Decision and Control (CDC)* (Kyoto University, Kyoto, 2018), Vol. 2018 pp. 18–44.

[16] L. Accardi and C. Cecchini, Conditional expectations in von Neumann algebras and a theorem of Takesaki, J. Funct. Anal. **45**, 245 (1982).

[17] D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, Commun. Math. Phys. **105**, 123 (1986).

[18] D. Petz, Conditional expectation in quantum probability, in *Quantum Probability and Applications III*, edited by Luigi Accardi and Wilhelm von Waldenfels (Springer, Berlin, Heidelberg, 1988), pp. 251–260.

[19] A. Jenčová, Quantum hypothesis testing and sufficient subalgebras, Lett. Math. Phys. **93**, 15 (2010).

[20] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*, 2nd ed. (Springer, Berlin, 2017).

[21] S. Gammelmark, B. Julsgaard, and K. Mølmer, Past Quantum States of a Monitored System, Phys. Rev. Lett. **111**, 160401 (2013).

[22] I. Guevara and H. Wiseman, Quantum State Smoothing, Phys. Rev. Lett. **115**, 180407 (2015).

[23] S. Personick, Application of quantum estimation theory to analog communication over quantum channels, IEEE Trans. Inf. Theory **17**, 240 (1971).

[24] V. P. Belavkin and B. A. Grishanin, Optimum estimation in quantum channels by the generalized Heisenberg inequality method, Prob. Peredachi Inf. **9**, 44 (1973).

[25] K. R. Parthasarathy, *Introduction to Probability and Measure* (Hindustan Book Agency, New Delhi, 2005).

[26] P. Billingsley, *Probability and Measure* (Wiley, New York, 1995).

[27] C. W. Gardiner and P. Zoller, *Quantum Noise* (Springer-Verlag, Berlin, 2004).

[28] Anonymous (Private communications).

[29] S. Kocsis, B. Braverman, S. Ravets, M. J. Stevens, R. P. Mirin, L. K. Shalm, and A. M. Steinberg, Observing the average trajectories of single photons in a two-slit interferometer, Science **332**, 1170 (2011).

[30] D. Petz and G. Toth, The Bogoliubov inner product in quantum statistics, Lett. Math. Phys. **27**, 205 (1993).

[31] J. Hamhalter, *Quantum Measure Theory* (Springer, Dordrecht, 2003).

[32] A. S. Holevo, Commutation superoperator of a state and its applications to the noncommutative statistics, Rep. Math. Phys. **12**, 251 (1977).

[33] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, Pisa, 2011).

[34] J. Dressel, Weak values as interference phenomena, Phys. Rev. A **91**, 032116 (2015).

[35] D. Petz, Monotone metrics on matrix spaces, Lin. Alg. Appl. **244**, 81 (1996).

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

[37] M. Tsang, Quantum limit to subdiffraction incoherent optical imaging. II. A parametric-submodel approach, Phys. Rev. A **104**, 052411 (2021).

[38] H. P. Yuen and M. Lax, Multiple-parameter quantum estimation and measurement of nonselfadjoint observables, IEEE Trans. Inf. Theory **19**, 740 (1973).

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

[40] V. Petersen and K. Mølmer, Estimation of fluctuating magnetic fields by an atomic magnetometer, Phys. Rev. A **74**, 043802 (2006).

[41] M. Tsang, H. M. Wiseman, and C. M. Caves, Fundamental Quantum Limit to Waveform Estimation, Phys. Rev. Lett. **106**, 090401 (2011).

[42] T. A. Wheatley, D. W. Berry, H. Yonezawa, D. Nakane, H. Arao, D. T. Pope, T. C. Ralph, H. M. Wiseman, A. Furusawa, and E. H. Huntington, Adaptive Optical Phase Estimation Using Time-Symmetric Quantum Smoothing, Phys. Rev. Lett. **104**, 093601 (2010); H. Yonezawa, D. Nakane, T. A. Wheatley, K. Iwasawa, S. Takeda, H. Arao, K. Ohki, K. Tsumura, D. W. Berry, T. C. Ralph, H. M. Wiseman, E. H. Huntington, and A. Furusawa, Quantum-enhanced optical-phase tracking, Science **337**, 1514 (2012); K. Iwasawa, K. Makino, H. Yonezawa, M. Tsang, A. Davidovic, E. Huntington, and A. Furusawa, Quantum-Limited Mirror-Motion Estimation, Phys. Rev. Lett. **111**, 163602 (2013); T. A. Wheatley, M. Tsang, I. R. Petersen, and E. H. Huntington, Improved mirror position estimation using resonant quantum smoothing, EPJ Quantum Technol. **2**, 13 (2015).

[43] C. N. Madsen, L. Valdetaro, and K. Mølmer, Quantum estimation of a time-dependent perturbation, Phys. Rev. A **104**, 052621 (2021).

[44] *Quantum State Estimation*, edited by Matteo Paris and Jaroslav Rehacek (Springer-Verlag, Berlin, Heidelberg, 2004).

[45] K. Macieszczak, M. Fraas, and R. Demkowicz-Dobrzański, Bayesian quantum frequency estimation in presence of collective dephasing, New J. Phys. **16**, 113002 (2014).

[46] C. Ferrie, Quasi-probability representations of quantum theory with applications to quantum information science, Rep. Prog. Phys. **74**, 116001 (2011).

[47] A. S. Holevo, *Quantum Systems, Channels, Information*, 2nd ed. (de Gruyter, Berlin, 2019).

[48] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, Rev. Mod. Phys. **84**, 621 (2012).

[49] J. Gough and M. R. James, The series product and its application to quantum feedforward and feedback networks, IEEE Trans. Autom. Control **54**, 2530 (2009); M. Tsang and C. M. Caves, Evading Quantum Mechanics: Engineering a Classical Subsystem within a Quantum Environment, Phys. Rev. X **2**, 031016 (2012).

[50] M. Tsang, J. H. Shapiro, and S. Lloyd, Quantum theory of optical temporal phase and instantaneous frequency. II. Continuous-time limit and state-variable approach to phase-locked loop design, Phys. Rev. A **79**, 053843 (2009).