Quantum Risk-Sensitive Estimation and Robustness

Naoki Yamamoto and Luc Bouten

***Abstract*—This paper studies a quantum risk-sensitive estimation problem and investigates robustness properties of the filter. This is a direct extension to the quantum case of analogous classical results. All investigations are based on a discrete approximation model of the quantum system under consideration. This allows us to study the problem in a simple mathematical setting. We close the paper with some examples that demonstrate the robustness of the risk-sensitive estimator.**

***Index Terms*—Quantum filtering, quantum probability, quantum systems, risk-sensitive estimation, robustness.**

I. Introduction

FILTERING, which in a broad sense is a method for extracting information from a noisy signal, is one of the principal tools in modern engineering science. In particular, when considering a partially observed dynamical system, we can construct an optimal filter that computes the least square estimate of a state variable of the dynamics. In the linear case, this results in the so-called *Kalman filter* [28]. This dynamical filtering theory was rigorously established using the classical Kolmogorov probability theory and its application to the theory of stochastic differential equations (e.g., [29]). Moreover, it is well known as the *separation theorem* [44] that the solution of a general optimal control problem for a partially observed system can be represented in terms of a corresponding information state of the filter. For this reason, the filtering theory is not only important in itself, but also essential in feedback control theory.

The situation is much the same in quantum mechanics. The *Heisenberg uncertainty principle* shows that any quantum system must possess fundamental uncertainty originating from the noncommutativity of its random variables. Therefore, we can never have complete observation in the quantum setting, which implies the necessity of filtering in the quantum case. Fortunately, there exists a *quantum filtering theory* as a beautiful parallel to the classical one. The theory was pioneered by Belavkin in the remarkable papers [4], [5], [6], and the *quantum filtering equation* or *stochastic master equation* is now widely used in the physics community [1], [8], [16], [21], [31], [38], [40], [47]. Moreover, as in the classical theory, it is possible to show that a separation principle holds in the quantum case [10].

Manuscript received September 29, 2007; revised March 07, 2008. Current version published January 14, 2009. This work was supported by the JSPS Grant-in-Aid 06693 and the ARO under Grant W911NF-06-1-0378. Recommended by Associate Editor J.-F. Zhang.

N. Yamamoto is with the Department of Applied Physics and Physico-Informatics, Keio University, Yokohama 223-8522, Japan (e-mail: yamamoto@appi.keio.ac.jp).

L. Bouten is with the Physical Measurement and Control Department, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: bouten@its.caltech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2008.2009571

The filtering for both classical and quantum cases is, as mentioned above, clearly an important tool in control theory. However, we have to point out that the optimal filter is in general quite fragile to unmodeled uncertainty of the system, and consequently the optimal estimation can be largely violated. This fact requires us to develop a theory of *robust estimation* that allows some model uncertainties and guarantees high-quality estimation performance. *Guaranteed-cost filtering* [34], [45] is one such robust estimation method in the classical theory. It guarantees that the variance of the estimation error is within a certain bound even when the linear system under consideration includes unknown parameters. Moreover, *risk-sensitive filtering* [13], [15], [32], [36] is known as a very efficient robust estimation method, for a wide class of classical linear and nonlinear systems [7], [42], [48]. Recently, one of the authors has obtained a quantum version of the guaranteed-cost filter mentioned above [46]. In this paper, we develop a quantum risk-sensitive estimation theory.

Let us first briefly introduce the classical theory of risk-sensitive estimation.

A. Classical Risk-Sensitive Estimation

We are given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a signal model of a discrete time system

$$x_l = a(x_{l-1}) + b(x_{l-1})w_l, \quad y_l = c(x_{l-1}) + v_l$$

where x_l is the signal state, y_l is the output, and w_l, v_l are *i.i.d.* random Gaussian processes. A version of the risk-sensitive estimator of x_l is defined as

$$\hat{x}_l^\mu := \operatorname*{argmin}_{z_l \in \mathcal{Y}_l} \mathbf{E}_{\mathbf{P}}[\Psi(z_l)]$$
$$\Psi(z_l) = \exp\left[\mu_1 \sum_{i=1}^{l-1} (x_i - \hat{x}_i^\mu)^2 + \mu_2 (x_l - z_l)^2\right] \tag{1}$$

where $\mathcal{Y}_l = \sigma\{y_i : 1 \le i \le l\}$ is the σ-algebra generated from the observation y_i, and $\mu = (\mu_1, \mu_2)$ are the weighting constants called the *risk-sensitive parameters*. Moreover, we use the notation $z_l \in \mathcal{Y}_l$ to indicate that z_l is a bounded $\mathcal{Y}_l$-measurable function. The risk-sensitive estimator (1) can be represented by $\hat{x}_l^\mu = \operatorname{argmin}_{z_l \in \mathcal{Y}_l} f_1(z_l, \alpha_l^\mu)$, where f_1 is a certain function and $\alpha_l^\mu(x)$ is an *information state* defined by

$$\mathbf{E}_{\mathbf{Q}}\left[\Lambda_l \zeta(x_l) \mathrm{e}^{\mu_1 \Sigma_{i=1}^{l-1} (x_i - \hat{x}_i^\mu)^2} \middle| \mathcal{Y}_l\right] = \int_{\mathbb{R}} \zeta(x) \alpha_l^\mu(x) dx$$

for all test functions $\zeta(x)$. Here, $\mathbf{Q}$ is a probability measure defined by

$$\Lambda_l := \frac{d\mathbf{P}}{d\mathbf{Q}} = \prod_{i=0}^{l} \exp\left[c(x_i) y_i - \frac{1}{2} c(x_i)^2\right].$$

Moreover, $\alpha_l^\mu(x)$ satisfies a recursive equation of the form $\alpha_l^\mu = f_2\left(\alpha_{l-1}^\mu, \hat{x}_{l-1}^\mu, y_l\right)$. Hence, running this equation with the measurement data y_l, we can recursively calculate $f_1\left(z_l, \alpha_l^\mu\right)$ and obtain the minimizer of this function, i.e., $\hat{x}_l^\mu$.

Note that $\hat{x}_l^\mu$ differs from the standard optimal (or *risk-neutral*) estimator $\hat{x}_l := \operatorname{argmin}_{z_l \in \mathcal{Y}_l} \mathbf{E}_{\mathbf{P}}[(x_l - z_l)^2]$ and is thus not optimal in the sense of the mean square error. However, the risk-sensitive estimator certainly has a great advantage over the risk-neutral one when we consider an uncertain system. This can be seen as follows. If the true probability measure $\mathbf{P}_{\text{true}}$ is unknown, then we need to use a known nominal measure $\mathbf{P}_{\text{nom}}$ and design a nominal filter based on $\mathbf{P}_{\text{nom}}$. However, since $\mathbf{P}_{\text{true}} \neq \mathbf{P}_{\text{nom}}$, there is no guarantee that the nominal estimator $\hat{x}_l^{\text{nom}}$ yields a bounded estimation error. The risk-sensitive estimator overcomes this issue. That is, the nominal risk-sensitive estimator $\hat{x}_l^{\mu,\text{nom}}$ (i.e., based on $\mathbf{P}_{\text{nom}}$) satisfies

$$\mathbf{E}_{\mathbf{P}_{\text{true}}}\left[\mu_1 \sum_{i=1}^{l-1} (x_i - \hat{x}_i^{\mu,\text{nom}})^2 + \mu_2 (x_l - \hat{x}_l^{\mu,\text{nom}})^2\right] \leq \log \mathbf{E}_{\mathbf{P}_{\text{nom}}}\left[\Psi(\hat{x}_l^{\mu,\text{nom}})\right] + R^c(\mathbf{P}_{\text{true}} \| \mathbf{P}_{\text{nom}}) \quad (2)$$

where $R^c(\mathbf{Q}\|\mathbf{P}) := \int \log(d\mathbf{Q}/d\mathbf{P}) d\mathbf{Q}$ is the classical relative entropy of $\mathbf{Q}$ and $\mathbf{P}$. Equation (2) implies that the unknown true estimation error is bounded if $R^c(\mathbf{P}_{\text{true}}\|\mathbf{P}_{\text{nom}})$ is finite. This robustness property is derived using the following *duality relation* (e.g., [17]) of two measures $\mathbf{P}$ and $\mathbf{Q}$:

$$\log \mathbf{E}_{\mathbf{P}}[\mathrm{e}^{\psi}] = \sup_{\mathbf{Q}}\left[\mathbf{E}_{\mathbf{Q}}(\psi) - R^c(\mathbf{Q}\|\mathbf{P}) : \mathbf{Q} \ll \mathbf{P}\right] \quad (3)$$

where $\mathbf{Q} \ll \mathbf{P}$ means that $\mathbf{Q}$ is absolutely continuous to $\mathbf{P}$.

B. Organization of the Paper

This paper provides a quantum version of the risk-sensitive estimation method presented above and shows its robustness properties against system uncertainty. The systems we consider are taken from quantum optics and consist of a quantum system in interaction with the quantized electromagnetic field. The field is described by a discretized model [9] that converges to a quantum stochastic dynamics [23] when the discretization step is taken to zero [2], [3], [11], [20], [30]. The discretized model has the advantage of being very tractable mathematically. The estimator is based on the risk-sensitive information state introduced by James [14], [24] in the context of quantum risk-sensitive control. We derive a bound on the estimation error in the presence of uncertainty. We illustrate the robustness of the estimator by simulations.

The paper is organized as follows. In Section II, we introduce quantum probability in a finite-dimensional context and a duality relation that will lead to the robustness property of the estimator. Section III is devoted to describe a discrete approximation model of the field. Section IV introduces the notion of composition of an operator and an operator valued function. In Section V we introduce the risk-sensitive estimator and derive the filter propagating the risk-sensitive information state. Section VI introduces a class of uncertain systems and derives a bound on the estimation error, showing robustness. In Section VII we present the results from simulations.

Notation

The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. For a classical probability measure $\mathbf{P}$, $\mathbf{E}_{\mathbf{P}}$ denotes the expectation with respect to $\mathbf{P}$. The commutator of two operators A and B is defined by $[A, B] := AB - BA$. I denotes the identity operator.

II. Quantum Probability Theory

A. Quantum Probability Space

In quantum mechanics, a random variable is represented by a linear self-adjoint operator on a Hilbert space. Due to the noncommutativity of such operators, we need to replace the conventional notion of a classical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ by the notion of a *quantum probability space* defined below.

Definition 2.1 ($*$-algebra): Let H be a finite-dimensional complex Hilbert space. A *$*$-algebra* $\mathcal{A}$ is a set of linear operators $\mathsf{H} \to \mathsf{H}$ such that $I, \alpha A + \beta B, AB, A^* \in \mathcal{A}$ for any $A, B \in \mathcal{A}$ and $\alpha, \beta \in \mathbb{C}$. $\mathcal{A}$ is called commutative if $[A, B] = 0$ for any $A, B \in \mathcal{A}$.

Recall that the adjoint of A is defined as the unique operator A^* that satisfies $\langle A^* x, y\rangle = \langle x, Ay\rangle, \forall\, x, y \in \mathsf{H}$.

Definition 2.2 (State): A *state* on $\mathcal{A}$ is a linear map $\mathbb{P} : \mathcal{A} \to \mathbb{C}$ that is positive $\mathbb{P}(A^*A) \geq 0, \forall\ A \in \mathcal{A}$ and normalized $\mathbb{P}(I) = 1$.

Let d be the dimension of H. Let $(e_1, \ldots, e_d)$ be an orthonormal basis of H. The *trace* is the state defined by $\operatorname{Tr}(A) = \sum_{i=1}^d \langle e_i, Ae_i\rangle$ for all $A \in \mathcal{A}$. It is well known that this definition does not depend on the basis.

Definition 2.3 (Quantum Probability Space): Let $\mathcal{A}$ be a $*$-algebra of operators on a finite-dimensional complex Hilbert space H and $\mathbb{P}$ be a state on $\mathcal{A}$. Then, $(\mathcal{A}, \mathbb{P})$ is called a (finite-dimensional) quantum probability space.

Let $(\mathcal{A}, \mathbb{P})$ be a quantum probability space. A self-adjoint element of $\mathcal{A}$ is called a *quantum random variable* or *observable*. If $\mathcal{A}$ is a commutative $*$-algebra, then we call $(\mathcal{A}, \mathbb{P})$ a commutative quantum probability space. In this case, all quantum random variables in $\mathcal{A}$ commute with each other, which is the same as in the classical case. It is therefore not surprising that a commutative quantum probability space is equivalent to a classical one. A formal statement of this assertion is provided by the well known *spectral theorem* (Theorem 2.1 below). Note that in the finite-dimensional setting of this article the spectral theorem follows trivially from diagonalizing the operators in $\mathcal{A}$ (see the proof of Theorem 2.1 below). In an infinite-dimensional setting an analogous result, which is closely related to Gelfand's Theorem for commutative C^*-algebras (see e.g., [35]), is true.

Definition 2.4 ($*$-isomorphism): Let Ω be a set and let $\mathcal{F}$ be a σ-algebra on Ω. A *$*$-isomorphism* between a commutative $*$-algebra $\mathcal{C}$ and the set of bounded $\mathcal{F}$-measurable functions $\ell^\infty(\mathcal{F})$ on Ω is a linear bijection $\iota : \mathcal{C} \to \ell^\infty(\mathcal{F})$ such that $\iota(A^*)(i) = \iota(A)(i)^*$ and $\iota(AB)(i) = \iota(A)(i)\iota(B)(i)$ for all $A, B \in \mathcal{C}$ and $i \in \Omega$.

Theorem 2.1 (Spectral Theorem): Let $(\mathcal{C}, \mathbb{P})$ be a finite-dimensional commutative quantum probability space. Then there exists a classical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a $*$-isomorphism $\iota : \mathcal{C} \to \ell^\infty(\mathcal{F})$ such that $\mathbb{P}(A) = \mathbf{E}_{\mathbf{P}}[\iota(A)], \forall\ A \in \mathcal{C}$.

Proof: The theorem is proved by construction. First, let $\mathsf{H} = \mathbb{C}^n$ and $\Omega = \{1, \ldots, n\}$. Since $[A, A^*] = 0\ \forall A \in \mathcal{C}$, all

the elements in $\mathcal{C}$ can be diagonalized simultaneously. Hence, we can set $A = \mathrm{diag}\{a_1, \ldots, a_n\}$ and define a classical random variable $\iota(A) : \Omega \to \mathbb{C}$ by $\iota(A)(i) = a_i$. Let P be a projection in $\mathcal{C}$, i.e., $P = P^* = P^2$, then $\iota(P)$ is the indicator function of a subset S_P of Ω. We define $\mathcal{F}$ as the set of subsets S_P of Ω where P runs through the projections in $\mathcal{C}$. Furthermore, we define a probability measure $\mathbf{P}$ on $\mathcal{F}$ by $\mathbf{P}(S_P) = \mathbb{P}(P)$, $\forall P \in \mathcal{C}$. As a result, we have constructed a classical probability space $(\Omega, \mathcal{F}, \mathbf{P})$. It is easy to verify $\mathbf{E}_{\mathbf{P}}[\iota(A)] = \mathbb{P}(A)$. ■

Note here that any observable $A = A^* \in \mathcal{A}$ is an element of the commutative $*$-subalgebra $\mathcal{C} \subset \mathcal{A}$ generated by A itself. Using the spectral theorem we see that we can always realize an observable A as a classical random variable $\iota(A)$ on a classical probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where the measure $\mathbf{P}$ is given by the state. If we perform a *measurement* of A, we obtain one of the values that $\iota(A)$ can take, distributed according to $\mathbf{P}$. Note that if two observables do not commute with each other, then we cannot represent them both as classical random variables on the same probability space. Such observables are called *incompatible*, they cannot both be measured in a single realization of an experiment.

Example 2.1 (Quantum Two-Level System): Let $\mathsf{H} = \mathbb{C}^2$ and let $\mathcal{M}$ be the $*$-algebra of 2×2 complex matrices. Moreover, let ψ be a state on $\mathcal{M}$. With the quantum probability space $(\mathcal{M}, \psi)$ we can model a *two-level system*. The state ψ can be written as $\psi(X) = \mathrm{Tr}(\rho A)$, $\forall\, A \in \mathcal{M}$ for some matrix ρ that is positive and normalized (i.e., $\mathrm{Tr}(\rho) = 1$). Let us now consider a commutative $*$-subalgebra $\mathcal{D} = \{D = \mathrm{diag}\{d_1, d_2\} : d_1, d_2 \in \mathbb{R}\} \subset \mathcal{M}$. From Theorem 2.1, we can construct a classical probability space that is in one-to-one correspondence with $(\mathcal{D}, \psi)$. The sample space is $\Omega = \{1, 2\}$, and the set of events is $\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \Omega\}$. A classical random variable $\iota(D)$ is then defined through $\iota(D)(1) = d_1$ and $\iota(D)(2) = d_2$. Now, $D \in \mathcal{D}$ has a spectral decomposition $D = \sum d_i P_i$ with the projection matrices $P_1 = \mathrm{diag}\{1, 0\}$ and $P_2 = \mathrm{diag}\{0, 1\}$, which yield classical indicator functions $\chi_{\{1\}} = \iota(P_1)$ and $\chi_{\{2\}} = \iota(P_2)$. Hence, the probability distribution of $\iota(D)$ is given by $\mathbf{P}(\{1\}) = \psi(P_1) = \mathrm{Tr}(\rho P_1) = \rho_{11}$ and $\mathbf{P}(\{2\}) = \rho_{22}$.

Let $(\mathcal{A}_1, \mathbb{P}_1)$ and $(\mathcal{A}_2, \mathbb{P}_2)$ be two quantum probability spaces, defined on the Hilbert spaces H_1 and H_2, respectively. We will now introduce the composite quantum probability space $(\mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$. Let $a_1 \otimes a_2$ be the tensor (Kronecker) product of two vectors $a_1 \in \mathsf{H}_1$ and $a_2 \in \mathsf{H}_2$. Introducing an inner product $\langle a_1 \otimes a_2, b_1 \otimes b_2 \rangle := \langle a_1, b_1 \rangle \langle a_2, b_2 \rangle$, we have a Hilbert space $\mathsf{H}_1 \otimes \mathsf{H}_2$. The composite quantum probability space $(\mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ is then defined on $\mathsf{H}_1 \otimes \mathsf{H}_2$ as follows. First, we define an element $A_1 \otimes A_2 \in \mathcal{A}_1 \otimes \mathcal{A}_2$ through the relation $(A_1 \otimes A_2)(a_1 \otimes a_2) = A_1 a_1 \otimes A_2 a_2$. Any element of $\mathcal{A}_1 \otimes \mathcal{A}_2$ is given as a linear combination of such elements. Second, the state $\mathbb{P}_1 \otimes \mathbb{P}_2$ is defined by $(\mathbb{P}_1 \otimes \mathbb{P}_2)(A_1 \otimes A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$.

B. Conditional Expectation

Let $(\mathcal{A}, \mathbb{P})$ be a quantum probability space. Let A and B be two commuting self-adjoint elements of $\mathcal{A}$. Using Theorem 2.1 we can present A and B as classical random variables $\iota(A)$ and $\iota(B)$ on a classical probability space $(\Omega, \mathcal{F}, \mathbf{P})$. This allows us to form the classical conditional expectation $\mathbf{E}_{\mathbf{P}}[\iota(A)|\iota(B)]$. The *quantum conditional expectation* $\mathbb{P}(A|B)$ can then be defined as its pull-back

$$\mathbb{P}(A|B) = \iota^{-1}\left(\mathbf{E}_{\mathbf{P}}[\iota(A)|\iota(B)]\right).$$

Now suppose that instead of the operator B, we want to condition A on a commutative $*$-subalgebra $\mathcal{C}$ of $\mathcal{A}$. As long as A commutes with every element in $\mathcal{C}$, we can apply the spectral theorem to the commutative $*$-algebra generated by $\mathcal{C}$ and A together, and define

$$\mathbb{P}(A|\mathcal{C}) = \iota^{-1}\left(\mathbf{E}_{\mathbf{P}}[\iota(A)|\sigma(\iota(\mathcal{C}))]\right) \tag{4}$$

where $\sigma(\iota(\mathcal{C}))$ stands for the classical σ-algebra generated by $\iota(K)$, $K \in \mathcal{C}$. This shows that given a commutative $*$-subalgebra $\mathcal{C}$, we can define the quantum conditional expectation onto $\mathcal{C}$ for every self-adjoint element in the *commutant* of $\mathcal{C}$. Here the commutant of $\mathcal{C}$ is given by

$$\mathcal{C}' := \{A \in \mathcal{A} : [A, C] = 0,\ \forall\, C \in \mathcal{C}\}.$$

The formal definition of the conditional expectation follows below. It coincides with the standard definition of the conditional expectation for operator algebras [41], [43] for the situation we are interested in. Note, however, that our definition is more restrictive since we only allow for conditional expectations from the commutant of a commutative algebra $\mathcal{C}$ onto $\mathcal{C}$ itself.

Definition 2.5 (Quantum Conditional Expectation): Let $(\mathcal{A}, \mathbb{P})$ be a quantum probability space, and let $\mathcal{C}$ be a commutative $*$-subalgebra of $\mathcal{A}$. Then the map $\mathbb{P}(\cdot|\mathcal{C}) : \mathcal{C}' \to \mathcal{C}$ is called (a version of) the quantum conditional expectation from $\mathcal{C}'$ to $\mathcal{C}$ if $\mathbb{P}(\mathbb{P}(A|\mathcal{C})K) = \mathbb{P}(AK)$, $\forall\, A \in \mathcal{C}'$, $\forall\, K \in \mathcal{C}$.

Note that for every self-adjoint element $A \in \mathcal{C}'$, we have given an explicit expression for the quantum conditional expectation in (4). Every element A in the commutant can be written in a unique way as $A = A_1 + iA_2$ with A_1 and A_2 self-adjoint. If we define the conditional expectation of A onto $\mathcal{C}$ by linear extension of the definition (4), then it is easy to see that it satisfies the formal definition given in Definition 2.5. This means we have shown existence of the quantum conditional expectation as defined in Definition 2.5.

Finally, we remark some basic properties that both the classical and quantum conditional expectations satisfy: (i) $\mathbb{P}(A|\mathcal{C})$ is unique with probability one, (ii) $\mathbb{P}(\mathbb{P}(A|\mathcal{C})) = \mathbb{P}(A)$, (iii) $\mathbb{P}(CA|\mathcal{C}) = C\mathbb{P}(A|\mathcal{C})$ if $C \in \mathcal{C}$ and $A \in \mathcal{C}'$ (module property), and (iv) $\mathbb{P}(\mathbb{P}(A|\mathcal{B})|\mathcal{C}) = \mathbb{P}(A|\mathcal{C})$ if $\mathcal{C} \subset \mathcal{B}$ (tower property). Note that it easily follows from the tower property that $\mathbb{P}(\cdot|\mathcal{C})$ is idempotent, i.e., it is a projection. Moreover, similar to the classical case, $\mathbb{P}(A|\mathcal{C})$ is the least mean square estimate of A given $\mathcal{C}$, i.e.

$$\begin{aligned} ||A - \mathbb{P}(A|\mathcal{C})||_{\mathbb{P}}^2 &\leq ||A - \mathbb{P}(A|\mathcal{C})||_{\mathbb{P}}^2 + ||\mathbb{P}(A|\mathcal{C}) - B||_{\mathbb{P}}^2 \\ &= ||A - B||_{\mathbb{P}}^2, \quad \forall\, B \in \mathcal{C} \end{aligned} \tag{5}$$

where we have defined $||X||_{\mathbb{P}}^2 := \mathbb{P}(X^*X)$.

C. Density Matrix and Quantum Relative Entropy

In Example 2.1, we have seen that the state $\mathbb{P}$ can be represented in terms of a matrix ρ. In the finite-dimensional case we can always find a unique *density matrix* ρ that satisfies

$$\mathbb{P}(A) = \mathrm{Tr}(\rho A), \quad \rho = \rho^* \geq 0, \quad \mathrm{Tr}\rho = 1.$$

The latter two conditions guarantee $\mathbb{P}(A^*A) \geq 0, \ \forall\, A \in \mathcal{A}$ and $\mathbb{P}(I) = 1$, respectively. In particular, when ρ is a rank-one projection matrix $\rho = bb^*$, $b \in \mathsf{H}$, then $\mathbb{P}(A)$ is expressed as

$$\mathbb{P}(A) = \mathrm{Tr}(bb^*A) = \langle b, Ab\rangle. \tag{6}$$

In analogy to the classical relative entropy, which has been introduced in Section I, we can define the *quantum relative entropy* of two states in terms of their density matrices as

$$R(\rho\|\rho') := \mathrm{Tr}\big[\rho(\log\rho - \log\rho')\big] \ \text{if}\ \mathrm{supp}\rho \subseteq \mathrm{supp}\rho' \tag{7}$$

where $\mathrm{supp}\rho$ represents the linear space spanned by the eigenvectors of ρ [33]. If $\mathrm{supp}\rho \nsubseteq \mathrm{supp}\rho'$, then $R(\rho\|\rho') := +\infty$. A quantum version of the duality relation (3) is given as follows.

Lemma 2.1 (Duality, See e.g., [33] Prop. 1.11): For any observable $A \in \mathcal{A}$ and density matrices ρ and ρ', the following relation holds:

$$\log \mathrm{Tr}(\mathrm{e}^{A+\log\rho'}) = \max_{\rho}\Big[\mathrm{Tr}(\rho A) - R(\rho\|\rho') : \mathrm{supp}\rho \subseteq \mathrm{supp}\rho'\Big]. \tag{8}$$

Proof: The proof is straightforward. Defining a density matrix $\rho_o = \mathrm{e}^{A+\log\rho'}/\mathrm{Tr}[\mathrm{e}^{A+\log\rho'}]$, we obtain

$$\mathrm{Tr}(\rho A) - R(\rho\|\rho') = \log \mathrm{Tr}(\mathrm{e}^{A+\log\rho'}) - R(\rho\|\rho_o).$$

Then, as the quantum relative entropy $R(\rho\|\rho_o)$ is always nonnegative and takes zero only when $\rho = \rho_o$, we observe that (8) holds and the maximum is attained only when $\rho = \rho_o$. ∎

We can derive a relaxed form of (8), expressed in terms of the corresponding states. From the Golden–Thompson inequality $\mathrm{Tr}(\mathrm{e}^{A+B}) \leq \mathrm{Tr}(\mathrm{e}^A\mathrm{e}^B)$ (see [19], [39]) with A, B self-adjoint, we have

$$\log \mathrm{Tr}(\mathrm{e}^{A+\log\rho'}) \leq \log \mathrm{Tr}(\mathrm{e}^A\mathrm{e}^{\log\rho'}) = \log \mathrm{Tr}(\mathrm{e}^A\rho').$$

Therefore, denoting the states corresponding to ρ and ρ' by $\mathbb{P}$ and $\mathbb{P}'$ respectively, we have

$$\mathbb{P}(A) \leq \log \mathbb{P}'(\mathrm{e}^A) + R(\rho\|\rho'). \tag{9}$$

This inequality will be used to show robustness properties of the quantum risk-sensitive estimator.

III. Discrete Field and Quantum Filtering

In this paper, we restrict ourselves to a system that consists of a two-level atom in interaction with the quantized electromagnetic field. This is merely for reasons of convenience, the theory can easily be extended to a large class of systems in interaction with the field. In this Section, we first introduce a discrete model for the electromagnetic field (see [9] and the references therein). Second, we describe the interaction between the atomic system and the field. Due to the interaction, the field carries off information about the system. In this way, by measuring the field, we can perform a noisy observation of the system. Finally, using quantum filtering theory we form optimal estimates of the atom observables. The quantum filtering equation recursively propagates these estimates in time.

A. Discrete Field

We first describe the quantum probability space with which we model the electromagnetic field in a discrete manner. Imagine a 1-D field traveling towards a photo detector. We divide the field into N time slices of length λ^2. The total measurement time is $T = N\lambda^2$. If N is large enough, the photo detector detects either zero or one photon in each time interval. Therefore, if N is large, each slice of the field can in good approximation be regarded as a two-level system $(\mathcal{M}, \phi)$, see Example 2.1. The *vacuum state* ϕ on $\mathcal{M}$ is given by $\phi(X) = \langle\Phi, X\Phi\rangle$, where $\Phi = [0\ 1]^{\mathsf{T}}$ denotes the so-called *vacuum vector*. The field can then be constructed as the N-fold tensor product of two-level systems representing the different time slices, i.e., $(\mathcal{W}_N, \phi^{\otimes N}) = (\mathcal{M}^{\otimes N}, \phi^{\otimes N})$. In particular, we assume that the system that interacts with the field is a two-level atomic system $(\mathcal{M}, \psi)$, i.e., the total quantum probability space for system and field together is given by

$$(\mathcal{M}\otimes\mathcal{W}_N, \mathbb{P}) = (\mathcal{M}\otimes\mathcal{M}^{\otimes N}, \psi\otimes\phi^{\otimes N}).$$

Let ρ be the density matrix corresponding to ψ. Then, $\mathbb{P}(X)$ can be written as $\mathbb{P}(X) = \mathrm{Tr}[X(\rho\otimes(\Phi\Phi^*)^{\otimes N})]$ for all $X \in \mathcal{M}\otimes\mathcal{W}_N$.

Next, we introduce discrete noises. To this end, we define

$$X_l := I^{\otimes(l-1)}\otimes X\otimes I^{\otimes(N-l)} \in \mathcal{W}_N, \quad 1 \leq l \leq N$$

where X is a 2×2 matrix and I is the 2×2 identity matrix. Using the above notation, let us define the following *noise matrices*:

$$\Delta A(l) = \lambda(\sigma_-)_l, \quad \Delta A(l)^* = \lambda(\sigma_+)_l$$
$$\Delta\Lambda(l) = (\sigma_+\sigma_-)_l, \quad \Delta t(l) = \lambda^2 I_l$$

where

$$\sigma_- = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \sigma_+ = \sigma_-^* = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{10}$$

Furthermore, we define the following so-called *fundamental noises* living in the first l slices:

$$A(l) = \sum_{i=1}^{l}\Delta A(i), \quad A(l)^* = \sum_{i=1}^{l}\Delta A(i)^*$$
$$\Lambda(l) = \sum_{i=1}^{l}\Delta\Lambda(i), \quad t(l) = \sum_{i=1}^{l}\Delta t(i)$$

with the convention $A(0) = A(0)^* = \Lambda(0) = t(0) = 0$. We now provide the following physical interpretation to the fundamental noises. First, $t_l := \iota(t(l))$ always takes the value $l\lambda^2 =$

$(l/N)T$, and thus, we may regard $t(l)$ as the time. Second, since $\Delta\lambda_l := \iota(\Delta\Lambda(l))$ takes either the value 0 or the value 1 at time l, it is reasonable to interpret $\Lambda(l)$ as the total number of photons counted by the photo detector up to time l. For the vacuum state, we have $\mathrm{Prob}(\Delta\lambda_l = 1) = \phi^{\otimes N}(\mathrm{diag}\{1,0\}_l) = 0$, which implies that the photo detector detects no photons. Finally, with regard to $A(l)$ and $A(l)^*$, we introduce an observable $\Delta W(l) := \Delta A(l) + \Delta A(l)^*$ and a commutative $*$-algebra *generated* from $\Delta W(i)$, $(0 \le i \le l)$

$$\mathcal{C}_l := \mathrm{alg}\{\Delta W(i) = \Delta A(i) + \Delta A(i)^* : 0 \le i \le l\}. \quad (11)$$

That is, $\mathcal{C}_l$ is the smallest $*$-algebra in $\mathcal{W}_l$ that contains $\Delta W(i)$, $(0 \le i \le l)$. Note that $\Delta W(l)$ has the following spectral decomposition:

$$\Delta W(l) = \Delta A(l) + \Delta A(l)^* = \begin{bmatrix} 0 & \lambda \\ \lambda & 0 \end{bmatrix}_l = \lambda P_l^+ + (-\lambda) P_l^-$$

with the projection matrices

$$P_l^+ := \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}_l, \quad P_l^- := \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}_l. \quad (12)$$

Thus, for the vacuum state, the classical random variable $\Delta w_l := \iota(\Delta W(l))$ takes $+\lambda$ with probability $\mathrm{Prob}(+\lambda) = \phi^{\otimes N}(P_l^+) = \langle \Phi^{\otimes N}, P_l^+ \Phi^{\otimes N}\rangle = 1/2$ or $-\lambda$ with probability $\mathrm{Prob}(-\lambda) = 1/2$ at each time. This implies that $\{w_l\}_{l=1,\dots,N}$ is a symmetric random walk. If we let N go to infinity and λ to 0, but keep the product $T = N\lambda^2$ constant, then it easily follows from Donsker's invariance principle (see e.g., [26]) that w_l converges weakly to a classical Brownian motion. Note that the relation $\Delta W(l)^2 = \Delta t(l)$ becomes $dw_t^2 = dt$ in the limit (see e.g., [37]). In physics the observable $A(l) + A(l)^*$ is known as a *field quadrature*, see e.g., [12], [18].

B. System-Field Interaction

Let H_1 and H_2 be Hilbert spaces with which we describe two separate quantum systems. The total interaction between these two systems over the first l time units can be described by a unitary transformation $U(l)$ that acts on the composite space $\mathsf{H}_1 \otimes \mathsf{H}_2$. The time evolution of an observable X of the composite system is given by the flow

$$j_l(X) = U(l)^* X U(l).$$

Suppose we start with an observable X that acts non-trivially only on the first system. At time l this observable is given by $j_l(X) = U(l)^*(X \otimes I)U(l)$ which in general will act nontrivially on both components in the tensor product $\mathsf{H}_1 \otimes \mathsf{H}_2$. This shows that information has been carried from the system that lives on H_1 to the system on H_2. Note that a unitary U can always be represented as $U = \mathrm{e}^{-iH}$ for some self-adjoint matrix H called the *Hamiltonian*.

In our model, a two-level atomic system repeatedly interacts with the slices of the field, i.e., $\mathsf{H}_1 = \mathbb{C}^2$ and $\mathsf{H}_2 = (\mathbb{C}^2)^{\otimes l}$. Let $H^{\mathrm{int}}(l) \in \mathcal{M} \otimes \mathcal{W}_l$ be the self-adjoint operator given by

$$H^{\mathrm{int}}(l) = j_{l-1}(L_1) \otimes \Delta\Lambda(l) + j_{l-1}(L_2) \otimes \Delta A(l)^* + j_{l-1}(L_2^*) \otimes \Delta A(l) + j_{l-1}(L_3) \otimes \Delta t(l) \quad (13)$$

where the L_i's are elements in $\mathcal{M}$ (for $i = 1, 2, 3$) such that L_1 and L_3 are self-adjoint. These system operators determine which kind of interaction between the two-level system and the field we are considering, i.e., they determine the physics of our problem. See Section VII for two examples: a dispersive interaction and spontaneous decay. We let $H^{\mathrm{int}}(l)$ be the Hamiltonian for the interaction between the system and the lth field slice, that is

$$U(l) = \overrightarrow{\prod_{i=1}^{l}} \mathrm{e}^{-iH^{\mathrm{int}}(i)} = \mathrm{e}^{-iH^{\mathrm{int}}(1)} \cdots \mathrm{e}^{-iH^{\mathrm{int}}(l)}$$
$$U(0) = I.$$

We define another unitary operator M_l by

$$M_l := \mathrm{e}^{-i\{L_1 \otimes \Delta\Lambda(l) + L_2 \otimes \Delta A^*(l) + L_2^* \otimes \Delta A(l) + L_3 \otimes \Delta t(l)\}}. \quad (14)$$

Since $\mathrm{e}^{-iH^{\mathrm{int}}(l)} = U(l-1)^* M_l U(l-1)$, the unitary operator $U(l)$ satisfies

$$U(l) = U(l-1)\mathrm{e}^{-iH^{\mathrm{int}}(l)} = M_l U(l-1). \quad (15)$$

The operator M_l acts nonidentically only on the system and the lth slice of the field. Thus, M_l can be expressed as

$$M_l := I + M^{\pm} \otimes \Delta\Lambda(l) + M^+ \otimes \Delta A(l)^* + M^- \otimes \Delta A(l) + M^{\circ} \otimes \Delta t(l), \quad (16)$$

for some system operators $M^i \in \mathcal{M}$, $(i = \pm, +, -, \circ)$, which are uniquely determined by L_i, $(i = 1, 2, 3)$. Note that the unitarity of M_l implies certain relations between the operators M^i, e.g., $M^{\circ} + M^{\circ *} + M^{+*}M^+ + \lambda^2 M^{\circ *} M^{\circ} = 0$. From now on, we will use M_l and M^i instead of $H^{\mathrm{int}}(l)$ and L_i to describe the interaction. We can write the following difference equation for the unitary $U(l)$:

$$\Delta U(l) = U(l) - U(l-1) = \Big[M^{\pm}\Delta\Lambda(l) + M^+\Delta A(l)^* + M^-\Delta A(l) + M^{\circ}\Delta t(l)\Big]U(l-1). \quad (17)$$

For simplicity we have omitted the tensor product $\otimes$ between M^i and the noise matrices. This rule will be applied throughout this paper. The dynamics (17) is called the *quantum stochastic difference equation*. It is a discrete version of the Hudson-Parthasarathy equation [23].

Next we describe a measurement performed on the field. Let us again consider the field observables $W(l) = A(l) + A(l)^*$, $(0 \le l \le N)$. After the interaction, these observables are given by

$$Y(l) := j_l(W(l)) = U(l)^*\Big[A(l) + A(l)^*\Big]U(l), \quad 0 \le l \le N. \quad (18)$$

The observation process $Y(l)$ satisfies the difference equation

$$\Delta Y(l) = U(l)^* \Big[\Delta A(l) + \Delta A(l)^* \Big] U(l) = j_l(\Delta W(l)). \quad (19)$$

Here we have used (15) and $[M_l, A(l-1)] = 0$. Moreover, using $[M_k, \Delta W(l)] = 0$, $(k \geq l+1)$ we find that

$$\Delta Y(l) = U(k)^* \Delta W(l) U(k) = j_k(\Delta W(l)), \quad (20)$$

for all $k \geq l$. Therefore we find

$$[\Delta Y(i), \Delta Y(j)] = 0, \quad \forall\, i, j.$$

This means that the algebra generated by the observations

$$\mathcal{Y}_l = \mathrm{alg}\{\Delta Y(i) : 0 \leq i \leq l\} \quad (21)$$

is a commutative $*$-algebra for all $0 \leq l \leq N$. This is called the *self-nondemolition* property of the observations $Y(l)$. Due to this property, we can define the classical process $\Delta y_l := \iota(\Delta Y(l))$, $(0 \leq l \leq N)$. This classical process represents the data that we obtain while running the measurement. Note that $\Delta Y(l)$ has the following spectral decomposition

$$\begin{aligned} \Delta Y(l) &= U(l)^* W(l) U(l) \\ &= \lambda U(l)^* P_l^+ U(l) + (-\lambda) U(l)^* P_l^- U(l) \end{aligned}$$

where the projection matrices P_l^+ and P_l^- are given by (12). Hence, from Theorem 2.1, the classical random variable $\Delta y_l = \iota(\Delta Y(l))$ takes $+\lambda$ with probability $\mathrm{Prob}(+\lambda) = \psi \otimes \phi^{\otimes N} \big(U(l)^* P_l^+ U(l) \big)$, which now depends on the interaction, or $-\lambda$ with probability $\mathrm{Prob}(-\lambda) = 1 - \mathrm{Prob}(+\lambda)$.

C. Quantum Filtering

The purpose of quantum filtering is to calculate the least mean square estimate of the observable $j_l(X) = U(l)^* X U(l)$ for a given system observable $X \in \mathcal{M}$. More specifically, we aim to find an element in the commutative $*$-algebra $\mathcal{Y}_l$ that minimizes the mean square error, i.e., $Z_l^{\mathrm{opt}} = \mathrm{argmin}_{Z_l \in \mathcal{Y}_l} \mathbb{P}\left(|j_l(X) - Z_l|^2 \right)$, where $|A|^2 := A^* A$ for an operator A on a Hilbert space. Note that (20) leads to the following *nondemolition* property

$$[j_l(X),\ \Delta Y(k)] = 0, \quad \forall\, l \geq k$$

which implies that $j_l(X) \in \mathcal{Y}_l'$, $\forall\, l$. Due to the self-nondemolition and nondemolition properties the quantum conditional expectation $\mathbb{P}(j_l(X)|\mathcal{Y}_l)$ exists. Moreover, in (5) we have seen that the quantum conditional expectation is an optimal estimator. Therefore Z_l^{opt} is given by $Z_l^{\mathrm{opt}} = \mathbb{P}(j_l(X)|\mathcal{Y}_l)$.

Since $\mathbb{P}(j_l(X)|\mathcal{Y}_l)$ is linear, positive with respect to X, and normalized, i.e., $\mathbb{P}(j_l(I)|\mathcal{Y}_l) = 1$, we can define an *information state* on the two-level atomic system by

$$\pi_l(X) := \mathbb{P}(j_l(X)|\mathcal{Y}_l), \quad X \in \mathcal{M}. \quad (22)$$

Note that the state π_l on $\mathcal{M}$ is stochastic, it depends on the observations up to time l. We are now going to derive a difference equation for $\pi_l(X)$, i.e., the quantum filter. The following *noncommutative Bayes formula* [10] is useful to derive the filter:

$$\pi_l(X) = \mathbb{P}(j_l(X)|\mathcal{Y}_l) = \frac{U(l)^* \mathbb{P}\left(V(l)^* X V(l) | \mathcal{C}_l\right) U(l)}{U(l)^* \mathbb{P}(V(l)^* V(l) | \mathcal{C}_l) U(l)}. \quad (23)$$

Here, $\mathcal{C}_l$ is the commutative $*$-algebra defined in (11) and $V(l)$ is the unique solution to the difference equation

$$\Delta V(l) = \Big[M^+ \Delta W(l) + M^\circ \Delta t(l) \Big] V(l-1), \quad V(0) = I. \quad (24)$$

We note that the conditional expectation in (23) is well defined, because $V(l)$ is driven by $\Delta t(l)$ and $\Delta W(l)$ and thus commutes with any element of $\mathcal{C}_l$. This means that $V(l)^* X V(l)$ is an element of $\mathcal{C}_l'$ for all system observables $X \in \mathcal{M}$. We now introduce an *unnormalized information state* σ_l by $\sigma_l(X) := U(l)^* \mathbb{P}\left(V(l)^* X V(l) | \mathcal{C}_l\right) U(l)$ for all $X \in \mathcal{M}$. Equation (23) now reads $\pi_l(X) = \sigma_l(X)/\sigma_l(I)$, which is a noncommutative analogue of the classical *Kallianpur-Striebel formula* [27]. It easily follows from (17) and (24) that $\sigma_l(X)$ satisfies the following difference equation:

$$\begin{aligned} \Delta\sigma_l(X) &= \sigma_{l-1}(\mathcal{L}(X)) \Delta t(l) + \sigma_{l-1}(\mathcal{J}(X)) \Delta Y(l) \\ \sigma_0 &= \psi \end{aligned} \quad (25)$$

where the operators $\mathcal{L}$ and $\mathcal{J}$ are given by

$$\begin{aligned} \mathcal{L}(X) &:= M^{+*} X M^+ + \lambda^2 M^{\circ *} X M^\circ + X M^\circ + M^{\circ *} X \\ \mathcal{J}(X) &:= \lambda^2 M^{+*} X M^\circ + \lambda^2 M^{\circ *} X M^+ \\ &\quad + X M^+ + M^{+*} X. \end{aligned} \quad (26)$$

The filter can now be obtained immediately from $\pi_l(X) = \sigma_l(X)/\sigma_l(I)$. We, however, will always use the unnormalized version of the filter given in (25).

Note that $\sigma_l(X)$, $\Delta t(l)$, and $\Delta Y(l)$ are all elements in the commutative $*$-algebra $\mathcal{Y}_l$. Due to Theorem 2.1 they can be diagonalized simultaneously, which yields the following classical random variables: $\iota(\sigma_l(X))$, $\Delta t_l = \iota(\Delta t(l)) = \lambda^2$, and $\Delta y_l = \iota(\Delta Y(l))$. Moreover, since $\iota(\sigma_l(X))$ is a linear and positive functional of X, we can define a 2×2 positive semidefinite matrix ϱ_l that satisfies $\iota(\sigma_l(X)) = \mathrm{Tr}(\varrho_l X)$. The unnormalized density matrix ϱ_l is called the *unnormalized information density matrix*. It is easy to derive a difference equation for ϱ_l:

$$\Delta\varrho_l = \bar{\mathcal{L}}(\varrho_{l-1}) \Delta t_l + \bar{\mathcal{J}}(\varrho_{l-1}) \Delta y_l, \quad \varrho_0 = \rho \quad (27)$$

where the operators $\bar{\mathcal{L}}$ and $\bar{\mathcal{J}}$ are given by

$$\bar{\mathcal{L}}(\varrho) := M^+ \varrho M^{+*} + \lambda^2 M^\circ \varrho M^{\circ *} + M^\circ \varrho + \varrho M^{\circ *} \quad (28)$$

$$\begin{aligned} \bar{\mathcal{J}}(\varrho) &:= \lambda^2 M^\circ \varrho M^{+*} + \lambda^2 M^+ \varrho M^{\circ *} \\ &\quad + M^+ \varrho + \varrho M^{+*}. \end{aligned} \quad (29)$$

IV. Composition of an Operator-Valued Function and an Observable

In Section V we will formulate risk-sensitive estimation as an optimal control problem for a given cost function, see (32) and (33) below. The idea of risk-sensitive control is to absorb

the running cost of the cost function into the dynamics, see (36) below. This means that the new dynamics depends on past estimates (the controls in the optimal control formulation of risk-sensitive estimation) which are a function of the observations thus far. Therefore we need to make precise mathematically what we mean by operator coefficients (see for example the coefficients in (37) below) that depend on a function of the observations thus far. We address this topic in this section.

Let $\mathcal{A}_1$ be a finite-dimensional $*$-algebra and let $\mathcal{A}_2$ be a commutative finite-dimensional $*$-algebra. Let K be an $\mathcal{A}_1$-valued function on $\mathbb{C}$, i.e., $K : \mathbb{C} \ni u \to K(u) \in \mathcal{A}_1$. Let $\mathfrak{u}$ be an element in $\mathcal{A}_2$. Note that since $\mathcal{A}_2$ is commutative, we have $\mathfrak{u}^*\mathfrak{u} = \mathfrak{u}\mathfrak{u}^*$, i.e., $\mathfrak{u}$ is *normal*. The spectral decomposition of $\mathfrak{u}$ can be written as $\mathfrak{u} = \sum_{x\in \mathrm{sp}(\mathfrak{u})} x P_{\mathfrak{u}}(x)$, where $\mathrm{sp}(\mathfrak{u})$ denotes the spectrum of $\mathfrak{u}$, i.e., the set of eigenvalues of $\mathfrak{u}$. The composition $K(\mathfrak{u}) \in \mathcal{A}_1 \otimes \mathcal{A}_2$ of K with $\mathfrak{u}$ is defined as [9]

$$K(\mathfrak{u}) := \sum_{x\in \mathrm{sp}(\mathfrak{u})} K(x) P_{\mathfrak{u}}(x) \in \mathcal{A}_1 \otimes \mathcal{A}_2. \tag{30}$$

This is a natural generalization of the composition of K with a classical random variable $\alpha : (\Omega, \mathcal{F}, \mathbf{P}) \to \mathbb{C}$, given by

$$K(\alpha)(\omega) := \sum_{x\in \mathrm{ran}(\alpha)} K(x) \chi_{\{\alpha = x\}}(\omega) \in \mathcal{A}_1. \tag{31}$$

Here $\mathrm{ran}(\alpha)$ denotes the range of α and $\chi_{\{\alpha=x\}}$ is the indicator function of the set $\{\omega \in \Omega : \alpha(\omega) = x\}$.

Let $\mathfrak{u}_l$ be an element of the observation algebra $\mathcal{Y}_l$, defined in (21). This means that we can write $\mathfrak{u}_l$ as a function of $\Delta Y(i)$, $(1 \le i \le l)$ by

$$\mathfrak{u}_l = f_l(\Delta Y(1), \ldots, \Delta Y(l)) \in \mathcal{Y}_l$$

for some function $f_l : \mathbb{R}^l \to \mathbb{C}$. Moreover, we can also write $\mathfrak{u}_l$ in terms of the observables $\Delta W(i) = \Delta A(i) + \Delta A(i)^*$, $(1 \le i \le l)$ as

$$\begin{aligned}\mathfrak{u}_l &= f_l(j_l(\Delta W(1)), \ldots, j_l(\Delta W(l))) \\ &= j_l(f_l(\Delta W(1), \ldots, \Delta W(l)))\end{aligned}$$

where we have used (20). Therefore, if we define an element $\check{\mathfrak{u}}_l$ in $\mathcal{C}_l$ by

$$\check{\mathfrak{u}}_l := f_l(\Delta W(1), \ldots, \Delta W(l)) \in \mathcal{C}_l$$

then $\mathfrak{u}_l$ can be written as $\mathfrak{u}_l = j_l(\check{\mathfrak{u}}_l)$. An $\mathcal{M}$-valued function $K : \mathbb{C} \ni u \to K(u) \in \mathcal{M}$ and an element $\check{\mathfrak{u}}_l$ in $\mathcal{C}_l$ give rise to the composition $K(\check{\mathfrak{u}}_l)$, which is an element in $\mathcal{M} \otimes \mathcal{C}_l$. Denoting the spectral decomposition of $\check{\mathfrak{u}}_l$ as $\check{\mathfrak{u}}_l = \sum_{x\in \mathrm{sp}(\check{\mathfrak{u}}_l)} x P_{\check{\mathfrak{u}}_l}(x)$, we obtain

$$\begin{aligned} j_l(K(\check{\mathfrak{u}}_l)) &= \sum_{x\in \mathrm{sp}(\check{\mathfrak{u}}_l)} U(l)^* K(x) U(l) U(l)^* P_{\check{\mathfrak{u}}_l}(x) U(l) \\ &= \sum_{x\in \mathrm{sp}(\check{\mathfrak{u}}_l)} j_l(K(x)) P_{U(l)^*\check{\mathfrak{u}}_l U(l)}(x) \\ &= \sum_{x\in \mathrm{sp}(\mathfrak{u}_l)} j_l(K(x)) P_{\mathfrak{u}_l}(x) =: j_l(K)(\mathfrak{u}_l)\end{aligned}$$

where we have introduced the notation $j_l(K)(\mathfrak{u}_l)$ in the last step. Note that $j_l(K)(\mathfrak{u}_l)$ is an element in $U(l)^*(\mathcal{M} \otimes \mathcal{C}_l)U(l)$.

V. Quantum Risk-Sensitive Filtering

In this section we study a quantum risk-sensitive estimation problem. Let X_{e} be a fixed element of the two-level atomic system $\mathcal{M}$. Then, the risk-sensitive estimator of $j_l(X_{\mathrm{e}})$ is defined as follows:

$$\left(\widehat{X_{\mathrm{e}}}^{\mu}(1), \ldots, \widehat{X_{\mathrm{e}}}^{\mu}(N)\right) := \operatorname*{argmin}_{\mathfrak{u}_1\in\mathcal{Y}_1, \ldots, \mathfrak{u}_N\in\mathcal{Y}_N} F(\mathfrak{u}_1, \ldots, \mathfrak{u}_N) \tag{32}$$

where the *cost function* F is given by

$$\begin{aligned} &F(\mathfrak{u}_1, \ldots, \mathfrak{u}_N) \\ &\quad := \mathbb{P}\left[R(N)^* \exp\left(\mu_2 |j_N(X_{\mathrm{e}}) - \mathfrak{u}_N|^2\right) R(N)\right], \end{aligned} \tag{33}$$

and the matrix $R(l) \in \mathcal{M} \otimes \mathcal{W}_{l-1}$ is given by

$$R(l) = \prod_{i=1}^{\overleftarrow{l-1}} \exp\left[\frac{\mu_1}{2}\lambda^2 |j_i(X_{\mathrm{e}}) - \mathfrak{u}_i|^2\right], \quad R(1) = R(0) = I. \tag{34}$$

Note again that $|A|^2 := A^*A$. Here, $\mu = (\mu_1, \mu_2)$ are weighting parameters that represent risk-sensitivity. Using the $\mathcal{M}$-valued function

$$K : \mathbb{C} \to \mathcal{M} : u \to K(u) = |X_{\mathrm{e}} - u|^2$$

we can write $K(\check{\mathfrak{u}}_l) = |X_{\mathrm{e}} - \check{\mathfrak{u}}_l|^2$ and $j_l(K)(\mathfrak{u}_l) = |j_l(X_{\mathrm{e}}) - \mathfrak{u}_l|^2$. Using these compositions, we can obtain a recursive form of $R(l)$:

$$\begin{aligned} R(l) &= \prod_{i=1}^{\overleftarrow{l-1}} \exp\left[\frac{\mu_1}{2}\lambda^2 j_i(K)(\mathfrak{u}_i)\right] \\ &= \prod_{i=1}^{\overleftarrow{l-1}} \exp\left[\frac{\mu_1}{2}\lambda^2 j_i(K(\check{\mathfrak{u}}_i))\right] \\ &= \exp\left[\frac{\mu_1}{2}\lambda^2 j_{l-1}(K(\check{\mathfrak{u}}_{l-1}))\right] R(l-1) \\ &= j_{l-1}\left(\mathrm{e}^{\mu_1\lambda^2 K(\check{\mathfrak{u}}_{l-1})/2}\right) R(l-1). \end{aligned} \tag{35}$$

Remark 5.1: If all matrices in (33) and (34) commute with each other, the quantum risk-sensitive estimator reduces to

$$\begin{aligned} &\left(\widehat{X_{\mathrm{e}}}^{\mu}(1), \ldots, \widehat{X_{\mathrm{e}}}^{\mu}(N)\right) \\ &\quad = \operatorname*{argmin}_{\mathfrak{u}_1\in\mathcal{Y}_1, \ldots, \mathfrak{u}_N\in\mathcal{Y}_N} \mathbb{P}\left[\exp\left(\mu_1\lambda^2 \sum_{i=1}^{N-1} |j_i(X_{\mathrm{e}}) - \mathfrak{u}_i|^2 \right.\right. \\ &\qquad\qquad \left.\left. + \mu_2 |j_N(X_{\mathrm{e}}) - \mathfrak{u}_N|^2\right)\right] \end{aligned}$$

which is identical to the definition of the (generalized) classical risk-sensitive estimator (1). Hence, (32) is a natural noncommutative extension of the classical risk-sensitive estimator to the quantum case.

Remark 5.2: In the limit of $\mu_1, \mu_2 \to 0$, $\widehat{X_{\rm e}}^{\mu}(l)$ coincides with the standard quantum optimal estimator $\pi_l(X)$ in (22). This is easily seen as follows. The estimation error cost function (33) is expanded to first order in μ_1 and μ_2 as

$$F(\mathfrak{u}_1,\ldots,\mathfrak{u}_N) = 1+\mu_1\lambda^2\sum_{i=1}^{N-1}\mathbb{P}\Big(|j_i(X_{\rm e})-\mathfrak{u}_i|^2\Big) + \mu_2\mathbb{P}\Big(|j_N(X_{\rm e})-\mathfrak{u}_N|^2\Big)+o(\mu_1,\mu_2).$$

Thus, in the limit $\mu_1, \mu_2 \to 0$, the minimizers of this function are given by $\mathfrak{u}_l^{\rm opt} = \pi_l(X_{\rm e})$, $(1 \le l \le N)$, i.e., we have

$$\lim_{\mu_1,\mu_2\to 0}\widehat{X_{\rm e}}^{\mu}(l) = \pi_l(X_{\rm e}).$$

For this reason, $\pi_l(X_{\rm e})$ is called the risk-neutral estimator.

The remainder of this section is organized as follows. First, we introduce a risk-sensitive information density matrix ϱ_l^μ, which is the quantum analogue to the classical information state $\alpha_l^\mu(x)$ discussed in Section I-A. Second, we derive a recursive equation for ϱ_l^μ. As in the standard quantum filtering case, ϱ_l^μ contains all information needed to calculate the estimator (32). More specifically, the cost function (33) can be represented only in terms of ϱ_l^μ, see Section V-C.

A. Quantum Risk-Sensitive Information State

We start by defining the following modification of the unitaries given by the difference (17):

$$U^\mu(l) := U(l)R(l). \tag{36}$$

Here $R(l)$ is given by (34). Note that $R(l)$ depends on $\mu = (\mu_1, \mu_2)$, but only through μ_1. Using (15) and (35), we find the difference equation for $U^\mu(l)$ as follows:

$$\begin{aligned}\Delta U^\mu(l) &= U(l)R(l) - U^\mu(l-1)\\ &= M_l U(l-1)U(l-1)^*\\ &\quad\times {\rm e}^{\mu_1\lambda^2 K_{l-1}/2}U(l-1)R(l-1) - U^\mu(l-1)\\ &= \Big[M_l{\rm e}^{\mu_1\lambda^2K_{l-1}/2} - I\Big]U^\mu(l-1)\\ U^\mu(0) &= I.\end{aligned}$$

Here, we have used K_{l-1} as a short hand for $K(\check{\mathfrak{u}}_{l-1})$. Using (16), this can be rewritten as

$$\begin{aligned}\Delta U^\mu(l) = \Bigg[&\Big\{M^\circ{\rm e}^{\mu_1\lambda^2K_{l-1}/2} + \frac{1}{\lambda^2}({\rm e}^{\mu_1\lambda^2K_{l-1}/2}-1)\Big\}\Delta t(l)\\ &+M^+{\rm e}^{\mu_1\lambda^2K_{l-1}/2}\Delta A(l)^* + M^-{\rm e}^{\mu_1\lambda^2K_{l-1}/2}\Delta A(l)\\ &+M^\pm{\rm e}^{\mu_1\lambda^2K_{l-1}/2}\Delta\Lambda(l)\Bigg]U^\mu(l-1).\end{aligned}$$

Now, let us define $V^\mu(l)$ as the solution to the difference equation

$$\begin{aligned}\Delta V^\mu(l) = \Bigg[&\Big\{M^\circ{\rm e}^{\mu_1\lambda^2K_{l-1}/2} + \frac{1}{\lambda^2}({\rm e}^{\mu_1\lambda^2K_{l-1}/2}-1)\Big\}\Delta t(l)\\ &+M^+{\rm e}^{\mu_1\lambda^2K_{l-1}/2}\Delta W(l)\Bigg]V^\mu(l-1),\end{aligned} \tag{37}$$

with $V^\mu(0) = I$. Note that this equation is identical to (24) when $\mu_1 = 0$. Two crucial properties of $V^\mu(l)$ are given in the following lemma.

Lemma 5.1: For all $1 \le l \le N$ the matrix $V^\mu(l)$ is an element of $\mathcal{M}\otimes\mathcal{C}_l \subset \mathcal{C}_l'$. Moreover, we have

$$\mathbb{P}\Big[U^\mu(l)^*XU^\mu(l)\Big] = \mathbb{P}\Big[V^\mu(l)^*XV^\mu(l)\Big] \tag{38}$$

for any X in $\mathcal{M}\otimes\mathcal{W}_N$.

Proof: To prove the first assertion, we assume that $V^\mu(l-1) \in \mathcal{M}\otimes\mathcal{C}_{l-1}$. Since $V^\mu(l)$ is calculated recursively, using $\Delta W(l)$, $\Delta t(l)$, and $V^\mu(l-1)$, all of which are included in $\mathcal{M}\otimes\mathcal{C}_l$, we obtain $V^\mu(l) \in \mathcal{M}\otimes\mathcal{C}_l$. The assertion follows by induction.

For the second claim, we note that $U^\mu(l)v\otimes\Phi^{\otimes N} = V^\mu(l)v\otimes\Phi^{\otimes N}$ holds for all vectors $v\in\mathbb{C}^2$ due to the relations $\Delta A(l)\Phi^{\otimes N} = \Delta\Lambda(l)\Phi^{\otimes N} = 0$ and $U^\mu(0) = V^\mu(0) = I$. Therefore, when the system density matrix is of the form $\rho = vv^*$, any $X\in\mathcal{M}\otimes\mathcal{W}_N$ satisfies

$$\begin{aligned}&\big\langle U^\mu(l)v\otimes\Phi^{\otimes N}, XU^\mu(l)v\otimes\Phi^{\otimes N}\big\rangle\\ &\qquad = \big\langle V^\mu(l)v\otimes\Phi^{\otimes N}, XV^\mu(l)v\otimes\Phi^{\otimes N}\big\rangle\end{aligned}$$

which directly implies (38) due to (6). Since every density matrix ρ is a convex combination of vector states, the lemma is proved. ∎

Definition 5.1: Since by Lemma 5.1 $V^\mu(l)$ is an element of the commutant of $\mathcal{C}_l$, we can define the following unnormalized risk-sensitive information state [14]:

$$\sigma_l^\mu(X) := U(l)^*\mathbb{P}(V^\mu(l)^*XV^\mu(l)|\mathcal{C}_l)U(l) \in \mathcal{Y}_l. \tag{39}$$

Moreover, we define ϱ_l^μ as the unnormalized risk-sensitive information density matrix corresponding to σ_l^μ by

$$\iota\left(\sigma_l^\mu(X)\right) = {\rm Tr}\left(\varrho_l^\mu X\right), \quad \forall\, X\in\mathcal{M}. \tag{40}$$

Lemma 5.2: Let $\check{\mathfrak{u}}_l$ be an element in $\mathcal{C}_l$. Let $Z:\mathbb{C}\to\mathcal{M}$ be an $\mathcal{M}$-valued function. Then we have

$$\iota\Big(\sigma_l^\mu(Z(\check{\mathfrak{u}}_l))\Big) = {\rm Tr}\Big[\varrho_l^\mu Z(u_l)\Big]$$

where $u_l = \iota(j_l(\check{\mathfrak{u}}_l)) = \iota(\mathfrak{u}_l)$ is a function of $\Delta y_i = \iota(\Delta Y(i))$, $(1 \le i \le l)$.

Proof: Denote the spectral decomposition of $\breve{\mathfrak{u}}_l$ by $\breve{\mathfrak{u}}_l = \sum_{x\in\mathrm{sp}(\breve{\mathfrak{u}}_l)} x P_{\breve{\mathfrak{u}}_l}(x) \in \mathcal{C}_l$. Then, it follows from the definitions (30) and (39) that we have

$$\begin{aligned}
&\sigma_l^\mu\left(Z(\breve{\mathfrak{u}}_l)\right)\\
&\quad= U(l)^*\mathbb{P}\left[V^\mu(l)^*\left(\sum_{x\in\mathrm{sp}(\breve{\mathfrak{u}}_l)} Z(x)P_{\breve{\mathfrak{u}}_l}(x)\right)V^\mu(l)\middle|\mathcal{C}_l\right]U(l)\\
&\quad= \sum_{x\in\mathrm{sp}(\breve{\mathfrak{u}}_l)} U(l)^*\mathbb{P}\left(V^\mu(l)^*Z(x)V^\mu(l)\middle|\mathcal{C}_l\right)U(l)\\
&\qquad\times U(l)^*P_{\breve{\mathfrak{u}}_l}(x)U(l)\\
&\quad= \sum_{x\in\mathrm{sp}(\breve{\mathfrak{u}}_l)} \sigma_l^\mu(Z(x))P_{U(l)^*\breve{\mathfrak{u}}_lU(l)}(x)\\
&\quad= \sum_{x\in\mathrm{sp}(\mathfrak{u}_l)} \sigma_l^\mu(Z(x))P_{\mathfrak{u}_l}(x).
\end{aligned}$$

In the first step we used $P_{\breve{\mathfrak{u}}_l}(x) \in \mathcal{C}_l$ and $[P_{\breve{\mathfrak{u}}_l}(x), V^\mu(l)] = 0$. Note that $\sigma_l^\mu(Z(x)) \in \mathcal{Y}_l$ and $P_{\mathfrak{u}_l}(x) \in \mathcal{Y}_l$ can be diagonalized simultaneously by a $*$-isomorphism ι. Using $\iota(P_{\mathfrak{u}_l}(x)) = \chi_{\{\iota(\mathfrak{u}_l)=x\}}$ (see Theorem 2.1), we get

$$\begin{aligned}
&\iota\left(\sigma_l^\mu\left(Z(\breve{\mathfrak{u}}_l)\right)\right)\\
&\quad= \sum_{x\in\mathrm{sp}(\mathfrak{u}_l)} \iota\left(\sigma_l^\mu(Z(x))\right)\iota\left(P_{\mathfrak{u}_l}(x)\right)\\
&\quad= \sum_{x\in\mathrm{sp}(\mathfrak{u}_l)} \mathrm{Tr}\left(\varrho_l^\mu Z(x)\right)\chi_{\{\iota(\mathfrak{u}_l)=x\}}\\
&\quad= \mathrm{Tr}\left[\varrho_l^\mu \sum_{x\in\mathrm{ran}(u_l)} Z(x)\chi_{\{u_l=x\}}\right]\\
&\quad= \mathrm{Tr}\left[\varrho_l^\mu Z(u_l)\right]
\end{aligned}$$

where we have used the definitions (31) and (6). Since $\mathfrak{u}_l \in \mathcal{Y}_l$, $u_l = \iota(\mathfrak{u}_l)$ is obviously a function of $\Delta y_1, \ldots, \Delta y_l$. This completes the proof. ∎

B. Dynamics of Risk-Sensitive Information Density Matrix

The objective here is to derive a recursive equation for ϱ_l^μ. Let X be an element of $\mathcal{M}$. A similar calculation to (19) yields the following difference equation for $\tilde{j}_l^\mu(X) := V^\mu(l)^*XV^\mu(l)$:

$$\begin{aligned}
\Delta\tilde{j}_l^\mu(X) =& \tilde{j}_{l-1}^\mu(\mathcal{L}^\mu(X, \breve{\mathfrak{u}}_{l-1}))\Delta t(l)\\
&+ \tilde{j}_{l-1}^\mu(\mathcal{J}^\mu(X, \breve{\mathfrak{u}}_{l-1}))\Delta W(l)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{L}^\mu(X,u) :=& \mathrm{e}^{\mu_1\lambda^2K(u)/2}\Big[M^{+*}XM^+ + \lambda^2M^{o*}XM^o\\
&+ XM^o + M^{o*}X\Big]\mathrm{e}^{\mu_1\lambda^2K(u)/2}\\
&+ \frac{1}{\lambda^2}\left(\mathrm{e}^{\mu_1\lambda^2K(u)/2}X\mathrm{e}^{\mu_1\lambda^2K(u)/2} - X\right)
\end{aligned}$$

$$\begin{aligned}
\mathcal{J}^\mu(X,u) :=& \mathrm{e}^{\mu_1\lambda^2K(u)/2}\Big[\lambda^2M^{+*}XM^o + \lambda^2M^{o*}XM^+\\
&+ XM^+ + M^{+*}X\Big]\mathrm{e}^{\mu_1\lambda^2K(u)/2}.
\end{aligned}$$

Note that $K(u) = |X_\mathrm{e} - u|^2$. Since $\tilde{j}_l^\mu(X)$ is an element of the commutant $\mathcal{C}_l'$, we can define the quantum conditional expectation $\tilde{\sigma}_l^\mu(X) := \mathbb{P}(\tilde{j}_l^\mu(X)|\mathcal{C}_l)$. This satisfies the difference equation

$$\begin{aligned}
\Delta\tilde{\sigma}_l^\mu(X) =& \tilde{\sigma}_{l-1}^\mu(\mathcal{L}^\mu(X, \breve{\mathfrak{u}}_{l-1}))\Delta t(l)\\
&+ \tilde{\sigma}_{l-1}^\mu(\mathcal{J}^\mu(X, \breve{\mathfrak{u}}_{l-1}))\Delta W(l).
\end{aligned}$$

Equation (39) can now be written as $\sigma_l^\mu(X) = U(l)^*\tilde{\sigma}_l^\mu(X)U(l)$. This means we find the difference equation

$$\begin{aligned}
\Delta\sigma_l^\mu(X) =& U(l-1)^*M_l^*\Delta\tilde{\sigma}_l^\mu(X)M_lU(l-1)\\
=& U(l-1)^*\tilde{\sigma}_{l-1}^\mu(\mathcal{L}^\mu(X, \breve{\mathfrak{u}}_{l-1}))U(l-1)\Delta t(l)\\
&+ U(l-1)^*\tilde{\sigma}_{l-1}^\mu(\mathcal{J}^\mu(X, \breve{\mathfrak{u}}_{l-1}))U(l-1)\\
&\times U(l)^*\Delta W(l)U(l)\\
=& \sigma_{l-1}^\mu(\mathcal{L}^\mu(X, \breve{\mathfrak{u}}_{l-1}))\Delta t(l)\\
&+ \sigma_{l-1}^\mu(\mathcal{J}^\mu(X, \breve{\mathfrak{u}}_{l-1}))\Delta Y(l) \qquad (41)
\end{aligned}$$

where in the last step (19) was used.

We can now represent (41) in terms of the unnormalized risk-sensitive information density matrix ϱ_l^μ. Since $\sigma_l^\mu(X)$, $\Delta t(l)$, and $\Delta Y(l)$ are elements in $\mathcal{Y}_l$, they can be simultaneously diagonalized by a $*$-isomorphism ι, which leads to

$$\begin{aligned}
\Delta\iota(\sigma_l^\mu(X)) =& \iota(\sigma_{l-1}^\mu(\mathcal{L}^\mu(X, \breve{\mathfrak{u}}_{l-1})))\Delta t_l\\
&+ \iota(\sigma_{l-1}^\mu(\mathcal{J}^\mu(X, \breve{\mathfrak{u}}_{l-1})))\Delta y_l\\
\sigma_0^\mu =& \psi.
\end{aligned}$$

where $\Delta t_l = \iota(\Delta t(l))$ and $\Delta y_l = \iota(\Delta Y(l))$. It then follows from Lemma 5.2 that the above equation leads to

$$\Delta\varrho_l^\mu = \bar{\mathcal{L}}^\mu(\varrho_{l-1}^\mu, u_{l-1})\Delta t_l + \bar{\mathcal{J}}^\mu(\varrho_{l-1}^\mu, u_{l-1})\Delta y_l, \quad \varrho_0^\mu = \rho \qquad (42)$$

where $u_{l-1} = \iota(j_{l-1}(\breve{\mathfrak{u}}_{l-1})) = \iota(\mathfrak{u}_{l-1})$ is a function of $\Delta y_1, \ldots, \Delta y_{l-1}$. The operators $\bar{\mathcal{L}}^\mu$ and $\bar{\mathcal{J}}^\mu$ are defined as follows:

$$\begin{aligned}
\bar{\mathcal{L}}^\mu(\varrho,u) :=& M^+\mathcal{H}(\varrho,u)M^{+*} + \lambda^2M^\circ\mathcal{H}(\varrho,u)M^{\circ*}\\
&+ M^\circ\mathcal{H}(\varrho,u) + \mathcal{H}(\varrho,u)M^{\circ*}\\
&+ \frac{1}{\lambda^2}(\mathcal{H}(\varrho,u) - \varrho)\\
\bar{\mathcal{J}}^\mu(\varrho,u) :=& \lambda^2M^+\mathcal{H}(\varrho,u)M^{\circ*} + \lambda^2M^\circ\mathcal{H}(\varrho,u)M^{+*}\\
&+ M^+\mathcal{H}(\varrho,u) + \mathcal{H}(\varrho,u)M^{+*}\\
\mathcal{H}(\varrho,u) :=& \mathrm{e}^{\mu_1\lambda^2K(u)/2}\varrho\mathrm{e}^{\mu_1\lambda^2K(u)/2}.
\end{aligned}$$

Equation (42) is a simple recursion for a 2×2 matrix and is thus easily implementable on a digital computer. Note that the operators $\bar{\mathcal{L}}^\mu$ and $\bar{\mathcal{J}}^\mu$ reduce when $\mu_1 = 0$ to $\bar{\mathcal{L}}^0 = \bar{\mathcal{L}}$ and $\bar{\mathcal{J}}^0 = \bar{\mathcal{J}}$, where $\bar{\mathcal{L}}$ and $\bar{\mathcal{J}}$ are given in (28) and (29). This implies that the solution of (42) converges to that of (27) when μ_1 goes to zero.

C. Calculating the Risk-Sensitive Estimator

We will now represent the cost function (33) in terms of ϱ_l^μ only. To this end, we define a new state $\mathbb{Q}^l$ on $\mathcal{M}\otimes\mathcal{W}_N$ by

$$\mathbb{Q}^l(X) := \mathbb{P}\Big[U(l)XU(l)^*\Big]$$

for $X\in\mathcal{M}\otimes\mathcal{W}_N$. Since $\mathcal{Y}_l$ is a commutative $*$-subalgebra of $\mathcal{M}\otimes\mathcal{W}_N$, we can apply Theorem 2.1 to $(\mathcal{Y}_l,\mathbb{Q}^l)$. That is, there exists a classical probability space $\big(\Omega_l,\mathcal{F}_l,\mathbf{Q}^l\big)$ and a $*$-isomorphism $\iota:\mathcal{Y}_l\to\ell^\infty(\mathcal{F}_l)$ such that $\mathbb{Q}^l(A)=\mathbf{E}_{\mathbf{Q}^l}[\iota(A)]$ for all $A\in\mathcal{Y}_l$. We now have the following theorem.

Theorem 5.1: The cost function (33) can be written as

$$F(\mathfrak{u}_1,\dots,\mathfrak{u}_l)=\mathbf{E}_{\mathbf{Q}^l}\left[\mathrm{Tr}\Big(\varrho_l^\mu \mathrm{e}^{\mu_2|X_\mathrm{e}-u_l|^2}\Big)\right] \tag{43}$$

where $u_l=\iota(\mathfrak{u}_l)$ is a function of the measurement data $\Delta y_1,\dots,\Delta y_l$, and ϱ_l^μ is the risk-sensitive information density matrix that satisfies (42).

Proof: We define $\check{\mathfrak{u}}_l=j_l^{-1}(\mathfrak{u}_l)\in\mathcal{C}_l$ as before. Since $|j_l(X_\mathrm{e})-\mathfrak{u}_l|^2=j_l(|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2)$, we find

$$\begin{aligned}F(\mathfrak{u}_1,\dots,\mathfrak{u}_l)&=\mathbb{P}\Big[R(l)^*\mathrm{e}^{\mu_2 j_l(|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2)}R(l)\Big]\\&=\mathbb{P}\Big[R(l)^*U(l)^*\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2}U(l)R(l)\Big]\\&=\mathbb{P}\Big[U^\mu(l)^*\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2}U^\mu(l)\Big]\end{aligned}$$

where $U^\mu(l)$ is defined by (36). Using (38) in Lemma 5.1 and the tower property of the conditional expectation, we find

$$\begin{aligned}&F(\mathfrak{u}_1,\dots,\mathfrak{u}_l)\\&\quad=\mathbb{P}\Big[V^\mu(l)^*\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2}V^\mu(l)\Big]\\&\quad=\mathbb{P}\left[\mathbb{P}\Big(V^\mu(l)^*\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2}V^\mu(l)\Big|\mathcal{C}_l\Big)\right]\\&\quad=\mathbb{Q}^l\left[U(l)^*\mathbb{P}\Big(V^\mu(l)^*\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2}V^\mu(l)\Big|\mathcal{C}_l\Big)U(l)\right]\\&\quad=\mathbb{Q}^l\left[\sigma_l^\mu\Big(\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2}\Big)\right]\end{aligned}$$

where we have used the definition of σ_l^μ in (39). Note that the above conditional expectation is well defined due to $X_\mathrm{e}-\check{\mathfrak{u}}_l\in\mathcal{C}_l'$ and $V^\mu(l)\in\mathcal{C}_l'$. Let u_l be given by $u_l=\iota(j_l(\check{\mathfrak{u}}_l))=\iota(\mathfrak{u}_l)$. It now follows from Lemma 5.2 that

$$\iota\Big(\sigma_l^\mu(\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2})\Big)=\mathrm{Tr}\Big[\varrho_l^\mu\mathrm{e}^{\mu_2|X_\mathrm{e}-u_l|^2}\Big].$$

Consequently, the cost function can be written as

$$\begin{aligned}F(\mathfrak{u}_1,\dots,\mathfrak{u}_l)&=\mathbf{E}_{\mathbf{Q}^l}\left[\iota\Big(\sigma_l^\mu\Big(\mathrm{e}^{\mu_2|X_\mathrm{e}-\check{\mathfrak{u}}_l|^2}\Big)\Big)\right]\\&=\mathbf{E}_{\mathbf{Q}^l}\left[\mathrm{Tr}\Big(\varrho_l^\mu\mathrm{e}^{\mu_2|X_\mathrm{e}-u_l|^2}\Big)\right].\end{aligned}$$

This completes the proof. ∎

As a result of Theorem 5.1, our estimation problem is now cast as a classical optimal control problem. The resulting problem can be solved systematically by dynamic programming. We will only provide a brief summary of this. Consider the following *optimal expected cost-to-go* $f_l(\varrho)$ at time l, given that $\varrho_l^\mu=\varrho$:

$$\begin{aligned}f_l(\varrho)&=\min_{u_l,\dots,u_N}\mathbf{E}_{\mathbf{Q}^N}\left[\mathrm{Tr}\Big(\varrho_N^\mu\mathrm{e}^{\mu_2|X_\mathrm{e}-u_N|^2}\Big)\Big|\varrho_l^\mu=\varrho\right]\\f_{N+1}(\varrho)&=\mathrm{Tr}(\varrho).\end{aligned}$$

This leads to the following dynamic programming equation; denoting (42) simply as $\varrho_l^\mu=\Gamma\big(\varrho_{l-1}^\mu,u_{l-1},\Delta y_l\big)$, we have

$$\begin{aligned}f_{l-1}(\varrho)&=\min_{u_{l-1}}\mathbf{E}_{\mathbf{Q}^N}\Big[f_l(\Gamma(\varrho,u_{l-1},\Delta y_l))\Big]\\&=\min_{u_{l-1}}\frac{1}{2}\Big[f_l(\Gamma(\varrho,u_{l-1},\lambda))+f_l(\Gamma(\varrho,u_{l-1},-\lambda))\Big].\end{aligned}$$

Note here that $\mathrm{Prob}_{\mathbf{Q}^N}(\Delta y_l=\lambda)=\mathbb{Q}^N\big[U(l)^*P_l^+U(l)\big]=\mathbb{P}\big(P_l^+\big)=1/2$. We can run the above algorithm efficiently in a digital computer and obtain the optimal sequence u_l^{opt}, $(1\le l\le N)$, which yields $\widehat{X_\mathrm{e}}^\mu(l)=\iota^{-1}(u_l^{\mathrm{opt}})$ through a verification theorem (e.g., see [25]). Theorem 6.1 below will lead to a robustness result for the risk-sensitive estimator.

Remark 5.3: Running the dynamic programming recursion on a digital computer is very costly computationally. Therefore we define a *suboptimal risk-sensitive estimator* by

$$\begin{aligned}&\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(l)\\&:=\operatorname*{argmin}_{\mathfrak{u}_l\in\mathcal{Y}_l}F\Big(\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(1),\dots,\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(l-1),\mathfrak{u}_l\Big).\end{aligned}$$

That is, $\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(l)$ is to be calculated based on the assumption that we have already performed the above minimization procedure up to time $l-1$ and obtained the suboptimal risk-sensitive estimators $\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(i)\in\mathcal{Y}_i$, $(1\le i\le l-1)$. As shown in [7] (Theorems 2.2 and 4.2), the minimizer of the trace function inside the expectation in (43), $u_l^{\min}$, leads to the suboptimal risk-sensitive estimator $\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(l)=\iota^{-1}\big(u_l^{\min}\big)$. Hence our algorithm in this case is represented simply as follows:

$$\begin{aligned}\iota\Big(\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(l)\Big)&=\operatorname*{argmin}_{u_l\in\mathbb{R}}\mathrm{Tr}\Big[\varrho_l^\mu\mathrm{e}^{\mu_2|X_\mathrm{e}-u_l|^2}\Big]\\\varrho_l^\mu&=\Gamma\Big(\varrho_{l-1}^\mu,\iota\Big(\widehat{X_\mathrm{e}}^{\mu,\mathrm{sub}}(l-1)\Big),\Delta y_l\Big)\end{aligned} \tag{44}$$

which is of the same structure as the classical algorithm presented in Section I-A. In Theorem 6.2 we will derive a bound for the conditional estimation error. The suboptimal risk-sensitive estimator minimizes this error bound. This provides a sound theoretical foundation for the suboptimal risk-sensitive estimator. Since the algorithm (44) is computationally much cheaper than the dynamic programming equation, we will consistently use the suboptimal estimator in the example part, Section VII.

VI. Quantum Uncertain Systems and Robustness of the Risk-Sensitive Estimator

In realistic situations, we often have to deal with a system that includes some model uncertainty. From the classical case, we expect that the risk-sensitive estimator has an enhanced robustness property against such uncertainty. In this section, we first describe a class of uncertain quantum systems for which the uncertainty is quantified by the quantum relative entropy. We will then show robustness properties of the estimator.

A. Quantum Uncertain Systems

Uncertainty can enter the system in many ways. It could for instance be the case that the state ψ is unknown to us. The uncertainty then enters the system density matrix ρ through the relation $\mathbb{P}(X) = \psi \otimes \phi^{\otimes N}(X) = \mathrm{Tr}[X(\rho \otimes (\Phi\Phi^*)^{\otimes N})]$. We assume that the field state is known and fixed to the vacuum ϕ. This, however, is not the only way uncertainty can enter our model. We will also allow for uncertainty in the coefficients of the dynamics, i.e., the difference (17). We can push this uncertainty into the initial state, as described below.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a classical probability space. Let p be an element of $L^\infty(\Omega, \mathcal{F}, \mathbf{P})$, i.e., p is a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$. Let $\mathfrak{p}$ be the operator on $L^2(\Omega, \mathcal{F}, \mathbf{P})$ given by pointwise multiplication with p, i.e.,

$$(\mathfrak{p}f)(\omega) = p(\omega)f(\omega), \quad f \in L^2(\Omega, \mathcal{F}, \mathbf{P}), \ \omega \in \Omega.$$

We denote the commutative $*$-algebra of all such multiplication operators with functions in $L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ by $\mathcal{P}$. On $\mathcal{P}$ we can define a state τ as integration with respect to the measure $\mathbf{P}$, i.e., $\tau(\mathfrak{p}) = \int_\Omega p(\omega)\mathbf{P}(d\omega)$. For simplicity we will take the operator $\mathfrak{p}$ to be self-adjoint, that is, it is a multiplication with a real-valued function. Next, let M^i, $(i = \pm, +, -, \circ)$ be $\mathcal{M}$-valued functions on $\mathbb{C}$, i.e., $M^i : \mathbb{C} \to \mathcal{M}$, such that the matrix M_l in (16) is unitary. Then, using the compositions of M^i and $\mathfrak{p} \in \mathcal{P}$, we can define the following difference equation:

$$\begin{aligned} \Delta U(l) = \Big[& M^{\pm}(\mathfrak{p})\Delta\Lambda(l) \\ & + M^{+}(\mathfrak{p})\Delta A(l)^* + M^{-}(\mathfrak{p})\Delta A(l) \\ & + M^{\circ}(\mathfrak{p})\Delta t(l)\Big] U(l-1) \\ U(0) = I & \end{aligned} \tag{45}$$

on the extended quantum probability space

$$(\mathcal{P} \otimes \mathcal{M} \otimes \mathcal{W}_N, \mathbb{P}) = (\mathcal{P} \otimes \mathcal{M} \otimes \mathcal{M}^{\otimes N}, \tau \otimes \psi \otimes \phi^{\otimes N}).$$

We now assume that the state $\tau \otimes \psi$ is unknown to us. This means that (45) is equivalent to the difference equation (17) such that its coefficients include uncertain parameter p. That is, the uncertainty in the model has been pushed completely into the state $\tau \otimes \psi$.

Now, let $\rho^{\mathrm{true}} = \rho_{\mathrm{p}}^{\mathrm{true}} \otimes \rho_{\mathrm{s}}^{\mathrm{true}}$ be the *true density matrix* corresponding to the unknown state $\tau \otimes \psi$. Then, the true filter is initialized to $\varrho_0 = \rho^{\mathrm{true}}$. However, as ρ^{true} is unknown, we fix a *nominal density matrix* $\rho^{\mathrm{nom}} = \rho_{\mathrm{p}}^{\mathrm{nom}} \otimes \rho_{\mathrm{s}}^{\mathrm{nom}}$, which in general differs from ρ^{true}, and construct the nominal filter that starts from $\varrho_0 = \rho^{\mathrm{nom}}$. The nominal estimator of $j_l(X)$ is then given by

$$\begin{aligned} \pi_l^{\mathrm{nom}}(X) &= \frac{\sigma_l^{\mathrm{nom}}(X)}{\sigma_l^{\mathrm{nom}}(I)} \\ \sigma_l^{\mathrm{nom}}(X) &= \iota^{-1}\Big(\mathrm{Tr}(\varrho_l^{\mathrm{nom}} X)\Big), \quad \varrho_0 = \rho^{\mathrm{nom}}. \end{aligned}$$

Example 6.1: If p is a discrete random variable that takes the values p_i, $(1 \le i \le m)$ with unknown probability r_i, then the corresponding multiplication operator is $\mathfrak{p} = \mathrm{diag}\{p_1, \ldots, p_m\} \in \mathcal{P}$, where the commutative $*$-algebra $\mathcal{P}$ is the set of $m \times m$ diagonal matrices, and the true density matrix is $\rho_{\mathrm{p}}^{\mathrm{true}} = \mathrm{diag}\{r_1, \ldots, r_m\}$. To design a nominal filter, we choose a nominal density matrix of the form $\rho_{\mathrm{p}}^{\mathrm{nom}} = \mathrm{diag}\,\{r_1', \ldots, r_m'\}$. In general, $r_i \neq r_i'$. It is easily seen that the quantum relative entropy between the above two distributions is equal to the classical one:

$$R\left(\rho_{\mathrm{p}}^{\mathrm{true}} \| \rho_{\mathrm{p}}^{\mathrm{nom}}\right) = \sum_{i=1}^{m} r_i \log \frac{r_i}{r_i'} = R^{\mathrm{c}}\left(\{r_i\} \| \{r_i'\}\right).$$

An important example for a true density matrix is $\rho_{\mathrm{p}}^{\mathrm{true}} = \mathrm{diag}\{1, 0, \ldots, 0\}$; that is, p is not a random variable but an unknown deterministic system parameter $p = p_1$. If we have no information about p at all, it is natural to take a uniform distribution $\rho_{\mathrm{p}}^{\mathrm{nom}} = \mathrm{diag}\{1/m, \ldots, 1/m\}$ as the nominal distribution.

B. Robustness Properties of the Risk-Sensitive and Suboptimal Risk-Sensitive Estimators

The nominal estimator $\pi_l^{\mathrm{nom}}(X_{\mathrm{e}})$ differs from the true one $\pi_l^{\mathrm{true}}(X_{\mathrm{e}})$. Hence, $\pi_l^{\mathrm{nom}}(X_{\mathrm{e}})$ is no longer the optimal estimator in the sense of the mean square error and thus can possibly take a large estimation error. However, as shown below, if one uses the nominal risk-sensitive estimator defined in (32), the estimation error is guaranteed to be within a certain bound. This implies that the risk-sensitive estimator does have a robustness property against unknown perturbation of the system state and the system parameters.

The quantum relative entropy (7) will be used to express the robustness property. We here assume that the unknown true density matrix $\rho^{\mathrm{true}} = \rho_{\mathrm{p}}^{\mathrm{true}} \otimes \rho_{\mathrm{s}}^{\mathrm{true}}$ is within a certain distance from a known nominal density matrix $\rho^{\mathrm{nom}} = \rho_{\mathrm{p}}^{\mathrm{nom}} \otimes \rho_{\mathrm{s}}^{\mathrm{nom}}$, that is,

$$R(\rho^{\mathrm{true}} \| \rho^{\mathrm{nom}}) < +\infty.$$

The following theorem will lead to a robustness property of the nominal risk-sensitive estimator (32).

Theorem 6.1: Let $\mathfrak{u}_l$, $(1 \le l \le N)$ be an element of $\mathcal{Y}_l$. Then, we have the following inequality:

$$\begin{aligned} & \mathbb{P}_{\mathrm{true}}\left[\log\left(R(N)^* \mathrm{e}^{\mu_2 |j_N(X_{\mathrm{e}}) - \mathfrak{u}_N|^2} R(N)\right)\right] \\ & \quad \le \log \mathbb{P}_{\mathrm{nom}}\left[R(N)^* \mathrm{e}^{\mu_2 |j_N(X_{\mathrm{e}}) - \mathfrak{u}_N|^2} R(N)\right] \\ & \qquad + R(\rho^{\mathrm{true}} \| \rho^{\mathrm{nom}}) \end{aligned} \tag{46}$$

where $R(N)$ is defined by (34), and $\mathbb{P}_{\rm true}(X) = {\rm Tr}[X(\rho^{\rm true} \otimes (\Phi\Phi^*)^{\otimes N})]$ and $\mathbb{P}_{\rm nom}(X) = {\rm Tr}[X(\rho^{\rm nom} \otimes (\Phi\Phi^*)^{\otimes N})]$.

Proof: Setting $\rho = \rho^{\rm true} \otimes (\Phi\Phi^*)^{\otimes N}$ and $\rho' = \rho^{\rm nom} \otimes (\Phi\Phi^*)^{\otimes N}$ in (9), we have

$$\mathbb{P}_{\rm true}(Z) \leq \log \mathbb{P}_{\rm nom}({\rm e}^Z) + R(\rho^{\rm true}||\rho^{\rm nom}) \quad \forall\, Z \in \mathcal{P} \otimes \mathcal{M} \otimes \mathcal{W}_N$$

where we have used the following additivity property:

$$\begin{aligned} &R(\rho^{\rm true} \otimes (\Phi\Phi^*)^{\otimes N}||\rho^{\rm nom} \otimes (\Phi\Phi^*)^{\otimes N}) \\ &\quad = R(\rho^{\rm true}||\rho^{\rm nom}) + R((\Phi\Phi^*)^{\otimes N}||(\Phi\Phi^*)^{\otimes N}) \\ &\quad = R(\rho^{\rm true}||\rho^{\rm nom}). \end{aligned}$$

Therefore, taking $Z = \log[R(N)^* {\rm e}^{\mu_2 |j_N(X_{\rm e}) - \mathfrak{u}_N|^2} R(N)]$ yields the theorem. ∎

Equation (46) is a quantum version of the classical robustness result (2), because the left hand side of (46) can be expanded up to second order in the estimation error as

$$\begin{aligned} &\mathbb{P}_{\rm true}\left[\mu_1 \sum_{l=1}^{N-1} |j_l(X_{\rm e}) - \mathfrak{u}_l|^2 + \mu_2 |j_N(X_{\rm e}) - \mathfrak{u}_N|^2\right] \\ &\quad + O(|j_l(X_{\rm e}) - \mathfrak{u}_l|^4) \\ &\qquad \leq \log \mathbb{P}_{\rm nom}\left[R(N)^* {\rm e}^{\mu_2 |j_N(X_{\rm e}) - \mathfrak{u}_N|^2} R(N)\right] \\ &\qquad\quad + R(\rho^{\rm true}||\rho^{\rm nom}). \end{aligned}$$

That is, as in the classical case, the nominal risk-sensitive estimator $\mathfrak{u}_l = \widehat{X_{\rm e}}^{\mu,{\rm nom}}(l)$, defined by (32), does have a robustness property, because it minimizes the upper bound of the estimation error under the unknown true state $\mathbb{P}_{\rm true}$.

We remark that the relative entropy in (46) can be written as

$$R(\rho^{\rm true}||\rho^{\rm nom}) = R(\rho_{\rm p}^{\rm true}||\rho_{\rm p}^{\rm nom}) + R(\rho_{\rm s}^{\rm true}||\rho_{\rm s}^{\rm nom}).$$

The first term is a classical relative entropy as shown in Example 6.1. Thus, if there is no uncertainty in the quantum state, the estimation error bound is written in terms of classical quantities only.

We now change our focus to the suboptimal risk-sensitive estimator $\widehat{X_{\rm e}}^{\mu,{\rm sub}}(l)$ defined in Remark 5.3. The following theorem shows that the conditional estimation error at time l also has an upper bound. This will lead to a robustness property for the nominal suboptimal risk-sensitive estimator.

Theorem 6.2: Let $\mathfrak{u}_l, (1 \leq l \leq N)$ be an element of $\mathcal{Y}_l$. Then, we have the following inequality:

$$\begin{aligned} &\iota\left(\mathbb{P}_{\rm true}\left[|j_l(X_{\rm e}) - \mathfrak{u}_l|^2 \middle| \mathcal{Y}_l\right]\right) \\ &\quad \leq \frac{1}{\mu_2} \log {\rm Tr}\left[\rho_l^{\mu,{\rm nom}} {\rm e}^{\mu_2 |X_{\rm e} - u_l|^2}\right] \\ &\qquad + \frac{1}{\mu_2} R(\rho_l^{\rm true}||\rho_l^{\mu,{\rm nom}}) \end{aligned} \tag{47}$$

where $\rho_l^{\rm true} = \varrho_l^{\rm true}/{\rm Tr}\,[\varrho_l^{\rm true}]$ and $\rho_l^{\mu,{\rm nom}} = \varrho_l^{\mu,{\rm nom}}/{\rm Tr}[\varrho_l^{\mu,{\rm nom}}]$ are the conditional density matrices corresponding to the true filter and the nominal risk-sensitive filter, and $u_l = \iota(\mathfrak{u}_l)$.

Proof: Using the definition of the optimal estimator $\pi_l(X_{\rm e})$, the left-hand side in (47) can be rewritten as

$$\begin{aligned} &\iota\left(\mathbb{P}_{\rm true}\left[j_l(|X_{\rm e} - \breve{\mathfrak{u}}_l|^2) \middle| \mathcal{Y}_l\right]\right) \\ &\quad = \iota\left(\pi_l^{\rm true}\left(|X_{\rm e} - \breve{\mathfrak{u}}_l|^2\right)\right) \\ &\quad = {\rm Tr}\left[\rho_l^{\rm true} |X_{\rm e} - u_l|^2\right] \end{aligned}$$

where the last equality follows directly from Lemma 5.2 with $u_l = \iota(j_l(\breve{\mathfrak{u}}_l)) = \iota(\mathfrak{u}_l)$. Then, from (9) we have the assertion. ∎

The first term of the right-hand side in (47) is minimized when choosing the nominal suboptimal risk-sensitive estimator $u_l = \iota\left(\widehat{X_{\rm e}}^{\mu,{\rm sub}}(l)\right)$ given by (44). Theorem 6.2 therefore shows a robustness property of the suboptimal risk-sensitive estimator.

VII. Examples

In this section, we study two examples in detail. The first example is a two-level atom that is coupled to the field via a *dispersive interaction*. This coupling can be obtained by putting the atom in a cavity that has a resonance frequency far detuned from the transition frequency of the two-level atom. The second example deals with a two-level atom that decays to the ground state due to *spontaneous emission* into its environment. We consider the situation where the quantum state of the two-level atom and a physical parameter are unknown to us. In particular, we employ the nominal suboptimal risk-sensitive estimator given by (44). We compare this estimator with both the true risk-neutral and nominal risk-neutral estimators.

A. Dispersive Interaction Model

The interaction Hamiltonian (13) in case of a dispersive interaction with the field, is given by the following system matrices:

$$L_1 = 0, \quad L_2 = i\sqrt{g}\sigma_z, \quad L_3 = 0$$

where $\sigma_z = {\rm diag}\{1, -1\}$ and $g > 0$ represents the interaction strength. From (14) and (16), we see that the matrices M^i, $(i = \pm, +, -, \circ)$ are given by

$$\begin{aligned} &M^{\pm}(g) = 0, \quad M^{+}(g) = \frac{\sin(g\lambda)}{\lambda}\sigma_z \\ &M^{-}(g) = -\frac{\sin(g\lambda)}{\lambda}\sigma_z, \quad M^{\circ}(g) = \frac{\cos(g\lambda) - 1}{\lambda^2} I. \end{aligned}$$

We assume that g is a classical random variable that takes the values g_i with unknown probabilities ${\rm Prob}(g_i) = r_i$. As seen in Section VI-A, g can be regarded as an observable $\mathfrak{g} = {\rm diag}\{g_1, \ldots, g_m\} \in \mathcal{P}$, where $\mathcal{P}$ is a commutative $*$-algebra given by the set of $m \times m$ diagonal matrices. The corresponding unknown true density matrix is $\rho_{\rm p}^{\rm true} = {\rm diag}\{r_1, \ldots, r_m\}$. In particular, we now study a toy model in which g can take 20 discrete values, $g_i = 0.4 + 0.03i$, $(i = 1, \ldots, 20)$. Moreover, we choose $\rho_{\rm p}^{\rm true}$ to be given by

$$\begin{aligned} \rho_{\rm p}^{\rm true} = {\rm diag}\{&0, 0.01, 0.04, 0.1, 0.7, 0.1, 0.04, 0.01, 0, 0, \\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0\} \end{aligned}$$

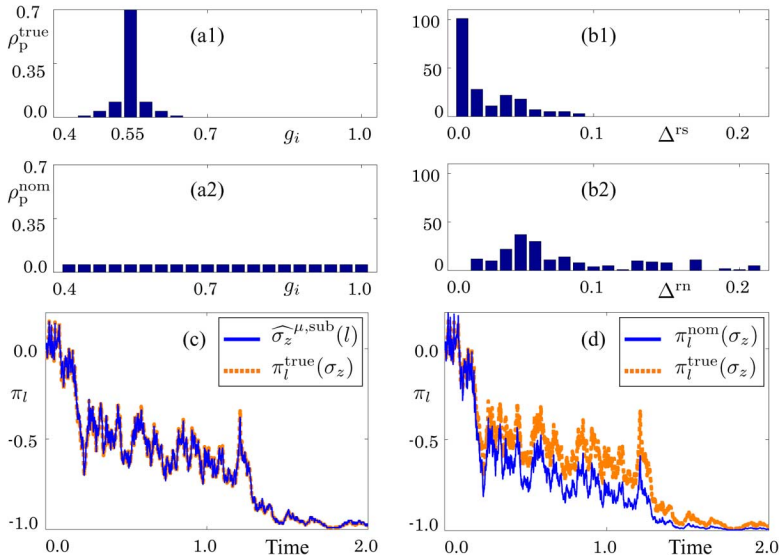

Fig. 1. For the dispersive interaction model of the atom, (a) the true and nominal parameter distributions, (b) the histogram of the averaged total estimation errors, (c) sample paths of the nominal risk-sensitive estimator (solid line) and the true risk-neutral one (thick dotted line), and (d) sample paths of the nominal risk-neutral estimator (solid line) and the true risk-neutral one (thick dotted line). For the figures (c) and (d), the notation (ι) is omitted.

which is illustrated in Fig. 1 (a1). For instance, g takes $g_3 = 0.49$ with probability $\mathrm{Prob}(g_3) = 0.04$. Furthermore, we assume that the true density matrix is given by

$$\rho_{\mathrm{s}}^{\mathrm{true}} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}. \tag{48}$$

Again, note that $\rho^{\mathrm{true}} = \rho_{\mathrm{p}}^{\mathrm{true}} \otimes \rho_{\mathrm{s}}^{\mathrm{true}}$ is unknown to us.

Now, let us consider estimating the system observable $X_{\mathrm{e}} = \sigma_z$. To design a nominal filter, we use the following nominal density matrix in $\mathcal{P} \otimes \mathcal{M}$:

$$\begin{aligned} \rho^{\mathrm{nom}} &= \rho_{\mathrm{p}}^{\mathrm{nom}} \otimes \rho_{\mathrm{s}}^{\mathrm{nom}} \\ \rho_{\mathrm{p}}^{\mathrm{nom}} &= \mathrm{diag}\left\{ \frac{1}{20}, \ldots, \frac{1}{20} \right\} \\ \rho_{\mathrm{s}}^{\mathrm{nom}} &= \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}. \end{aligned} \tag{49}$$

$\rho_{\mathrm{p}}^{\mathrm{nom}}$ is depicted in Fig. 1 (a2). The nominal risk-neutral estimator $\pi_l^{\mathrm{nom}}(\sigma_z)$ and the nominal risk-sensitive one $\widehat{\sigma_z}^{\mu,\mathrm{sub}}(l)$ are then calculated from (27) with $\varrho_0 = \rho^{\mathrm{nom}}$ and (44) with $\varrho_0^{\mu} = \rho^{\mathrm{nom}}$, respectively. The risk-sensitive parameters are chosen to be $(\mu_1, \mu_2) = (0.1, 0.182)$. Note that the filter equations include the composition $M^i(\mathfrak{g})$ and are driven by the true output data Δy_l. We compare those two nominal estimators with the ideal true risk-neutral estimator $\pi_l^{\mathrm{true}}(\sigma_z)$, which is calculated from (27) with $\varrho_0 = \rho^{\mathrm{true}}$. To do this, we use the averaged total estimation errors

$$\begin{aligned} \Delta^{\mathrm{rn}} &:= \frac{1}{N} \sum_{l=1}^{N} \left| \iota\left(\pi_l^{\mathrm{true}}(\sigma_z)\right) - \iota\left(\pi_l^{\mathrm{nom}}(\sigma_z)\right) \right| \\ \Delta^{\mathrm{rs}} &:= \frac{1}{N} \sum_{l=1}^{N} \left| \iota\left(\pi_l^{\mathrm{true}}(\sigma_z)\right) - \iota\left(\widehat{\sigma_z}^{\mu,\mathrm{sub}}(l)\right) \right|. \end{aligned} \tag{50}$$

The histogram for these values are depicted in Fig. 1(b) for 200 sample paths with $\lambda^2 = 0.001$ and $N = 2000$. Overall, Δ^{rs} is smaller than Δ^{rn}, showing the better performance of the risk-sensitive estimator over the risk-neutral one. Fig. 1(c) and (d) illustrate an example of sample paths of the estimators; in Fig. 1(c) the solid line shows $\iota\left(\widehat{\sigma_z}^{\mu,\mathrm{sub}}(l)\right)$, while in Fig. 1(d) the solid line is $\iota\left(\pi_l^{\mathrm{nom}}(\sigma_z)\right)$. In both figures, the thick dotted line is $\iota\left(\pi_l^{\mathrm{true}}(\sigma_z)\right)$. In Fig. 1(c), both estimators are quite close to each other in spite of the difference in their initial states. On the other hand, as depicted in Fig. 1(d), the nominal risk-neutral estimator fails in the estimation, although it finally converges to the true value -1. As a summary, the risk-sensitive estimator outperforms the nominal risk-neutral estimator in the presence of uncertainty.

Remark 7.1: The performance of the nominal estimator depends on the magnitude of uncertainty. For example, if there is no uncertainty in the nominal distribution, the nominal risk-neutral estimator coincides with the true optimal estimator and clearly works better than the risk-sensitive one. However, under the existence of some uncertainty, the risk-neutral estimator is no longer optimal and will be inferior to the risk-sensitive one, as seen in Fig. 1. To make a more quantitative observation, we consider the following nominal distribution characterized by one parameter $\beta \in [0, 1]$ (this means $0 \leq \beta \leq 1$) that represents the uncertainty magnitude:

$$\begin{aligned} \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{1,1} &= \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{9,9} = \cdots = \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{20,20} = 0.05\beta \\ \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{2,2} &= \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{8,8} = 0.04\beta + 0.01 \\ \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{3,3} &= \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{7,7} = 0.01\beta + 0.04 \\ \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{4,4} &= \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{6,6} = -0.05\beta + 0.1 \\ \left(\rho_{\mathrm{p}}^{\mathrm{nom},\beta}\right)_{5,5} &= -0.65\beta + 0.7 \\ \rho_{\mathrm{s}}^{\mathrm{nom},\beta} &= \begin{bmatrix} 0.5 & 0.5 - 0.25\beta \\ 0.5 - 0.25\beta & 0.5 \end{bmatrix}. \end{aligned}$$

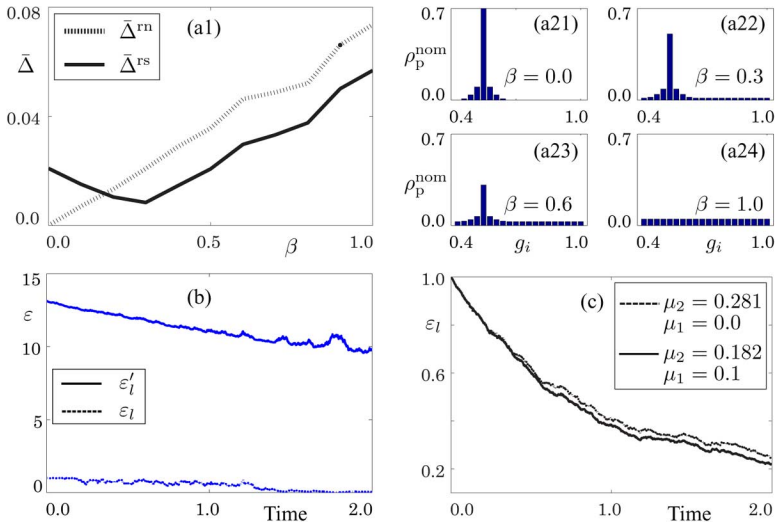

Fig. 2. For the dispersive interaction model of the atom, (a1) the averaged total estimation errors of the nominal risk-sensitive filter (solid line) and the nominal risk-neutral one (dotted line), (a2) examples of the nominal parameter distributions, (b) the conditional error (dotted line) and the guaranteed error bound (solid line) in Theorem 6.2, and (c) the averaged conditional estimation errors of the nominal risk-sensitive filters with $(\mu_1, \mu_2) = (0.0, 0.281)$ (upper dotted line) and $(\mu_1, \mu_2) = (0.1, 0.182)$ (lower solid line).

When $\beta = 0$, the nominal distribution is equal to the true one; $\rho_{\rm p}^{{\rm nom},0} \otimes \rho_{\rm s}^{{\rm nom},0} = \rho_{\rm p}^{\rm true} \otimes \rho_{\rm s}^{\rm true}$. Hence $\beta = 0$ implies there is no uncertainty. On the other hand when $\beta = 1$, the nominal distribution is the one given in (49). We consider the nominal risk-neutral estimator and the risk-sensitive one with $\mu_1 = 0.01$, $\mu_2 = 0.05$. Note that these two estimators are close to each other due to the small risk-sensitive parameter. To evaluate their performances, we calculate the averaged total estimation errors (50) and compare them. In Fig. 2 (a1), the horizontal axis shows the uncertainty magnitude β, while the vertical axis shows the average of $\Delta^{\rm rn}$ and $\Delta^{\rm rs}$ over 100 sample paths, which are denoted by $\bar{\Delta}^{\rm rn}$ and $\bar{\Delta}^{\rm rs}$, respectively. Fig. 2 (a2) shows examples of the nominal parameter distribution $\rho_{\rm p}^{{\rm nom},\beta}$. The risk-sensitive estimator clearly shows a better performance than the risk-neutral one, except in the case of a small β.

Remark 7.2: The robustness property of the risk-sensitive estimator is based on the fact that the estimation error is upper bounded, as presented in Theorems 6.1 and 6.2. Fig. 2(b) illustrates sample paths of the conditional estimation error and its upper bound given in Theorem 6.2:

$$\varepsilon_l := \iota\left(\mathbb{P}_{\rm true}\left[|j_l(X_{\rm e}) - {\rm u}_l|^2 \middle| \mathcal{Y}_l\right]\right)$$
$$\varepsilon'_l := \frac{1}{\mu_2}\log {\rm Tr}\left[\rho_l^{\mu,{\rm nom}} {\rm e}^{\mu_2 |X_{\rm e} - u_l|^2}\right] + \frac{1}{\mu_2} R\left(\rho_l^{\rm true} || \rho_l^{\mu,{\rm nom}}\right). \quad (51)$$

From this, we see that the bound is much larger than the actual estimation error. This is very similar to the classical case where one also often finds a very conservative upper bound.

Remark 7.3: As in the classical case, there is no theoretical procedure to determine the best risk-sensitive parameters (μ_1, μ_2). We here only maintain that a non-zero μ_1, the weighting parameter of the running estimation error cost, is actually helpful in obtaining a high-quality risk-sensitive estimator. To show this fact, we apply a nominal risk-sensitive filter with $\mu_1 = 0$, initialized to (49), to the same uncertain system as discussed above. For this filter, $\mu_2 = 0.281$ appears to be the best parameter. Fig. 2(c) illustrates the mean values over 200 sample paths of the conditional estimation error ε_l in (51). The upper dotted line and lower solid one corresponds to the risk-sensitive estimation with $(\mu_1, \mu_2) = (0.0, 0.281)$ and $(\mu_1, \mu_2) = (0.1, 0.182)$, respectively. This shows that a non-zero μ_1 does improve the performance of the estimator.

B. Spontaneous Emission Model

In the case of spontaneous decay, the interaction Hamiltonian (13) is given by

$$L_1 = 0, \quad L_2 = i\sqrt{e}\sigma_-, \quad L_3 = 0$$

where σ_- is defined in (10) and $e > 0$ represents the emission rate. The matrices M^i, $(i = \pm, +, -, \circ)$ are determined from (14) and (16) and read

$$M^{\pm}(e) = (1 - \cos(e\lambda))\sigma_z, \quad M^{+}(e) = \frac{\sin(e\lambda)}{\lambda}\sigma_-$$
$$M^{-}(e) = -\frac{\sin(e\lambda)}{\lambda}\sigma_+, \quad M^{\circ}(e) = \frac{\cos(e\lambda) - 1}{\lambda^2}\sigma_+\sigma_-.$$

Similar to the dispersive interaction case, we here assume that e behaves as a classical discrete random variable with an unknown probability distribution; e is then replaced by an observable $\mathfrak{e} = {\rm diag}\{e_1, \ldots, e_m\} \in \mathcal{P}$. In particular, we assume $e_i = 0.2 + 0.04i$, $(i = 1, \ldots, 20)$ with the true density matrix

$$\rho_{\rm p}^{\rm true} = {\rm diag}\{0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0.01,0.04,0.9,0.04,0.01,0\}$$

which is illustrated in Fig. 3(a). The system true density matrix $\rho_{\rm s}^{\rm true}$ is given by (48). For a nominal density matrix, we take $\rho_{\rm p}^{\rm nom}$ in (49) and assume that $\rho_{\rm s}^{\rm nom} = \rho_{\rm s}^{\rm true}$. In the above setting, we consider estimating $X_{\rm e} = \sigma_y := i(\sigma_- - \sigma_+)$, investigate the performance of the nominal risk-sensitive filter, and

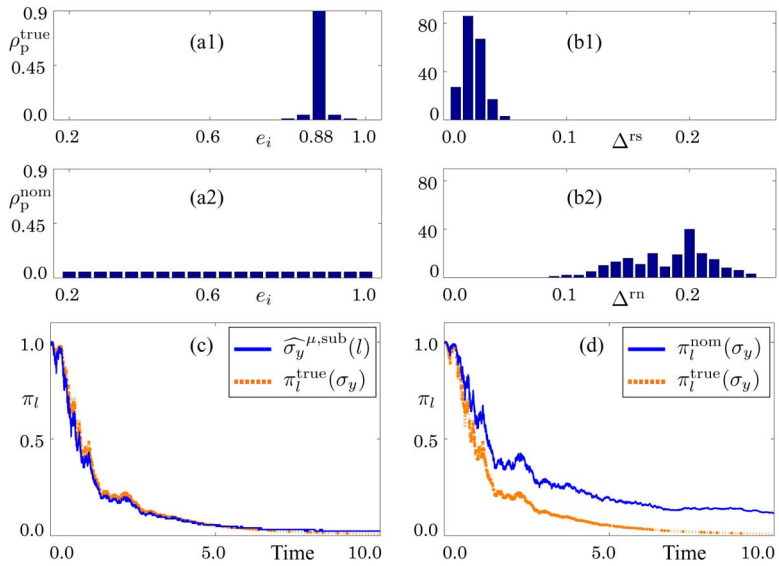

Fig. 3. For the spontaneous emission model of the atom, (a) the true and nominal parameter distributions, (b) the histogram of the averaged total estimation errors, (c) sample paths of the nominal risk-sensitive estimator (solid line) and the true risk-neutral one (thick dotted line), and (d) sample paths of the nominal risk-neutral estimator (solid line) and the true risk-neutral one (thick dotted line). For the figures (c) and (d), the notation (ι) is omitted.

compare it with the nominal risk-neutral one. The risk-sensitive parameters are chosen as $(\mu_1, \mu_2) = (0.15, 0.25)$. Fig. 3(b) shows the histogram for the averaged error $\Delta^{\rm rs}$ and $\Delta^{\rm rn}$ for 200 sample paths with $\lambda^2 = 0.001$ and $N = 5000$. Fig. 3(c) shows the sample paths of $\iota(\pi_l^{\rm true}(\sigma_y))$ and $\iota\left(\widehat{\sigma_y}^{\mu,{\rm sub}}(l)\right)$, while in Fig. 3(d) $\iota\left(\pi_l^{\rm true}(\sigma_y)\right)$ and $\iota\left(\pi_l^{\rm nom}(\sigma_y)\right)$ are shown. These figures clearly show that the nominal risk-sensitive estimator is superior to the nominal risk-neutral estimator.

Remark 7.4: While in this paper we have considered the estimation problem over the finite-time horizon, let us here look at the asymptotic behaviour as $l \to \infty$ of the quantity

$$\delta_l = \left| l\left(\pi_l^{\rm true}(X_{\rm e})\right) - l\left(\pi_l^{\rm nom}(X_{\rm e})\right) \right|$$

where $\pi_l^{\rm true}(X_{\rm e})$ and $\pi_l^{\rm nom}(X_{\rm e})$ correspond to the standard risk-neutral estimator for the true and nominal initial states, respectively. If $\lim_{l\to\infty} \delta_l = 0$ for all observables $X_{\rm e}$, then we say the filter is *stable*. Recently, Van Handel [22] has provided the following characterization for filter stability in continuous time. That is, for all $X_{\rm e}$ included in the *observable space*

$$\mathcal{O} = {\rm span}\left\{\mathcal{L}^{c_1}\mathcal{J}^{d_1}\mathcal{L}^{c_2}\cdots\mathcal{L}^{c_k}\mathcal{J}^{d_k}(I) : k, c_i, d_i \geq 0\right\}$$

we have $\delta_l \to 0$. Here, $\mathcal{L}$ and $\mathcal{J}$ are the continuous time analogues of the quantities defined in (26). Therefore, the filter is stable if $\dim\mathcal{O} = \dim\mathcal{A}$.

In our examples the observable spaces are given by

$$\mathcal{O}^{\rm dis} = {\rm span}\{I,\ \sigma_z\}, \quad \mathcal{O}^{\rm spon} = {\rm span}\{I,\ \sigma_x,\ \sigma_z\}.$$

Therefore, for a dispersive interaction where we estimate $\sigma_z \in \mathcal{O}^{\rm dis}$, it is guaranteed by Van Handel's theorem that $\pi_l^{\rm nom}(\sigma_z)$ with any initial state converges to the true estimator. On the other hand, in the spontaneous decay case, due to $\sigma_y \notin \mathcal{O}^{\rm spon}$, we cannot expect that $\delta_l \to 0$. This could be the reason why the increase in performance by the nominal risk-sensitive estimator over the risk-neutral one is more pronounced in Fig. 3 than in Fig. 1. We must note here that in simulations we do see that, with the settings used in Fig. 3, $\iota\left(\pi_l^{\rm nom}(\sigma_y)\right)$ eventually converges to the true value 0. However, this convergence is very slow.

ACKNOWLEDGMENT

The authors wish to thank I. R. Petersen and V. A. Ugrinovskii for insightful discussion and for bringing the quantum duality relation to our attention. We also thank M. R. James for helpful comments on simulation results.

REFERENCES

[1] C. Ahn, A. C. Doherty, and A. J. Landahl, "Continuous quantum error correction via quantum feedback control," *Phys. Rev. A*, vol. 65, p. 04-2301, 2002.

[2] S. Attal, "Approximating the Fock space with the toy Fock space," in *Seminaire de Probabilites, XXXVI*. Berlin, Germany: Springer, 2003, vol. 1801, pp. 477–491.

[3] S. Attal and Y. Pautrat, "From repeated to continuous quantum interactions," *Ann. Henri Poincare*, vol. 7, pp. 59–104, 2006.

[4] V. P. Belavkin, "Quantum filtering of markov signals with white quantum noise," *Radiotechnika i Electronika*, vol. 25, pp. 1445–1453, 1980.

[5] V. P. Belavkin, "Quantum stochastic calculus and quantum nonlinear filtering," *J. Multivariate Anal.*, vol. 42, pp. 171–201, 1992.

[6] V. P. Belavkin, "Nondemolition measurements, nonlinear filtering, and dynamic programming of quantum stochastic processes," in *Proc. Bellman Continuum, Sophia-Antipolis*, ser. Lecture Notes in Control and Information Sciences. New York: Springer-Verlag, 1988, vol. 121, pp. 245–265.

[7] R. K. Boel, M. R. James, and I. R. Petersen, "Robustness and risk-sensitive filtering," *IEEE Trans. Automat. Control*, vol. 47, no. 3, pp. 451–461, Mar. 2002.

[8] L. M. Bouten, M. I. Guta, and H. Maassen, "Stochastic Schrodinger equations," *J. Phys. A*, vol. 37, pp. 3189–3209, 2004.

[9] L. M. Bouten, R. Van Handel, and M. R. James, "A Discrete Invitation to Quantum Filtering and Feedback Control," *SIAM Rev.*, to be published.

[10] L. M. Bouten and R. Van Handel, "On the separation principle of quantum control," in *Quantum Stochastics and Information: Statistics, Filtering and Control*. London, U.K.: World Scientific, 2008, pp. 206–238.

[11] T. D. Brun, "A simple model of quantum trajectories," *Amer. J. Phys.*, vol. 70, pp. 719–737, 2002.

[12] H. J. Carmichael, *An Open Systems Approach to Quantum Optics*. Berlin, Germany: Springer-Verlag, 1993.

[13] C. D. Charalambous, S. Dey, and R. J. Elliott, "New finite-dimensional risk-sensitive filters: Small noise limits," *IEEE Trans. Automat. Control*, vol. 43, no. 10, pp. 1424–1429, Oct. 1998.
[14] C. D'Helon, A. C. Doherty, M. R. James, and S. D. Wilson, "Quantum risk-sensitive control," in *Proc. 45th IEEE Conf. Decision Control*, San Diego, CA, 2006, pp. 3132–3137.
[15] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing via reference probability methods," *IEEE Trans. Automat. Control*, vol. 42, no. 11, pp. 1587–1591, Nov. 1997.
[16] A. C. Doherty and K. Jacobs, "Feedback control of quantum systems using continuous state-estimation," *Phys. Rev. A*, vol. 60, p. 2700, 1999.
[17] P. Dupuis and R. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: Wiley, 1997.
[18] C. Gardiner and P. Zoller, *Quantum Noise*, 3rd ed. Berlin, Germany: Springer, 2004.
[19] S. Golden, "Lower bounds for the helmholtz function," *Phys. Rev.*, vol. 137, pp. B1127–B1128, 1965.
[20] J. Gough and Sobolev, "Stochastic schrodinger equations as limit of discrete filtering," *Open Syst. Inf. Dyn.*, vol. 11, pp. 235–255, 2004.
[21] R. V. Handel, J. K. Stockton, and H. Mabuchi, "Feedback control of quantum state reduction," *IEEE Trans. Automat. Control*, vol. 50, no. 6, pp. 768–780, Jun. 2005.
[22] R. V. Handel, "The stability of quantum Markov filters," *Infin. Dimens. Anal. quant. Probab. Rel. Top.*, to be published.
[23] R. L. Hudson and K. R. Parthasarathy, "Quantum Ito's formula and stochastic evolution," *Commun. Math. Phys.*, vol. 93, pp. 301–301, 1984.
[24] M. R. James, "A quantum langevin formulation of risk-sensitive optimal control," *J. Opt. B: Quantum Semiclass. Opt.*, vol. 7, p. 198, 2005.
[25] M. R. James, J. S. Baras, and R. J. Elliott, "Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems," *IEEE Trans. Automat. Control*, vol. 39, no. 4, pp. 780–792, Apr. 1994.
[26] O. Kallenberg, *Foundations of Modern Probability*. Berlin, Germany: Springer, 1997.
[27] G. Kallianpur and C. Striebel, "Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors," *Annu. Math. Statist.*, vol. 39, pp. 785–801, 1968.
[28] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, Ser. D, J. Basic Eng.*, vol. 82, pp. 35–45, 1960.
[29] H. J. Kushner, "Differential equations for optimal nonlinear filtering," *J. Differential Equations*, vol. 3, pp. 179–190, 1967.
[30] J. M. Lindsay and K. R. Parthasarathy, "The passage from random walk to diffusion in quantum probability ii," *Sankhya Ser. A*, vol. 50, pp. 151–170, 1988.
[31] M. Mirrahimi and R. V. Handel, "Stabilizing feedback controls for quantum systems," *SIAM J. Control Optim.*, vol. 46, pp. 445–467, 2007.
[32] J. B. Moore, R. J. Elliott, and S. Dey, "Risk-sensitive generalization of minimum variance estimation and control," *J. Math. Syst., Estim., Control*, vol. 7, pp. 1–15, 1997.
[33] M. Ohya and D. Petz, *Quantum Entropy and Its Use*. Berlin, Germany: Springer-Verlag, 1991.
[34] I. R. Petersen and A. V. Savkin, *Robust Kalman Filtering for Signals and Systems With Large Uncertainties*. Boston, MA: Birkhauser, 1999.
[35] S. Sakai, *C^*-Algebras and W^*-Algebras*. Berlin, Germany: Springer-Verlag, 1998.
[36] J. L. Speyer, C. Fan, and N. Banavar, "Optimal stochastic estimation with exponential cost criteria," in *Proc. 31st IEEE Conf. Decision Control*, 1992, pp. 2293–2298.
[37] J. M. Steele, *Stochastic Calculus and Financial Applications*. Berlin, Germany: Springer, 2001.
[38] J. K. Stockton, "Continuous Quantum Measurement of Cold Alkali-Atom Spins," Ph.D. dissertation, California Institute of Technology, Pasadena, 2006.
[39] C. J. Thompson, "Inequality with applications in statistical mechanics," *J. Math. Phys.*, vol. 6, pp. 1812–1813, 1965.
[40] L. Thomsen, S. Mancini, and H. M. Wiseman, "Continuous quantum nondemolition feedback and unconditional atomic spin squeezing," *J. Phys. B*, vol. 35, p. 4937, 2002.
[41] J. Tomiyama, "On the projections of norm one in W^*-algebras," in *Proc. Jpn. Acad.*, 1957, vol. 33, pp. 608–612.
[42] V. A. Ugrinovskii and I. R. Petersen, "Robust filtering of stochastic uncertain systems on an infinite time horizon," *Int. J. Control*, vol. 75, pp. 614–626, 2002.
[43] H. Umegaki, "Conditional expectations in an operator algebra," *Tohoku Math. J.*, vol. 6, pp. 177–181, 1954.
[44] W. M. Wonham, "On the separation theorem of stochastic control," *SIAM J. Control Optim.*, vol. 6, pp. 312–326, 1968.
[45] L. Xie and Y. C. Soh, "Robust Kalman filtering for uncertain systems," *Syst. Control Lett.*, vol. 22, pp. 123–129, 1994.
[46] N. Yamamoto, "Robust observer for uncertain linear quantum systems," *Phys, Rev. A*, vol. 74, p. 032107, 2006.
[47] N. Yamamoto, K. Tsumura, and S. Hara, "Feedback control of quantum entanglement in a two-spin system," *Automatica*, vol. 43, pp. 981–992, 2007.
[48] M.-G. Yoon, V. A. Ugrinovskii, and I. R. Petersen, "Robust finite horizon minimax filtering for discrete-time stochastic uncertain systems," *Syst. Control Lett.*, vol. 52, pp. 99–112, 2004.

Naoki Yamamoto received the B.S. degree in engineering, and the M.S. and Ph.D. degrees in information physics and computing from the University of Tokyo, Tokyo, Japan, in 1999, 2001, and 2004, respectively.

From 2003 to 2007, he was a Research Fellow of the Japan Society for the Promotion of Science. From 2004 to 2006, he was a Visiting Researcher at the California Institute of Technology, Pasadena, and from 2007 to 2008 he was a Research Fellow at Australian National University, Canberra, Australia. He is currently an Assistant Professor at Keio University, Yokohama, Japan. His current research interests are in quantum control, quantum filtering, and quantum information theory.

Luc Bouten received the M.Sc. degree in mathematics, the M.Sc. degree in physics, and the Ph.D. degree in mathematics from Radboud University, Nijmegen, The Netherlands, in 2000, 2002, and 2004, respectively.

Currently, he is a Post-Doctoral Researcher in the Physics Department, California Institute of Technology, Pasadena. His research focuses on quantum filtering and feedback control, singular perturbation of quantum stochastic differential equations, and applications to quantum optics.