Open access • Journal Article • DOI:10.1137/060651239

An Introduction to Quantum Filtering

Luc Bouten, Ramon van Handel, Matthew R. James

Institutions: California Institute of Technology, Australian National University

Published on: 01 Nov 2007 - Siam Journal on Control and Optimization (Society for Industrial and Applied Mathematics)

Topics: Quantum probability, Open quantum system, Quantum algorithm, Quantum process and Quantum operation

SIAM J. CONTROL OPTIM.
Vol. 46, No. 6, pp. 2199–2241

AN INTRODUCTION TO QUANTUM FILTERING*

LUC BOUTEN†, RAMON VAN HANDEL†, AND MATTHEW R. JAMES‡

Abstract. This paper provides an introduction to quantum filtering theory. An introduction to quantum probability theory is given, focusing on the spectral theorem and the conditional expectation as a least squares estimate, and culminating in the construction of Wiener and Poisson processes on the Fock space. We describe the quantum Itô calculus and its use in the modeling of physical systems. We use both reference probability and innovations methods to obtain quantum filtering equations for system-probe models from quantum optics.

Key words. quantum filtering, quantum probability, quantum stochastic processes

AMS subject classifications. 93E11, 81P15, 81S25, 81Q10, 81R15, 34F05

DOI. 10.1137/060651239

1. Introduction.

Since even before the industrial revolution, feedback control has played a major role in the development of technology. Nowadays, many machines and devices that make up our everyday lives use feedback to provide efficient and reliable performance despite ever increasing complexity and miniaturization, and a rich control theory has been developed to aid in the design of feedback controllers based on device models from classical physics. As microtechnology is making way for nanotechnology, however, we are now rapidly approaching the boundary of the classical world past which the effects of quantum mechanics cannot be neglected.

The laws of quantum mechanics tell us that any description of the phenomena at small scales is inherently nondeterministic in nature. This opens new areas of application for stochastic control theory, which could play a role in a future generation of technology. In particular, as observations of quantum systems are inherently noisy, the theory of filtering—the extraction of information from a noisy signal—forms an integral part of quantum feedback control theory.

Quantum filtering was already implicit in early work on quantum measurement theory by Davies in the 1960s [24, 25]. In its modern form, the study of quantum filtering and control was pioneered by Belavkin in a series of articles dating back to the early 1980s [9, 10, 11, 12, 13]. The theory developed by Belavkin provides an essential foundation for statistical inference in, e.g., quantum optical systems, and much of what we will discuss in the second half of this article is based on his work. The theory gained popularity in the physics community after it was independently developed on a more heuristic level by Carmichael in the early 1990s [22] under the name "quantum trajectory theory" and has since been widely applied in the description of quantum optical experiments and as a computational tool.

*Received by the editors January 30, 2006; accepted for publication (in revised form) June 20, 2007; published electronically December 21, 2007.
http://www.siam.org/journals/sicon/46-6/65123.html

†Physical Measurement and Control 266-33, California Institute of Technology, Pasadena, CA 91125 (bouten@its.caltech.edu, ramon@its.caltech.edu). The research of these authors are supported by Army Research Office grant DAAD19-03-1-0073. The first author is additionally supported by National Science Foundation grant PHY-0456720.

‡Department of Engineering, Australian National University, Canberra, ACT 0200, Australia (matthew.james@anu.edu.au). The research of this author is supported by the Australian Research Council.

Based on the foundations of quantum filtering theory, methods from classical nonlinear and stochastic control can be developed and applied to design feedback control laws for quantum systems. These methods may be optimal in some sense or otherwise designed with relevant considerations in mind (e.g., stability). The resulting controllers are intended to be implemented with some classical technology (e.g., digital or analog electronics). Recent experiments implementing quantum feedback controls [3, 35, 58, 21] have led to renewed interest in the field, which is now rapidly expanding [10, 64, 28, 27, 17, 43, 15, 61, 16, 44, 37, 30, 60, 62, 18]. We believe that a fruitful interaction between stochastic control and theoretical and experimental physics will be essential in paving the way towards the engineering of quantum technologies.

This paper provides an introduction to quantum filtering theory. There are three key ingredients that are required for the development of the theory. First, we need to capture both classical probability and quantum mechanics within the framework of a generalized probability theory, called noncommutative or quantum probability theory. The central object in this theory, the spectral theorem, provides a link between quantum systems and the associated probabilistic measurement outcomes. Second, we need a noncommutative generalization of the concept of conditional expectations. As in classical probability, we will find that a suitably restricted definition of the quantum conditional expectation is none other than a least squares estimator, which elucidates its role in quantum filtering theory. Finally, we need a noncommutative analog of stochastic calculus and quantum stochastic differential equations (QSDEs). This provides a broad class of models for which we can obtain filtering equations.

A typical physical scenario, to which the theory that we will develop can be applied, is illustrated schematically in Figure 5.1. A cloud of (usually cold, trapped) atoms interacts with the electromagnetic field in free space; this can be coherent light from a laser, or even the vacuum. Depending on their internal state the atoms can, for example, emit radiation into the field. If we detect this radiation using an optical detection setup, we can try to infer some information on the internal state of the atoms—this is precisely the goal of quantum filtering theory. If we wanted to control the state of the atoms, we could then feed back some function of the state estimates through a suitable actuator. Recent laboratory experiments (e.g., [58]) implement precisely such a setup and provide a motivating example for the theory.

We begin in section 2 by providing some background for quantum filtering. This includes a discussion of the quantum mechanics and quantum probability in the simplest, finite-dimensional context. In section 3 quantum probability is developed in detail. Then in section 4 we show how Wiener and Poisson processes emerge in a particular quantum probabilistic model based on the Fock space, and how these can be used to develop a noncommutative stochastic calculus. In section 5 we introduce a class of system-observation models that describe typical experiments in quantum optics. Section 6 deals with the derivation of quantum filtering equations using the reference probability approach, while section 7 gives an alternative derivation using the innovations or martingale method.

Scope. It has been our aim to make quantum probability and filtering theory accessible, modulo a set of technicalities, to readers with a minimal number of prerequisites. We (only) presume some familiarity with probability theory and elementary functional analysis. We have put an emphasis on introducing the mathematical structures of quantum probability theory and on demonstrating their significance and their use. As a consequence we do not everywhere achieve the highest level of rigor; we are particularly lax in the use of unbounded operators and their domains. It is

our hope that skimming over these technicalities has enabled us to paint a clearer picture of the pillars of the theory and of the essential techniques involved. That being said, we should point out that many of the tools described in this paper are applied regularly and successfully by physicists without paying any attention to the technical issues involved; the reader should not hesitate to get his feet wet!

It is an ambitious project to introduce an unfamiliar probability theory, a new stochastic calculus, and to even solve a nontrivial problem (filtering) within the confines of about 40 pages. Though we have tried to give a pedagogical treatment, the explanations are sometimes necessarily terse; we hope that the reader will be sufficiently compelled to work his way through the paper. Needless to say there are many omissions; one that particularly deserves mention is the linear case: indeed, the quantum Kalman filter, and the corresponding theory of quantum LQG control, can be developed along similar lines to the filters we will discuss. We have chosen to omit this topic in order to avoid the technicalities of QSDEs with unbounded coefficients, but refer instead to [30] and the references therein.

Notation. The sets of natural, real, and complex numbers are denoted $\mathbf{N}$, $\mathbf{R}$, and $\mathbf{C}$, respectively. In general, script symbols (e.g., $\mathscr{Y}$) are used for von Neumann algebras, while calligraphic symbols (e.g., $\mathcal{Y}$) stand for σ-algebras. $\mathcal{B}$ is the Borel σ-algebra on $\mathbf{R}$. Classical probability spaces are denoted as $(\Omega, \mathcal{F}, \mathbf{P})$, and $E_{\mathbf{P}}$ denotes the expectation with respect to the measure $\mathbf{P}$. Blackboard symbols (e.g., $\mathbb{P}$) denote states on von Neumann algebras. Sans serif symbols (e.g., H) are used for Hilbert spaces. Hilbert space adjoints, as well as the scalar complex conjugate, are indicated by *, and the Hilbert space inner product is denoted by $\langle \cdot, \cdot \rangle$. The commutator of two bounded operators is denoted by $[X, Y] = XY - YX$. I is the identity operator.

2. Background and motivation.

In this article we adopt a modern quantum probability formulation of quantum mechanics. *Quantum probability* is the noncommutative counterpart of Kolmogorov's axiomatic characterization of classical probability theory. In addition to the natural interpretation and mathematical tools provided by Kolmogorov's formalism, one of its major successes is that conditioning is a derived concept rather than an additional axiom. The situation is much the same in quantum probability; in particular, the conditioning axiom or "projection postulate," as it is traditionally posed in quantum mechanics, can emerge as a consequence of conditional expectation and the physical idea that in a single experiment one only has direct access to information contained in a commutative subalgebra of observables.

Considering the success of the classical (Kolmogorov) theory, it should come as no surprise that the mathematical abstraction provided by the framework of quantum probability pays off significantly (as we will see throughout the article). Introductory physics textbooks on quantum mechanics rarely use such a description, however. In this section we introduce the basic concepts of quantum probability in their simplest form and attempt to provide contact with ideas about quantum mechanics that readers may be familiar with. This is intended to provide a reference point for interpreting the quantum probabilistic framework used in this paper.

2.1. Some textbook quantum mechanics.

According to the textbook by Merzbacher [51, p. 1], "Quantum mechanics is the theoretical framework within which it has been found possible to describe, correlate, and predict the behavior of a vast range of physical systems, from particles through nuclei, atoms, and radiation to molecules and condensed matter." Central to quantum mechanics are the notions of *observables*, which are mathematical representations of physical quantities that can (in principle) be measured, and *states*, which summarize the status of physical systems

and permit the calculation of statistical quantities (such as probabilities, expectations, correlations) of observables.

Indeed, the reader may be familiar with the *Schrödinger wavefunction* $\psi(q,t)$ for a particle of mass m moving in a force field $V(q)$ (dependent on position q, in one dimension for simplicity). If Q is the observable representing position (defined in Example 3.9), the expected position of the particle when in a state described by $\psi(q,t)$ at time t is defined to be

$$\langle Q\rangle = \int q|\psi(q,t)|^2 dq. \tag{2.1}$$

The wavefunctions are normalized to one $\int |\psi(q,t)|^2 dq = 1$, so that $|\psi(q,t)|^2$ could be interpreted as the probability density of the position of the particle. The dynamics of the particle are described by the famous *Schrödinger wave equation*

$$i\hbar\frac{\partial\psi(q,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(q,t)}{\partial q^2} + V(q)\psi(q,t), \tag{2.2}$$

where $\hbar = h/2\pi$, h is *Planck's constant*, and $i^2 = -1$.

The key distinction between classical (i.e., nonquantum) and quantum mechanics is that quantum mechanics is *noncommutative*, meaning that there exist observables that do not commute, a fact which has deep implications. The momentum observable P (defined in Example 3.9) does not commute with the position observable Q; in fact $[Q,P] = QP - PQ = i\hbar I$. The most famous implication of this failure of commutativity is *Heisenberg's uncertainty relation*, which asserts that

$$\Delta Q\Delta P \geq \frac{1}{2}|\langle i[Q,P]\rangle| = \frac{\hbar}{2}, \tag{2.3}$$

where the variances are defined by $\Delta Q = (\langle Q^2\rangle - \langle Q\rangle^2)^{1/2}$, $\Delta P = (\langle P^2\rangle - \langle P\rangle^2)^{1/2}$. Naive interpretation of the Heisenberg uncertainty relation can be misleading; we will discuss its precise meaning in the following section. Nonetheless, it evidently implies that there is a fundamental irreducible randomness in quantum mechanics. This is in contrast to classical randomness, which in principle can be eliminated with enough effort and information. Experimental evidence has repeatedly confirmed the irreducible randomness of quantum mechanical observations.

Let us make this somewhat vague discussion a little more precise. For simplicity, we will work in this section only in a finite-dimensional setting (in which observations can only take a finite number of values; i.e., they are finite-state random variables). First, recall that if $A = A^*$ is a self-adjoint operator on a finite-dimensional Hilbert space $\mathsf{H} = \mathbf{C}^n$, it has at most n (distinct) real eigenvalues. The set $\mathrm{spec}(A) = \{a_j\}$ of eigenvalues of A is called the spectrum of A, and A can be written as

$$A = \sum_{a\in\mathrm{spec}(A)} a\,P_a, \tag{2.4}$$

where P_a is the projection operator onto the subspace of H spanned by vectors with eigenvalue a. The projections resolve the identity $\sum_{a\in\mathrm{spec}(A)} P_a = I$.

In this finite-dimensional setting, the following operational characterization of quantum mechanical models (often referred to as the "postulates" of quantum mechanics) can be found in most introductory textbooks.

Observables. Physical quantities like position, momentum, spin, etc. are represented by self-adjoint operators on the Hilbert space H and are called *observables*. These are the noncommutative counterparts of random variables.

States. A state is meant to provide a summary of the status of a physical system that enables the calculation of statistical quantities associated with observables. A generic state is specified by a *density matrix* ρ, which is a self-adjoint operator on H that is positive $\rho \geq 0$ and normalized $\mathrm{Tr}[\rho] = 1$. This is the noncommutative counterpart of a probability density.

Measurement. A *measurement* is a physical procedure or experiment that produces numerical results related to observables. In any given measurement, the allowable results take values in the spectrum $\mathrm{spec}(A)$ of a chosen observable A. Given the state ρ, the value $a \in \mathrm{spec}(A)$ is observed with probability $\mathrm{Tr}[\rho P_a]$. Consequently, the expectation of an observable A is given by $\langle A \rangle = \mathrm{Tr}[\rho A]$.

Conditioning. Suppose that a measurement of A gives rise to the observation $a \in \mathrm{spec}(A)$. Then we must condition the state in order to predict the outcomes of subsequent measurements by updating the density matrix ρ using

$$\rho \mapsto \rho'[a] = \frac{P_a \rho P_a}{\mathrm{Tr}[\rho P_a]}. \tag{2.5}$$

This is known as the "projection postulate."

Evolution. A *closed* (i.e., isolated) quantum system evolves in a *unitary* fashion: a physical quantity that is described at time $t = 0$ by an observable A is described at time $t > 0$ by $A(t) = U(t)^* A U(t)$, where $U(t)$ is a unitary operator for each time t. The unitary is generated by the *Schrödinger equation*

$$i\hbar \frac{d}{dt} U(t) = H(t) U(t), \tag{2.6}$$

where the (time-dependent) Hamiltonian $H(t)$ is a self-adjoint operator for each t.

Before continuing, we make the following remarks.

Remark 2.1. *Pure states.* The set of density matrices ρ is convex; we can thus wonder what are the extremal points in this set, i.e., those that correspond to the most informative states. It is not difficult to show that the set of extremal density matrices is the set of projections onto one-dimensional subspaces. Thus we can specify any extremal state uniquely (up to a phase factor $e^{i\varphi}$) by a single unit vector $\psi \in \mathsf{H}$ in the corresponding subspace, and $\mathrm{Tr}[\rho X] = \langle \psi, X\psi \rangle$ for any operator X. In classical probability theory, the set of probability measures is also convex and the extremal measures are deterministic (Dirac) measures. In the quantum mechanical setting, on the other hand, the Heisenberg uncertainty relation implies that even extremal states do not give deterministic measurement outcomes for all observables.

Historically, and in most textbooks, quantum mechanics is first formulated in terms of the extremal states (called *pure states*) and the description is later generalized to density matrices (*mixed states*). The Schrödinger wavefunction $\psi(q, t)$ is an example of a pure state vector in an infinite-dimensional Hilbert space setting. □

Remark 2.2. *Heisenberg vs. Schrödinger picture.* In the above description of time evolution we work with a fixed state while the observables change in time. This conforms to the usual treatment in classical probability theory, where the underlying probability measure is fixed at the outset and the random variables are time dependent (stochastic processes). In quantum mechanics this is known as the *Heisenberg picture*; equally (or perhaps more) popular is the *Schrödinger picture*, in which the observables

are considered fixed and the density matrix evolves as $\rho(t) = U(t)\rho U(t)^*$. The two pictures are essentially equivalent as $\mathrm{Tr}[\rho A(t)] = \mathrm{Tr}[\rho(t)A]$ for any observable A.

Note that if we start in a pure state, then unitary evolution preserves this property; in terms of the state vector, $\psi(t) = U(t)\psi$. Intuitively, this enforces the physical idea that no information is lost from an isolated system. Together with (2.6) we obtain the traditional Schrödinger equation for $\psi(t)$, of which (2.2) is a special case (for a specific choice of H, in infinite dimensions). *We will always work in the Heisenberg picture*, however, as we will be dealing with (quantum) stochastic processes. ∎

As a basic illustration we discuss the following simple example.

Example 2.3. One of the classic experimental demonstrations of the necessity of quantum mechanics was performed in 1922 by Stern and Gerlach. A silver atom is subjected to an inhomogeneous magnetic field. The atom possesses an intrinsic magnetic moment and hence experiences a force that is proportional to the component of its magnetic moment in the direction of the field gradient. As Stern and Gerlach did not prepare the atom in a particular orientation, they expected it to be deflected randomly in a continuous range of directions corresponding to a random orientation of the magnetic moment. Repeated runs of the experiment showed, however, that the atom is randomly deflected into two discrete directions only—the reason being that in quantum mechanics the intrinsic magnetic moment (or spin) observable is discrete, rather than continuous. Atoms deflected in the upper direction are said to have "spin up," while those in the lower direction have "spin down."

A simple model of a spin is as follows. Let $\mathsf{H} = \mathbf{C}^2$, and consider the observable

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.7}$$

representing spin in the z direction. We have $\mathrm{spec}(\sigma_z) = \{-1, 1\}$, which correspond to spin down and spin up, respectively. In terms of the eigenprojections

$$P_{z,1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_{z,-1} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

we can write $\sigma_z = P_{z,1} - P_{z,-1}$. The next step is to introduce a state. Consider a pure state, given by the vector $\psi = (c_1\ c_{-1})^T$ with $|c_1|^2 + |c_{-1}|^2 = 1$. If we observe σ_z, we obtain the outcome 1 (spin up) with probability $\langle \psi, P_{z,1}\psi\rangle = |c_1|^2$, or the outcome -1 with probability $\langle \psi, P_{z,-1}\psi\rangle = |c_{-1}|^2$. ∎

2.2. A first look at quantum probability. The description of quantum mechanics in the previous section contains the rudiments of a viable probability theory. We will now formalize these ideas, once again restricting ourselves to the finite-dimensional case for simplicity (the general theory, which will be discussed in section 3, is conceptually very similar). Two key ideas, which we elaborate on below, form the essence of the formalism: the first is that a set of measurements made in a single realization[1] of a quantum experiment corresponds to a particular choice of a commutative algebra of observables; and the second is that any such commutative algebra is entirely equivalent to a classical (Kolmogorov) probability model.

A classical probability model is described by a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Here Ω, the sample space, is not of essential importance; the basic ingredients of the theory

[1]By a *realization* or an *experiment* we mean that random variables are assigned a definite value, as is the case if we perform measurements on a single physical system. In classical probability this corresponds to the choice of a sample point $\omega \in \Omega$; the quantum case is a little more subtle.

are the events that can occur, contained in the σ-algebra $\mathcal{F}$, and their probabilities, which are determined by the measure $\mathbf{P}$. Equivalently, we could describe an event $F \in \mathcal{F}$ by a random variable χ_F which takes the value 1 if F occurs and 0 otherwise (the indicator function on F), and the probability of the event is simply the expectation of χ_F. We have already encountered such objects in the previous section: events are precisely those observables that are projection operators ($P = P^* = P^2$), and the probability of an event P is given by $\mathbb{P}(P) = \mathrm{Tr}[\rho P]$. Thus the set of projections, together with the linear map $\mathbb{P}$, play much the same role as the classical pair $\mathcal{F}, \mathbf{P}$.

We run into trouble in the quantum case when we try to ascribe joint probabilities to certain events. This is always possible in classical probability theory: the joint probability of the events A and B is $\mathbf{P}(A \cap B) = E_{\mathbf{P}}(\chi_A \chi_B)$. But given two projection operators P, Q, the operator PQ is not guaranteed to be a projection or even an observable ($(PQ)^* = QP$), *unless* P and Q commute. This simple observation is no coincidence; it has the following physical interpretation: in a single realization of a quantum probability model, we can only verify the truth of a set of commuting events. This is in contrast with classical probability where in every realization any event is either true or false, whether we choose to observe it or not. In quantum probability we can a priori choose to verify the truth of an arbitrary event, but subsequently some of the other events (those that do not commute with the observed event, said to be *incompatible*) become meaningless within the same realization.

The incompatibility of events is a significant conceptual departure from classical probability and requires a little getting used to. In many ways, however, this is the only essential departure from classical probability theory. We now begin to construct the mathematical formalism of quantum probability, and we will show that it is indeed very close to Kolmogorov's theory.

Consider the following idea. Suppose we decide to measure an observable A and obtain a particular outcome $a \in \mathrm{spec}(A)$. Then we do not need to perform another measurement to know that any function $f(A)$ would give the outcome $f(a)$; in essence, this is merely a relabeling of the measurement outcomes of A. Indeed,

$$A = \sum_{a \in \mathrm{spec}(A)} a\, P_a \quad \Longrightarrow \quad f(A) = \sum_{a \in \mathrm{spec}(A)} f(a)\, P_a, \tag{2.8}$$

and all such operators commute with each other. Thus measuring A "automatically" measures all functions $f(A)$. The set of operators $\mathscr{A} = \{X : X = f(A),\ f : \mathbf{R} \to \mathbf{C}\}$ forms a *commutative $*$-algebra*, i.e., arbitrary (complex) linear combinations, products, and adjoints of operators in $\mathscr{A}$ are still in $\mathscr{A}$, $I \in \mathscr{A}$, and all elements of $\mathscr{A}$ commute. We will call $\mathscr{A}$ the $*$-algebra *generated*[2] *by* A. A linear map $\mathbb{P} : \mathscr{A} \to \mathbf{C}$ that is positive ($\mathbb{P}(A) \geq 0$ if $A \geq 0$) and normalized ($\mathbb{P}(I) = 1$) is called a *state* on $\mathscr{A}$ (clearly we can always write such a state as $A \mapsto \mathrm{Tr}[\rho A]$ for some density matrix ρ). Note that the projections $P \in \mathscr{A}$ are precisely those events that we can distinguish by measuring A, and $\mathbb{P}(P)$ gives their probabilities. We can similarly generate the commutative $*$-algebra of functions of an arbitrary set of commuting observables.

The algebraic structure we have introduced is of fundamental importance as it provides us with a direct connection to the classical theory, as follows.

THEOREM 2.4 (spectral theorem, finite-dimensional case). *Let $\mathscr{A}$ be a commutative $*$-algebra of operators on a finite-dimensional Hilbert space, and let $\mathbb{P}$ be a state*

[2] In fact, it is the smallest $*$-algebra of operators that contains A.

on $\mathscr{A}$. Then there are a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a map ι from $\mathscr{A}$ onto the set of measurable functions on Ω that is a $*$-isomorphism, i.e., a linear bijection with $\iota(AB) = \iota(A)\iota(B)$ (pointwise) and $\iota(A^*) = \iota(A)^*$, and moreover $\mathbb{P}(A) = E_{\mathbf{P}}(\iota(A))$.

Proof. The proof is an elementary exercise in linear algebra. As the Hilbert space H has dimension $n < \infty$, we can, without loss of generality, suppose that $\mathsf{H} = \mathbf{C}^n$ and that $\mathscr{A}$ is a commutative $*$-algebra of complex $n \times n$ matrices. As all the elements of $\mathscr{A}$ commute, we can find a unitary matrix U such that U^*AU is a diagonal matrix for every $A \in \mathscr{A}$. Let $\Omega = \{1, \dots, n\}$. Define $\iota(A) : \Omega \to \mathbf{C}$ by $\iota(A)(i) = (U^*AU)_{ii}$ for every $A \in \mathscr{A}$. Next, define $\mathcal{F} = \sigma\{\iota(A) : A \in \mathscr{A}\}$. Finally, define $\mathbf{P}(S) = \mathbb{P}(\iota^{-1}(\chi_S))$ for every $S \in \mathcal{F}$. We have now explicitly constructed $(\Omega, \mathcal{F}, \mathbf{P})$ and ι. ∎

Evidently the commutative $*$-algebra structure is completely equivalent to classical probability theory; by simultaneously diagonalizing all the operators in the algebra, we obtain an explicit representation of measurable random variables as the functions on the diagonals. We also note the following. Suppose we are given some (large) commutative $*$-algebra $\mathscr{A}$, and consider a subalgebra $\mathscr{B} \subset \mathscr{A}$ generated by a single element $B \in \mathscr{A}$. If we apply the map ι to $\mathscr{B}$, we obtain precisely the subset of functions on Ω that are measurable with respect to $\sigma\{\iota(B)\}$. Thus subalgebras play the same role in quantum probability as sub-σ-algebras in classical probability; they allow us to keep track of particular subsets of information.

We do not a priori have a basis for specifying a particular commutative $*$-algebra; given a quantum system, we could decide to measure any of a large set of incompatible observables. The discussion up to this point motivates the following definition.

DEFINITION 2.5 (quantum probability space, finite-dimensional case). *A pair $(\mathscr{N}, \mathbb{P})$, where $\mathscr{N}$ is a (not necessarily commutative) $*$-algebra of operators on a finite-dimensional Hilbert space and $\mathbb{P}$ is a state on $\mathscr{N}$, is called a (finite-dimensional) quantum probability space.*

Usually we will choose $\mathscr{N}$ to be the set of all (bounded) operators $\mathscr{B}(\mathsf{H})$ on some underlying Hilbert space H. The principles of quantum probability now amount to the following. In each realization, we must make a choice of commutative $*$-subalgebra $\mathscr{A} \subset \mathscr{N}$ which fixes the observations. Every statistic that pertains to these observations (e.g., the statistics compiled by repeating the experiment many times with the same choice of $\mathscr{A}$) is now described by the classical probability model obtained through the spectral theorem. The reader should convince himself that the operational description given in the previous section fits neatly within this model (with the exception of conditioning, which we discuss in section 2.4).

Notice that in contrast to a classical probability space $(\Omega, \mathcal{F}, \mathbf{P})$, there are no sample points $\omega \in \Omega$ in a quantum probability space. The sample points emerge through the spectral theorem after the choice of a commutative $*$-subalgebra.

Example 2.6. Let us reformulate Example 2.3. Set $\mathsf{H} = \mathbf{C}^2$ and choose $\mathscr{N} = \mathscr{B}(\mathsf{H}) = M_2$, the $*$-algebra of 2×2 complex matrices. The pure state is defined by $\mathbb{P}(A) = \langle \psi, A\psi \rangle = \psi^* A \psi$ (recall that $\psi = (c_1 \; c_{-1})^T$ with $|c_1|^2 + |c_{-1}|^2 = 1$).

The observable σ_z, used to represent spin measurement in the z direction, generates a commutative $*$-subalgebra $\mathscr{A}_z \subset \mathscr{N}$. It is not difficult to see that $\mathscr{A}_z$ is simply the linear span of the events $P_{z,1}$ and $P_{z,-1}$. Let us now apply the spectral theorem; we obtain the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ where $\Omega = \{1, 2\}$, $\mathcal{F} = \{\varnothing, \{1\}, \{2\}, \Omega\}$, $\mathbf{P}(\{1\}) = |c_1|^2$, etc., and $\iota(P_{z,1}) = \chi_{\{1\}}$, $\iota(P_{z,-1}) = \chi_{\{2\}}$. In particular, the random variable $\iota(\sigma_z) : (1, 2) \mapsto (1, -1)$ has precisely the right properties.

Now suppose we do not wish to measure the intrinsic angular momentum (spin)

in the z direction, but in the x direction. This corresponds to the observable

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2.9}$$

which has the spectral decomposition $\sigma_x = P_{x,1} - P_{x,-1}$ with

$$P_{x,1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad P_{x,-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The observable σ_x also generates a commutative $*$-subalgebra $\mathscr{A}_x = \mathrm{span}\{P_{x,1}, P_{x,-1}\}$ to which we can apply the spectral theorem. However, as σ_x and σ_z do not commute, they cannot be jointly represented on a classical probability space through the spectral theorem. In other words, σ_x and σ_z are incompatible and their joint statistics are undefined; hence they cannot both be observed in the same realization. □

To conclude this section, let us say a few words about the interpretation of the Heisenberg uncertainty relation. The relation says that the product of the variances of two noncommuting observables is bounded from below by a positive constant. It is important to realize, however, that the two observables cannot be measured in the same realization as they are incompatible—in particular, the *co*variance of the observables is undefined. Rather, the uncertainty relation is a statement about the properties of quantum states: for any state, the statistics of the two observables, compiled in the course of separate realizations in each of which only one of the observables is measured, must obey the Heisenberg inequality.[3]

2.3. Composite systems. We will often wish to form a composite probability model from two separate probability spaces. In classical probability theory, two probability spaces $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ can be merged into a single probability space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbf{P}_1 \times \mathbf{P}_2)$ where $\mathbf{P}_1 \times \mathbf{P}_2$ is the product measure. We now briefly describe the noncommutative counterpart.

Consider a composite system constructed from two quantum probability spaces $(\mathscr{N}_1, \mathbb{P}_1)$, $(\mathscr{N}_2, \mathbb{P}_2)$ of operators on the Hilbert spaces H_1 and H_2, respectively. The composite quantum probability space consists of operators on the tensor product Hilbert space $\mathsf{H}_1 \otimes \mathsf{H}_2$; for vectors $\psi_1, \phi_1 \in \mathsf{H}_1$ and $\psi_2, \phi_2 \in \mathsf{H}_2$, the inner product on $\mathsf{H}_1 \otimes \mathsf{H}_2$ is given by

$$\langle \psi_1 \otimes \psi_2, \phi_1 \otimes \phi_2 \rangle = \langle \psi_1, \phi_1 \rangle \langle \psi_2, \phi_2 \rangle,$$

which is extended by linearity to any vector in the tensor product space. The algebra $\mathscr{N}_1 \otimes \mathscr{N}_2$ is generated by elements of the form

$$(A_1 \otimes A_2)(\psi_1 \otimes \psi_2) = A_1\psi_1 \otimes A_2\psi_2,$$

where $A_1 \in \mathscr{N}_1$ and $A_2 \in \mathscr{N}_2$. Finally, the product state is defined by

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)(A_1 \otimes A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$$

[3] In the physics literature one often find statements to the effect that the Heisenberg uncertainty relation limits the precision with which we can "imperfectly" observe two noncommuting observables simultaneously, i.e., within the same realization. This is a misconception. Though the idea of an imperfect measurement can be implemented rigorously (see, e.g., [38]), this gives rise to an uncertainty relation which is different from Heisenberg's uncertainty relation [4].

and is extended by linearity. The quantum probability space $(\mathscr{N}_1 \otimes \mathscr{N}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ of operators on the Hilbert space $\mathsf{H}_1 \otimes \mathsf{H}_2$ describes the composite system. The reader should verify that if $\mathscr{N}_1$ and $\mathscr{N}_2$ are commutative, then applying the spectral theorem to the composite system is equivalent to applying the spectral theorem to the individual subsystems, then forming the composite classical probability space.

2.4. Conditional expectations. Let us recall for a moment the Stern–Gerlach experiment of Examples 2.3 and 2.6. We have introduced the observables σ_z and σ_x, corresponding to spin in the z and x directions. These observables are incompatible, so we cannot measure them in the same realization. Recall that in order to measure σ_z, Stern and Gerlach apply a field gradient in the z direction; the atom then acquires momentum in that direction proportional to σ_z, and we can determine the value of σ_z in that realization by observing whether the atom is deflected up (1) or down (-1). Similarly, σ_x is measured by orienting the field gradient along the x axis.

We would not be measuring both σ_z and σ_x by applying both field gradients simultaneously, but rather as magnetic fields add vectorially, this would measure the spin in some other direction in the x-z plane whose observable commutes with neither σ_z nor σ_x. On the other hand, we could first apply the field gradient in the z direction until we can resolve σ_z, then turn this field off and switch on a field in the x direction to resolve σ_x. It is a characteristic feature of quantum mechanics that the measurement outcomes in such a procedure can differ drastically depending on what order we apply the fields. It is thus of crucial importance to specify precisely how such measurements are performed by including in the quantum probability space a model of the measurement apparatus (or *probe*).

We defer the discussion of the Stern–Gerlach measurement with magnetic fields until we have developed the necessary machinery in section 3. For the sake of example, we develop in this section a simpler probe model which shows the main features of the procedure. We will see that this probe model, together with the concept of conditional expectations, reproduces precisely the traditional projection postulate of section 2.1.

Let us begin by discussing *conditional expectations* in the noncommutative context. The key observation we need is the following. The conditional probability of an event B given an event A is the probability that B is true given that A is true in the same realization. Hence the concept of conditioning inherently makes sense only in the context of quantities that can be observed in the same realization of an experiment. This means that we can only define conditional expectations in commutative subalgebras of a quantum probability space; but as long as we are restricted to the commutative case, the spectral theorem allows us to define any probabilistic operation directly in terms of the associated classical probability space (see [18]).

To be more precise, let $(\mathscr{N}, \mathbb{P})$ be a quantum probability space, $\mathscr{A} \subset \mathscr{N}$ be a commutative subalgebra, and $B \in \mathscr{N}$ be a self-adjoint element that commutes with every $A \in \mathscr{A}$. Then B and $\mathscr{A}$ generate a larger commutative subalgebra $\mathscr{C} \subset \mathscr{N}$, to which we can apply the spectral theorem to obtain a $*$-isomorphism ι. The conditional expectation is now simply inherited from the classical space as $\mathbb{P}(B|\mathscr{A}) = \iota^{-1}(E_{\mathbf{P}}(\iota(B)|\sigma\{\iota(\mathscr{A})\}))$. Note, however, that if B, C are two self-adjoint operators that commute with every $A \in \mathscr{A}$, this does not necessarily imply that B and C commute. The set $\mathscr{A}' = \{B \in \mathscr{N} : AB = BA \ \forall A \in \mathscr{A}\}$, the *commutant* of $\mathscr{A}$ (in $\mathscr{N}$), is the largest $*$-subalgebra of operators that can be conditioned on $\mathscr{A}$. The conditional expectation is defined as above for its self-adjoint elements, and extends to all of $\mathscr{A}'$ by linearity.

From this discussion and the definition of the classical conditional expectation, we extract the following definition directly in terms of the quantum probability space.

DEFINITION 2.7 (conditional expectation, finite-dimensional case). *Let $(\mathscr{N}, \mathbb{P})$ be a finite-dimensional quantum probability space and let $\mathscr{A} \subset \mathscr{N}$ be a commutative $*$-subalgebra. Then $\mathbb{P}(\cdot|\mathscr{A}) : \mathscr{A}' \to \mathscr{A}$ is called (a version of) the conditional expectation from $\mathscr{A}'$ onto $\mathscr{A}$ if $\mathbb{P}(\mathbb{P}(B|\mathscr{A})A) = \mathbb{P}(BA)$ for all $A \in \mathscr{A}$, $B \in \mathscr{A}'$.*

As we will see in section 3, the discussion above generalizes directly to the infinite-dimensional case. In finite dimensions it is convenient to give an explicit expression for the conditional expectation. Note that a finite-dimensional $*$-algebra is a finite-dimensional linear space. Then $\langle A, B\rangle_{\mathbb{P}} = \mathbb{P}(A^*B)$ turns the algebra into a pre-Hilbert space; i.e., it is a Hilbert space except that $A \mapsto \langle A, A\rangle_{\mathbb{P}} = \|A\|_{\mathbb{P}}^2$ may have a nontrivial null space. In particular, the fundamental property $\mathbb{P}(\mathbb{P}(B|\mathscr{A})A) = \mathbb{P}(BA)$ for all $A \in \mathscr{A}$ is precisely that of orthogonal projection from $\mathscr{A}'$ onto the linear subspace $\mathscr{A}$, which in a pre-Hilbert space is uniquely determined up to an event of zero probability. Note that the classical characterization of $\mathbb{P}(B|\mathscr{A})$ as the least mean square estimate of B in $\mathscr{A}$ follows immediately. We will elaborate on this point in section 3.

An explicit expression for $\mathbb{P}(B|\mathscr{A})$ is easily obtained if we find an orthogonal basis for $\mathscr{A}$. Any commutative $*$-algebra in finite dimensions is spanned by a set of projections that resolve the identity. This is easily seen: in n dimensions any self-adjoint operator is a linear combination of at most n projections that resolve the identity, and as all the operators in the $*$-algebra commute they must be expressible as linear combinations of the same projections. Let $\mathscr{A} = \operatorname{span}\{P_a\}$ for some set of orthogonal projections P_a resolving the identity. Then a version of the conditional expectation is given by

$$\mathbb{P}(B|\mathscr{A}) = \sum_{P \in \{P_a\} : \mathbb{P}(P) \neq 0} \frac{P}{\|P\|_{\mathbb{P}}} \left\langle \frac{P}{\|P\|_{\mathbb{P}}}, B \right\rangle_{\mathbb{P}} = \sum_{P \in \{P_a\} : \mathbb{P}(P) \neq 0} \frac{\mathbb{P}(PB)}{\mathbb{P}(P)} P. \tag{2.10}$$

Note what could happen if we naively fill in some $B \notin \mathscr{A}'$. Then $\langle P, B\rangle_{\mathbb{P}} \neq \langle B, P\rangle_{\mathbb{P}}$ for some $P \in \{P_a\}$, which implies that we obtain complex coefficients in the sum even if B is an observable. Hence the expression does not make sense unless $B \in \mathscr{A}'$.

Example 2.8. This example serves to illustrate conditional expectations; it is not meant to represent a particular physical scenario. Consider $\mathsf{H} = \mathbf{C}^3$, $\mathscr{N} = M_3$, and $\mathbb{P}(X) = \langle \psi, X\psi\rangle$ with $\psi = (1\ 1\ 1)^T/\sqrt{3}$. Define $A, B \in \mathscr{N}$ by

$$A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix} = 4 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 5 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Let $\mathscr{A}$ be the $*$-algebra generated by A. Then

$$\mathscr{A}' = \left\{ \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & x \end{pmatrix} \ : \ a, b, c, d, x \in \mathbf{C} \right\}.$$

Note that $\mathscr{A}'$ is not a commutative algebra, despite that every element of $\mathscr{A}'$ commutes with every element of $\mathscr{A}$. As $B \in \mathscr{A}'$, we can use (2.10) to calculate

$$\mathbb{P}(B|\mathscr{A}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} = 1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathscr{A}.$$

The observable $\mathbb{P}(B|\mathscr{A})$ is the orthogonal projection of B onto $\mathscr{A}$ with respect to the inner product $\langle A, B\rangle_{\mathbb{P}} = \mathbb{P}(A^*B)$. By the projection theorem, $\mathbb{P}(B|\mathscr{A})$ is an element of $\mathscr{A}$ that minimizes the mean square error $\|B - \mathbb{P}(B|\mathscr{A})\|_{\mathbb{P}}$. ∎

We now proceed to develop a simple probe model that reproduces the projection postulate. Recall that the conditional *probability* of an event P given a commuting event Q is simply given by $\mathbb{P}(PQ)/\mathbb{P}(Q)$. This is equivalent to $\mathbf{P}(A \cap B)/\mathbf{P}(B)$ by the spectral theorem, where A and B are the sets corresponding to P and Q.

Example 2.9. *Simple probe model.* We will work in a generic n-dimensional setting, $n < \infty$. Let $\mathsf{H} = \mathbf{C}^n$, $\mathscr{N} = M_n$ (the set of $n \times n$ complex matrices), and let $\mathbb{P}(X) = \mathrm{Tr}[\rho X]$ be some state on $\mathscr{N}$. Let A, B be two observables in $\mathscr{N}$ that do not commute. Hence we cannot measure A and B directly in the same realization. However, we can have the system interact with an external probe system, in such a way that the observable A is copied to some probe observable A' after the interaction. If A' commutes with B, we interpret this procedure (like in the Stern–Gerlach example) as an (indirect) measurement of A followed by a (direct) measurement of B.

The strategy is simple. First, we describe the probe system by a separate probe quantum probability space $(\mathscr{N}_p, \mathbb{P}_p)$ and form the composite space $(\mathscr{N} \otimes \mathscr{N}_p, \mathbb{P} \otimes \mathbb{P}_p)$. Next, we introduce an interaction. Recall from section 2.1 that the evolution of an isolated system is described by a unitary transformation. Hence, we will choose a probe observable $I \otimes A'$ and construct a suitable unitary operator U so that the probe observable $U^*(I \otimes A')U$ after the interaction gives the same outcome as $A \otimes I$ would have before the interaction. Note that, by construction, the system observable $B \otimes I$ commutes with $I \otimes A'$ after the interaction, $[U^*(I \otimes A')U, U^*(B \otimes I)U] = 0$. Hence we can measure them within the same realization.

We now fill out the details of this model. Let $A = \sum_{a \in \mathrm{spec}(A)} a\, P_a$, and we denote by m the number of elements in $\mathrm{spec}(A)$ (the number of possible measurement outcomes). For the probe algebra, we choose $\mathsf{H}_p = \mathbf{C}^m$, $\mathscr{N}_p = M_m$. Now fix an observable $A' \in \mathscr{N}_p$ that has the same spectrum as A. $A' = \sum_{a \in \mathrm{spec}(A)} a\, P'_a$ and that P'_a are projections onto one-dimensional subspaces of H_p; hence we can fix an orthonormal basis of vectors $\psi_a \in \mathsf{H}_p$ such that $P'_a = \psi_a \psi_a^*$. Now define the operator $X'_{ab} = \psi_b \psi_a^* + \psi_a \psi_b^* + \sum_{c \neq a,b} \psi_c \psi_c^* \in \mathscr{N}_p$ for $a \neq b$, and $X'_{aa} = I$; these operators switch the events P'_a and P'_b in the sense $X'_{ab} P'_a X'_{ab} = P'_b$, $X'_{ab} P'_b X'_{ab} = P'_a$, and $X'_{ab} P'_c X'_{ab} = P'_c$ for $c \neq a, b$. Finally, set $\mathbb{P}_p(X) = \mathrm{Tr}[X P'_p]$ where we have fixed some $p \in \mathrm{spec}(A)$ at the outset.

Now consider the operator $U \in \mathscr{N} \otimes \mathscr{N}_p$ defined by $U = \sum_{a \in \mathrm{spec}(A)} P_a \otimes X'_{ap}$. As $(X'_{ap})^2 = I$ it follows that $U^*U = UU^* = U^2 = I$; i.e., U is unitary. Note that $U^*(I \otimes P'_c)U = P_c \otimes P'_p + (1 - P_c) \otimes P'_c$ if $c \neq p$, $U^*(I \otimes P'_p)U = \sum_a P_a \otimes P'_a$. We calculate $(\mathbb{P} \otimes \mathbb{P}_p)(U^*(I \otimes P'_c)U(P_c \otimes I))/(\mathbb{P} \otimes \mathbb{P}_p)(P_c \otimes I) = 1$ for every c; i.e., the conditional probability that $U^*(I \otimes A')U$ gives the outcome c, given that we have observed $A \otimes I$ with outcome c, is one. Thus the unitary interaction U precisely copies the system observable A onto the probe observable A'.

We can now measure the system observable B after interaction with the probe. In particular, let us calculate the expectation of B conditioned on the probe measurement. Define $\mathscr{A}$ as the commutative $*$-algebra generated by $U^*(I \otimes A')U$, and note that $U^*(B \otimes I)U \in \mathscr{A}'$. Thus we can use (2.10) to calculate

$$(\mathbb{P} \otimes \mathbb{P}_p)(U^*(B \otimes I)U|\mathscr{A}) = \sum_c \frac{(\mathbb{P} \otimes \mathbb{P}_p)(U^*(B \otimes P'_c)U)}{(\mathbb{P} \otimes \mathbb{P}_p)(U^*(I \otimes P'_c)U)}\, U^*(I \otimes P'_c)U$$

$$= \sum_c \frac{\mathbb{P}(P_c B P_c)}{\mathbb{P}(P_c)}\, U^*(I \otimes P_c')U = \sum_c \mathrm{Tr}[\rho_c B]\, U^*(I \otimes P_c')U,$$

where $\rho_c = P_c \rho P_c / \mathrm{Tr}[\rho P_c]$. This is precisely the projection postulate of section 2.1.

This example may be somewhat bewildering, and we encourage the reader to work through the procedure for a particular model (e.g., that of Example 2.8), paying particular attention to which operators do and do not commute. The reader should convince himself that different answers are obtained if one first measures B then A.

Finally, we note that though we have here measured A through a probe and B directly, there is no reason to stop here. If, in addition to A and B, we want to measure an observable C that does not commute with B, we would introduce a second probe to measure B as well. Now suppose that $C = A$. If we first measure A through the probe, then measure A again, we would (obviously) obtain the same outcome. However, if we first probe A, then probe B, and then measure A, we obtain a different outcome from that of the first measurement of A! The reader is encouraged to work out also this case. The reason for this phenomenon is that the interaction with the probe that is used for the observation of B disturbs the system in such a way that its value of A is changed. This effect is known as "measurement back action." □

The previous example, in particular the construction of the probe and the corresponding interaction, may seem rather ad hoc, and indeed we have only chosen this rather artificial example to reproduce the projection postulate. This is not a shortcoming of the theory we have outlined, however, but rather highlights the importance of including a reasonable model of the probe in the quantum probability space. Indeed, most realistic measurement setups are not of this type and the projection postulate of section 2.1 cannot be used to describe such systems. For example, we will see in section 3 that the Stern–Gerlach measurement is only approximately described by the projection postulate. Later we will describe even more complicated optical measurements in which we wish to condition system observables based on the observation of stochastic processes in continuous time (the signal from a photodetector). It is the latter, most practically useful case where we need quantum filtering theory.

Remark 2.10. It is important to realize that statements like the projection postulate do not really implement the notion of conditioning; they consist of a pure conditioning component and of a particular physical probe model which has no statistical significance. One also finds in the literature generalizations of the projection postulate, called instruments, which implement different types of probes [25, 40]. In the quantum probability context of this paper it is most natural to separate the two parts; we will take existing probe models from physics and concentrate on the calculation of the associated conditional expectations (filtering). □

3. Noncommutative probability theory. In the finite-dimensional case, we have seen in section 2 that quantum mechanics can be modeled as a noncommutative probability theory. In this section we present a general formulation for quantum probability that has wide applicability. We give a general definition of quantum probability space, prove the existence and uniqueness of conditional expectations, and prove a quantum version of Bayes' rule that is very helpful for quantum filtering.

Almost all of the features of the full theory can already be seen in the finite-dimensional case discussed in section 2; the main difficulties in the general case are the technicalities involved in the theory of infinite-dimensional Hilbert spaces. This parallels the difficulties in classical probability theory—though finite-state random

variables can be treated by almost trivial (counting, combinatoric) methods, the description of continuous random variables requires us to upgrade our machinery using methods of real analysis. Similarly, the elementary linear algebra that underlies finite-dimensional quantum probability must be upgraded to functional analysis if we wish to treat the infinite-dimensional case. Conceptually, however, the two cases are very similar, and the reader is encouraged to develop an intuitive understanding of the finite-dimensional case before tackling the full formalism. For a thorough introduction to functional analysis we refer the reader to the excellent textbook [56].

3.1. Quantum probability spaces. Let H be a complex Hilbert space, and denote by $\mathscr{B}(\mathsf{H})$ the set of all bounded (linear) operators on H. We restrict ourselves (for the time being) to bounded operators as we wish to construct $*$-algebras of such operators: attempting to do this with unbounded operators would get us into no end of trouble, as we would surely run into domain problems. Recall that for $A \in \mathscr{B}(\mathsf{H})$, the usual Hilbert space adjoint $A^* \in \mathscr{B}(\mathsf{H})$ is defined by $\langle \psi, A\phi\rangle = \langle A^*\psi, \phi\rangle \ \forall \psi, \phi \in \mathsf{H}$. With this involution $\mathscr{B}(\mathsf{H})$ is a $*$-algebra in the sense of section 2.

We wish to introduce a structure that plays the same role as a $*$-algebra in the finite-dimensional case. It turns out, however, that the $*$-algebra structure in itself is not sufficient in the infinite-dimensional case; we need to impose an additional technical condition in order to be able to prove an infinite-dimensional version of the spectral theorem (Theorem 2.4). The additional condition has a natural interpretation which we will discuss below; however, the reader should not be too worried about this technicality, particularly if he is not familiar with nets or locally convex topologies. In practice we will rarely need to verify this property directly.

DEFINITION 3.1. *A positive linear functional $\mu : \mathscr{B}(\mathsf{H}) \to \mathbf{C}$ is said to be* normal *if $\mu(\sup_\alpha A_\alpha) = \sup_\alpha \mu(A_\alpha)$ for any upper bounded increasing net $\{A_\alpha\}$ of positive elements in $\mathscr{B}(\mathsf{H})$. The locally convex topology on $\mathscr{B}(\mathsf{H})$ defined by the family of seminorms $\{A \mapsto |\mu(A)| : \mu \text{ normal}\}$ is called the* normal topology.

For a detailed discussion of nets, locally convex topologies, etc., see [56].

DEFINITION 3.2 (von Neumann algebra). *A von Neumann algebra $\mathscr{N}$ is a $*$-subalgebra of $\mathscr{B}(\mathsf{H})$ that is closed in the normal topology. A state $\mathbb{P}$ on $\mathscr{N}$ is normal if it is the restriction to $\mathscr{N}$ of a normal state on $\mathscr{B}(\mathsf{H})$.*

We can now extend the spectral theorem to the infinite-dimensional case, essentially showing that commutative von Neumann algebras with normal states are equivalent to classical probability spaces. See, e.g., [57, Proposition 1.18.1] for a proof. Conceptually, we are guided by the finite-dimensional case; Theorem 3.3 extends the idea of simultaneous diagonalization to infinite-dimensional operators. Though technically much more involved, the flavor of the procedure remains the same.[4]

THEOREM 3.3 (spectral theorem). *Let $\mathscr{C}$ be a commutative von Neumann algebra. Then there is a measure space $(\Omega, \mathcal{F}, \mu)$ and a $*$-isomorphism ι from $\mathscr{C}$ to $L^\infty(\Omega, \mathcal{F}, \mu)$, the algebra of bounded measurable complex functions on Ω up to μ-a.s. equivalence. Moreover, a normal state $\mathbb{P}$ on $\mathscr{C}$ defines a probability measure $\mathbf{P}$, which is absolutely continuous with respect to μ such that $\mathbb{P}(C) = E_{\mathbf{P}}(\iota(C))$ for all $C \in \mathscr{C}$.*

Before we continue, let us demonstrate the significance of the additional technical conditions on a von Neumann algebra. First, we give an example of a $*$-subalgebra

[4]The additional measure μ that shows up in the theorem has no direct physical significance; its job is to identify "enough" null sets in $L^\infty(\Omega)$ so we can construct the $*$-isomorphism ι. We can generally not use $\mathbf{P}$ for this purpose as there may be projections $P \in \mathscr{C}$ with $\mathbb{P}(P) = 0$; if ι were to map to $L^\infty(\Omega, \mathcal{F}, \mathbf{P})$, then necessarily $\iota(P) = 0$ and hence ι would not be invertible. The precise details of the construction are never an issue, as we will never use μ and only prove results $\mathbf{P}$-a.s.

of $\mathscr{B}(\mathsf{H})$ that is not a von Neumann algebra.

Example 3.4. Let $\mathsf{H} = L^2([0,1])$ and $\mathscr{A} = C([0,1])$, the commutative algebra of continuous functions on the unit interval. We can consider $A \in \mathscr{A}$ as an operator on H under pointwise multiplication; i.e., $(A\psi)(x) = A(x)\psi(x)$ for every $\psi \in \mathsf{H}$. Then $\mathscr{A}$ satisfies all the requirements of a von Neumann algebra except that it is not closed in the normal topology. Indeed, one can construct, for example, an increasing sequence of continuous functions that converges to $\chi_{[0,1/2]}$, which is discontinuous.

The problem is that the only indicator functions in $\mathscr{A}$ are $\chi_\varnothing$ and $\chi_{[0,1]}$: all other indicator functions on $[0,1]$ are discontinuous. Hence from a probabilistic point of view $\mathscr{A}$ defines a trivial theory, as the only events in $\mathscr{A}$ are the trivial ones. Nonetheless $\mathscr{A}$ is much larger than the algebra $\mathbf{C}$ that is generated by $\chi_\varnothing$ and $\chi_{[0,1]}$. Hence $\mathscr{A}$ cannot be $*$-isomorphic to the set of measurable functions on some measure space. The role of normal closure is to avoid this complication. Indeed, this property guarantees that any von Neumann algebra is generated by its projections [45]. □

Like normal closure, normality of the state is also required in order for the spectral theorem to hold. Note that for normal states the expectation of an increasing set of observables converges to the expectation of their least upper bound; i.e., the monotone convergence property holds. This corresponds to the more basic property of countable additivity. In the following example we construct a state which is not normal.

Example 3.5. Let $\mathsf{H} = \ell^2(\mathbf{N})$ and $\mathscr{A} = \ell^\infty(\mathbf{N})$, acting on H by pointwise multiplication. $\mathscr{A}$ is closed in the normal topology; i.e., it is a commutative von Neumann algebra. Now introduce a state on $\mathscr{A}$ which is given by the expression[5]

$$\mathbb{P}(A) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} A(n), \qquad A \in \mathscr{D} \subset \mathscr{A}, \tag{3.1}$$

on a suitably chosen linear subspace $\mathscr{D}$. $\mathbb{P}$ is not a normal state; to see this, let us introduce the events $P_n \in \mathscr{A}$ defined by $(P_n\psi)(k) = \psi(k)$ if $k \le n$, and zero otherwise. $\{P_n\}$ is an increasing sequence of projections in $\mathscr{A}$ whose least upper bound is the identity $P_\infty = I$. However, straightforward calculation shows that $\mathbb{P}(P_n) = 0$ for any finite n, whereas $\mathbb{P}(I) = 1$. We conclude that the state $\mathbb{P}$ is not normal.

Note that what we have constructed is precisely the classical model of a uniform distribution over the natural numbers $\mathbf{N}$. This does not give rise to a well-defined probability model in the sense of Kolmogorov, however, as the uniform distribution on $\mathbf{N}$ does not obey the property that the probability of a countable union of disjoint events is the sum of the probabilities of these events (which is exactly what went wrong above). Requiring that the state be normal is equivalent to requiring that it gives rise to a countably additive measure [46], which rules out our example. □

Remark 3.6. Definition 3.2 is one of many equivalent definitions of a von Neumann algebra. We have emphasized normality as it is close to the probabilistic notion of monotone convergence. Normal closure turns out to be equivalent to closure in several other topologies, notably the weak and strong operator topologies on $\mathscr{B}(\mathsf{H})$. We will not concern ourselves with topological issues in this article; see, e.g., [20, section 2.4].

The following definition should come as no surprise.

DEFINITION 3.7 (quantum probability space). *A quantum probability space is a pair $(\mathscr{N}, \mathbb{P})$, where $\mathscr{N}$ is a von Neumann algebra and $\mathbb{P}$ is a normal state.*

[5] Equation (3.1) does not by itself define a state, as there are many $A \in \mathscr{A}$ for which the limit does not exist. However, note that (3.1) is well defined on a linear subspace, e.g., $\mathscr{D} = \{A \in \mathscr{A} : \exists c \in \mathbf{C} \text{ s.t. } \lim_{n\to\infty} A(n) = c\}$. Now $\mathbb{P}$ can be extended from $\mathscr{D}$ to $\mathscr{A}$ using the Hahn–Banach theorem.

The structure has precisely the same interpretation as in section 2, of which we briefly remind the reader. In each realization we must choose a commutative von Neumann subalgebra $\mathscr{A} \subset \mathscr{N}$ which fixes the observations. Every statistic that pertains to these observations is then described by the classical probability model obtained by applying the spectral theorem to $(\mathscr{A}, \mathbb{P})$. The equivalence between commutative quantum probability spaces and classical probability spaces is the foundation of the theory; a commutative quantum probability model *is* a classical probabilistic model, and we will often implicitly identify these two pictures.

In this article we will only use three types of von Neumann algebras. We list these below; they will be used throughout without comment.

(i) $\mathscr{A} = \mathscr{B}(\mathsf{H})$ is a von Neumann algebra. Moreover, any *vector state* on $\mathscr{A}$ ($\mathbb{P}(A) = \langle \psi, A\psi \rangle$ for fixed $\psi \in \mathsf{H}$), or any convex combination of vector states, is a normal state. Many models from quantum mechanics are described by such a model.

(ii) $\mathscr{A} = L^\infty(\Omega, \mathcal{F}, \mathbf{P})$, acting on $\mathsf{H} = L^2(\Omega, \mathcal{F}, \mathbf{P})$ by pointwise multiplication, is a commutative von Neumann algebra. Moreover, any state of the form $\mathbb{P}(X) = E_{\mathbf{P}}(X)$ is a normal state. This is a *classical probability model*.

(iii) Given $\mathscr{S} \subset \mathscr{B}(\mathsf{H})$, recall that $\mathscr{S}' = \{X \in \mathscr{B}(\mathsf{H}) : XS = SX \ \forall S \in \mathscr{S}\}$ is called the *commutant* of $\mathscr{S}$ in $\mathscr{B}(\mathsf{H})$. The following theorem (see [45, Theorem 5.3.1] for a proof) allows us to construct von Neumann subalgebras of $\mathscr{B}(\mathsf{H})$.

THEOREM 3.8 (double commutant theorem). *Let $\mathscr{S} \subset \mathscr{B}(\mathsf{H})$ be any self-adjoint set, i.e., $S \in \mathscr{S} \Rightarrow S^* \in \mathscr{S}$. Then $\mathscr{A} = \mathscr{S}''$ is the smallest von Neumann subalgebra of $\mathscr{B}(\mathsf{H})$ that contains $\mathscr{S}$. In particular, $\mathscr{S}$ is a von Neumann algebra iff $\mathscr{S} = \mathscr{S}''$.*

Given any $\mathscr{S} \subset \mathscr{B}(\mathsf{H})$, we call $\mathrm{vN}(\mathscr{S}) = (\mathscr{S} \cup \mathscr{S}^*)''$ the von Neumann algebra generated by $\mathscr{S}$. We will repeatedly use this construction in the following. For example, suppose that we decide to measure in one realization some commuting set of observables $A_1, \ldots, A_n$. Then $\mathscr{A} = \mathrm{vN}(A_1, \ldots, A_n)$ is a commutative von Neumann algebra which, through the spectral theorem, describes the associated classical probability model. $\mathscr{A}$ is the quantum probability equivalent of the σ-algebra generated by a set of random variables.

3.2. Random variables. Now that we have a general definition of a quantum probability space, we can develop some tools to deal with random variables. Recall from section 2 that any self-adjoint element of a quantum probability space can be decomposed into events using (2.4), which gives its interpretation as an observable (random variable). Let us show how to do this in the infinite-dimensional case.

Let $(\mathscr{N}, \mathbb{P})$ be a quantum probability space and consider an element $A \in \mathscr{N}$ which is self-adjoint $A = A^*$. Then $\mathscr{A} = \mathrm{vN}(A) \subset \mathscr{N}$ is a commutative von Neumann algebra. By the spectral theorem, there is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a $*$-isomorphism ι that maps A to some (measurable) random variable $a : \Omega \to \mathbf{R}$. We can now apply classical probability theory; in particular, for any Borel set $B \in \mathcal{B}$ we have the event $[a \in B] = \{\omega \in \Omega : a(\omega) \in B\} = a^{-1}(B) \in \mathcal{F}$. To map this event back to $\mathscr{A}$ we simply invert ι; the projection corresponding to $[a \in B]$ is denoted by $P_A(B) = \iota^{-1}(\chi_{[a \in B]})$, and we call the map P_A from $\mathcal{B}$ to the projections in $\mathscr{N}$ the *spectral measure* of A. But this object is a familiar one from functional analysis [56]; in fact, it is well known that we can express A in terms of its spectral measure by

$$A = \int_{\mathbf{R}} \lambda \, P_A(d\lambda), \tag{3.2}$$

where the integral is defined in a suitable sense [56, sections 7.3 and 8.3]. Equation (3.2) is precisely the infinite-dimensional counterpart of (2.4). We emphasize the

physical interpretation of $P_A(B)$: it is the event [A takes a value in B], which occurs with probability $\mathbb{P}(P_A(B))$.

This would be all there is to it, were it not for the fact that our algebras contain only bounded operators (recall that unbounded operators cannot be defined on the entire Hilbert space, and hence cannot be added or multiplied at will). Evidently we did not lose much by this choice, as the probabilistic model is already contained in an algebra of bounded operators by the spectral theorem. An unfortunate side effect, however, is that self-adjoint operators in the algebra can only represent *bounded* random variables, whereas many observations of interest are quite naturally unbounded (think of a Gaussian random variable). This means that we need to deal with unbounded observables separately. We briefly discuss one way of doing this.

Consider a von Neumann algebra $\mathscr{N} \subset \mathscr{B}(\mathsf{H})$. In general, an observable is defined by a (not necessarily bounded) self-adjoint operator A on some dense domain in H. We need to relate the unbounded operator A to $\mathscr{N}$. The trick we use is remarkably simple: we compute a bounded function of A. Define $T_A = (A+iI)^{-1}$. By elementary spectral theory [56], any self-adjoint A has a real spectrum, and hence $A+iI$ is invertible with bounded inverse. We say that A is *affiliated* to $\mathscr{N}$ if $T_A \in \mathscr{N}$. This is the equivalent of the classical notion of a random variable that is measurable with respect to some σ-algebra $\mathcal{G}$. Note that every self-adjoint A is affiliated to $\mathscr{B}(\mathsf{H})$, and if A is also bounded, then A is affiliated to $\mathscr{N}$ iff $A \in \mathscr{N}$.

We wish to represent A as a classical (unbounded) random variable. To this end, define the von Neumann algebra generated by A as $\mathrm{vN}(A) = \mathrm{vN}(T_A)$. Now note that T_A commutes with its adjoint, and hence $\mathrm{vN}(A)$ is a commutative von Neumann algebra to which we can apply the spectral theorem. All we need to do is to "package" A into T_A, apply ι, and "unpack" it on the other end; in other words, we define $\iota(A) = \iota(T_A)^{-1} - i$. Once we have done this, we can define a spectral measure P_A for A in the usual way, and indeed (3.2) still holds even for unbounded A [56]. We remark that A being affiliated to $\mathscr{N}$ corresponds to the fact that $P_A(B) \in \mathscr{N}$ for every $B \in \mathcal{B}$; this is precisely the classical notion of measurability.

Unbounded operators are a nuisance, but unfortunately they are a fact of life in mathematical physics. In this article, particularly in the later sections, we will occasionally add and multiply unbounded operators without justification; a detailed analysis of the operator domains is beyond our scope. Though this does not often cause trouble, the reader should keep in mind that a fully rigorous treatment must verify that any addition or multiplication of unbounded operators is indeed well defined. We quote one useful result: operators affiliated to a *commutative* von Neumann algebra can be added and multiplied at will (see [45, Theorem 5.6.15] and [54]).

Example 3.9. We take $\mathsf{H} = L^2(\mathbf{R})$ and $\mathscr{N} = \mathscr{B}(\mathsf{H})$. The vector

$$\psi \in \mathsf{H}, \qquad \psi(x) = (2\pi)^{-1/4}\sigma^{-1/2}\exp\left(-\frac{(x-\mu)^2}{4\sigma^2}\right)$$

defines the (pure) state $\mathbb{P}(X) = \langle \psi, X\psi \rangle$. Now consider the self-adjoint operators

$$(Q\psi)(x) = x\psi(x), \qquad (P\psi)(x) = -i\hbar\frac{d}{dx}\psi(x),$$

which are prototypical observables for the position Q and momentum P of a quantum particle. Both are unbounded observables, but their domains include at least the set of smooth functions with compact support which is dense in $L^2(\mathbf{R})$.

What random variables do these represent? We can read off from the definition that Q is a Gaussian random variable with mean μ and variance σ^2—as Q is already in "diagonal" form (Q is affiliated to $L^\infty(\mathbf{R}) \subset \mathscr{N}$), its spectral measure is given by

$$(P_Q(B)\psi)(x) = \chi_B(x)\psi(x)$$

and it is evident that $\mathbb{P}(P_Q(B))$ is a Gaussian measure with mean μ and variance σ^2. Alternatively, consider the characteristic function $q(k) = \mathbb{P}(e^{ikQ})$ of Q. Unlike Q, e^{ikQ} is a bounded operator, and we can directly compute

$$q(k) = \langle \psi, e^{ikQ}\psi\rangle = (2\pi)^{-1/2}\sigma^{-1}\int_{-\infty}^{\infty} e^{ikx}\, e^{-(x-\mu)^2/2\sigma^2}\, dx = e^{ik\mu - k^2\sigma^2/2}$$

which is the characteristic function of a Gaussian random variable with mean μ and variance σ^2. Similarly, e^{ikP} is a bounded operator, and we compute

$$p(k) = \mathbb{P}(e^{ikP}) = \langle \psi, e^{ikP}\psi\rangle = \int_{-\infty}^{\infty} \psi(x)\,\psi(x+\hbar k)\, dx = e^{-\hbar^2 k^2/8\sigma^2}$$

which is the characteristic function of a Gaussian random variable with mean zero and variance $\hbar^2/4\sigma^2$. Thus both Q and P are Gaussian random variables, but their joint distribution is undefined as they do not commute. Note that we cannot choose σ so that both Q and P have arbitrarily small variance: this is a manifestation of the Heisenberg uncertainty relation (compare (2.3)). ∎

The following example plays a central role in the physics of harmonic oscillators; we will encounter a very similar construction later for continuous-time quantum stochastic processes. We will need the following classic result (see, e.g., [45] for a proof).

THEOREM 3.10 (Stone's theorem). *Let $\mathscr{N}$ be a von Neumann algebra and let $\{U_t\}_{t\in\mathbf{R}} \subset \mathscr{N}$ be a group of unitary operators that is strongly continuous. Then there is a unique self-adjoint A affiliated to $\mathscr{N}$, the* Stone generator*, such that $U_t = e^{itA}$.*

Example 3.11. Let $\mathsf{H} = \ell^2(\mathbf{N})$ and $\mathscr{N} = \mathscr{B}(\mathsf{H})$. Define the complete orthonormal basis $\{\psi_n,\ n = 0, 1, \ldots\} \subset \mathsf{H}$, where $\psi_n(k) = 1$ if $k = n$ and $\psi_n(k) = 0$ otherwise. Moreover, we define for every $\alpha \in \mathbf{C}$ the *exponential vector* $e(\alpha) \in \mathsf{H}$ by $e(\alpha)(k) = \alpha^k/\sqrt{k!}$, and we remark that the linear span D of all exponential vectors is dense in H. The normalized exponential vectors $e(\alpha)e^{-|\alpha|^2/2}$ are called coherent vectors and can be used to define the coherent states $\mathbb{P}_\alpha(X) = \langle e(\alpha), Xe(\alpha)\rangle\, e^{-|\alpha|^2}$.

The simplest random variable we can investigate is defined by $(\lambda\psi)(k) = k\psi(k)$—i.e., this is the natural diagonal operator affiliated to $\ell^\infty(\mathbf{N}) \subset \mathcal{N}$. The spectral measure of λ is given by $(P_\lambda(B)\psi)(k) = \chi_B(k)\psi(k)$, from which we obtain directly

$$\mathbb{P}_\alpha(P_\lambda(B)) = \langle e(\alpha), P_\lambda(B)e(\alpha)\rangle\, e^{-|\alpha|^2} = \sum_{k\in B} \frac{e^{-|\alpha|^2}(|\alpha|^2)^k}{k!}.$$

Thus, evidently, λ is a Poisson-distributed random variable with intensity $|\alpha|^2$.

Can we find other interesting observables affiliated to $\mathscr{N}$? In many cases, physically relevant observables are found to be the Stone generators of particular unitary symmetry groups; see, e.g., [38] for a lucid discussion. Let us try to implement this procedure with the two-dimensional translation group. As a first attempt, let us define a translation operator by $D_\gamma e(\alpha) = e(\alpha+\gamma)\, e^{|\alpha|^2/2 - |\alpha+\gamma|^2/2}$ for $\gamma \in \mathbf{C}$; the constant

factor ensures that $\|D_\gamma e(\alpha)\| = \|e(\alpha)\|$, as must be the case for any unitary operator. Unfortunately, D_γ is not in fact unitary; a straightforward calculation shows

$$\langle e(\beta), D_\gamma^* D_\gamma e(\alpha)\rangle = \langle D_\gamma e(\beta), D_\gamma e(\alpha)\rangle = e^{\beta^*\alpha} e^{i\,\mathrm{Im}(\beta^*\gamma) - i\,\mathrm{Im}(\alpha^*\gamma)}$$

which contradicts unitarity $D_\gamma^* D_\gamma = I$, i.e., $\langle e(\beta), D_\gamma^* D_\gamma e(\alpha)\rangle = \langle e(\beta), e(\alpha)\rangle = e^{\beta^*\alpha}$. To fix this, define the *Weyl operator*

$$W_\gamma e(\alpha) = e(\alpha+\gamma)\, e^{|\alpha|^2/2 - |\alpha+\gamma|^2/2} e^{i\,\mathrm{Im}(\alpha^*\gamma)} = e(\alpha+\gamma)\, e^{-\gamma^*\alpha - |\gamma|^2/2}.$$

The Weyl operator is unitary and provides a projective unitary representation [38] in the sense that $W_\alpha W_\beta = W_{\alpha+\beta} e^{i\,\mathrm{Im}(\beta^*\alpha)}$. Note that it is sufficient to define the action of W_α only on exponential vectors; we can then extend to D by linearity, and as D is dense and W_α is bounded the Weyl operators are uniquely extended to all of H.

Now fix $\beta \in \mathbf{C}$ and consider the unitary group $\{W_{t\beta}\}_{t\in\mathbf{R}}$. This group is continuous ($W_{t\beta} e(\gamma) \to e(\gamma)$ as $t \to 0$) and hence by Stone's theorem, there exists a self-adjoint operator B_β such that $W_{t\beta} = e^{itB_\beta}$. Finding the distribution of the observable B_β is straightforward, as the characteristic function of B_β is given by

$$b_\beta(k) = \mathbb{P}_\alpha(W_{k\beta}) = \langle e(\alpha), e(\alpha+k\beta)\rangle\, e^{-k\beta^*\alpha - k^2|\beta|^2/2 - |\alpha|^2} = e^{2ik\,\mathrm{Im}(\alpha^*\beta) - k^2|\beta|^2/2}.$$

Hence B_β is a Gaussian random variable with mean $2\,\mathrm{Im}(\alpha^*\beta)$ and variance $|\beta|^2$.

Our next task is to obtain an explicit representation of B_β. We proceed as follows:

$$B_\beta e(\alpha) = \frac{1}{i}\frac{d}{dt} W_{t\beta} e(\alpha)\bigg|_{t=0} = i\beta^*\alpha\, e(\alpha) - i\,\frac{d}{dt} e(\alpha+t\beta)\bigg|_{t=0}.$$

One can verify explicitly that this expression makes sense, i.e., $B_\beta e(\alpha) \in \mathsf{H}$. Note that we cannot extend B_β to all of H, as B_β is unbounded. However, we see that the domain of B_β contains at least the exponential domain D.

Let us introduce the following notation. Define $q = B_i$, $p = B_{-1}$, and $a = (q+ip)/2$. Note that q and p are self-adjoint by Stone's theorem, whereas a has the adjoint $a^* = (q-ip)/2$. Moreover, we find that $a\, e(\alpha) = \alpha\, e(\alpha)$. But then

$$(a\, e(\alpha))(k) = \alpha\,\frac{\alpha^k}{\sqrt{k!}} = \sqrt{k+1}\,\frac{\alpha^{k+1}}{\sqrt{(k+1)!}} = \sqrt{k+1}\, e(\alpha)(k+1).$$

This implies that we can extend the domain of a to include also the $\{\psi_n\}$ by defining $a\,\psi_{k+1} = \sqrt{k+1}\,\psi_k$ (where $a\,\psi_0 = 0$). Furthermore, from

$$\langle \psi_m, a^*\psi_k\rangle = \langle a\psi_m, \psi_k\rangle = \sqrt{m}\,\delta_{(m-1)k} = \sqrt{k+1}\,\delta_{m(k+1)}$$

we can read off $a^*\psi_k = \sqrt{k+1}\,\psi_{k+1}$. a^* is known as the creation (or raising) operator and a as the annihilation (or lowering) operator.

Finally, note that $\lambda = a^*a$. From a classical probability point of view this is very remarkable indeed. Not only do both Poisson and Gaussian random variables emerge from the same state $\mathbb{P}_\alpha$, but there is even a *continuous* map $q, p \mapsto (q-ip)(q+ip)/4 = \lambda$ that transforms two Gaussian random variables into a Poisson random variable. One could never continuously transform a continuous classical random variable into a discrete classical random variable; however, we get away with it here because p, q, and λ do not commute with one another. Thus in each realization we can choose to measure either a discrete or a continuous random variable, but not both. ∎

Remark 3.12. Though presented rather differently, the last two examples are in fact $*$-isomorphic in the case that $\sigma^2 = \frac{1}{2}$ in the first example. For example, if $\alpha \in \mathbf{R}$, we can map $p \mapsto 2^{1/2}\hbar^{-1}P$, $q \mapsto 2^{1/2}Q$, and $\mathbb{P}_\alpha \mapsto \mathbb{P}_{\mu=2^{1/2}\alpha,\sigma=2^{-1/2}}$. From the expression for $b_\beta(k)$ we see that in a coherent state both p and q must have the same variance. In the first example we allowed for the variance of Q to shrink, though this necessarily increases the variance of P. This results in a "squeezed state" which can also be introduced in the context of the second example. We will not construct such states here; in the following, we will only use coherent states. ∎

3.3. Conditional expectation. We now consider conditional expectations, following the treatment of [18]. The following definition is identical to the one in section 2.

DEFINITION 3.13 (conditional expectation). *Let $(\mathscr{N}, \mathbb{P})$ be a quantum probability space and let $\mathscr{A} \subset \mathscr{N}$ be a commutative von Neumann subalgebra. Then the map $\mathbb{P}(\cdot|\mathscr{A}) : \mathscr{A}' \to \mathscr{A}$ is called (a version of) the conditional expectation from $\mathscr{A}'$ onto $\mathscr{A}$ if $\mathbb{P}(\mathbb{P}(B|\mathscr{A})A) = \mathbb{P}(BA)$ for all $A \in \mathscr{A}$, $B \in \mathscr{A}'$.*

We briefly recall the significance of $\mathscr{A}'$. $\mathscr{A}$ is the algebra generated by our observations: it must be commutative, as we cannot observe incompatible events in a single experiment. We now wish to find the conditional statistics of an observable B that is not affiliated to $\mathscr{A}$. However, as we have already observed $\mathscr{A}$, this is only sensible if B commutes with every element in $\mathscr{A}$—there would be no physical way to test our predictions if we could not subsequently measure B in the same realization.

Remark 3.14. Recall that if $B = B^*$, we can use the spectral theorem to obtain explicitly $\mathbb{P}(B|\mathscr{A}) = \iota^{-1}(E_{\mathbf{P}}(\iota(B)|\sigma\{\iota(\mathscr{A})\}))$. This representation extends even to the case that B is an unbounded self-adjoint operator that is affiliated to $\mathscr{A}'$. For simplicity we will discuss below the properties of $\mathbb{P}(B|\mathscr{A})$ assuming that B is bounded, but with suitable care the treatment extends also to the unbounded case. ∎

Remark 3.15. A more general definition (see, e.g., [59]), of which Definition 3.13 is a special case, is often used in quantum probability. Unlike our definition, which is motivated by statistical inference and filtering, the more general "conditional expectation" allows for conditioning on noncommutative algebras and does not have a direct statistical interpretation. The more general definition is used, e.g., in the theory of noncommutative Markov processes [47]. We will not dwell on this further. ∎

THEOREM 3.16. *The conditional expectation of Definition* 3.13 *exists and is unique with probability one* (*any two versions P and Q of $\mathbb{P}(B|\mathscr{A})$ satisfy $\|P-Q\|_{\mathbb{P}} = 0$, where $\|X\|_{\mathbb{P}}^2 = \mathbb{P}(X^*X)$*). *Moreover, $\mathbb{P}(B|\mathscr{A})$ is the least mean square estimate of B given $\mathscr{A}$ in the sense that $\|B - \mathbb{P}(B|\mathscr{A})\|_{\mathbb{P}} \le \|B - A\|_{\mathbb{P}}$ for all $A \in \mathscr{A}$.*

Proof. (i) *Existence.* We have already established that for self-adjoint $B \in \mathscr{A}'$, we can explicitly define a $\mathbb{P}(B|\mathscr{A})$ that satisfies the conditions of Definition 3.13 using the spectral theorem. The classical conditional expectation exists, and moreover the conditional expectation of a bounded random variable is bounded. Hence $\mathbb{P}(B|\mathscr{A})$ exists in $\mathscr{A}$ for self-adjoint $B \in \mathscr{A}'$. But any $B \in \mathscr{A}'$ can be written as $B = B_1 + iB_2$ with self-adjoint $B_1 = (B + B^*)/2$ and $B_2 = i(B^* - B)/2$. As $\mathbb{P}(B_1|\mathscr{A})$ and $\mathbb{P}(B_2|\mathscr{A})$ exist and $\mathbb{P}(B|\mathscr{A}) = \mathbb{P}(B_1|\mathscr{A}) + i\mathbb{P}(B_2|\mathscr{A})$ satisfies the conditions of Definition 3.13, existence is proved.

(ii) *Uniqueness with probability one.* Define the pre-inner product $\langle X, Y\rangle = \mathbb{P}(X^*Y)$ on $\mathscr{A}'$ (it might have nontrivial kernel). Then $\langle A, B - \mathbb{P}(B|\mathscr{A})\rangle = \mathbb{P}(A^*B) - \mathbb{P}(A^*\mathbb{P}(B|\mathscr{A})) = 0$ for all $A \in \mathscr{A}$ and $B \in \mathscr{A}'$, i.e., $B - \mathbb{P}(B|\mathscr{A})$ is orthogonal to $\mathscr{A}$. Now let P and Q be two versions of $\mathbb{P}(B|\mathscr{A})$. It follows that $\langle A, P - Q\rangle = 0$ for all $A \in \mathscr{A}$. But $P - Q \in \mathscr{A}$, so $\langle P - Q, P - Q\rangle = \|P - Q\|_{\mathbb{P}}^2 = 0$.

(iii) *Least squares.* Let P be a version of $\mathbb{P}(B|\mathscr{A})$. Then for all $K \in \mathscr{A}$

$$\|B-K\|_{\mathbb{P}}^2 = \|B-P+P-K\|_{\mathbb{P}}^2 = \|B-P\|_{\mathbb{P}}^2 + \|P-K\|_{\mathbb{P}}^2 \geq \|B-P\|_{\mathbb{P}}^2,$$

where, in the second step, we used that $(B-\mathbb{P}(B|\mathscr{A})) \perp (\mathbb{P}(B|\mathscr{A})-K) \in \mathscr{A}$. ∎

Remark 3.17. The usual elementary properties of classical conditional expectations and their proofs [63] carry over directly. In particular, we have linearity, positivity, invariance of the state $\mathbb{P}(\mathbb{P}(B|\mathscr{A})) = \mathbb{P}(B)$, invariance of $\mathscr{A}$ ($\mathbb{P}(B|\mathscr{A}) = B$ if $B \in \mathscr{A}$), the tower property $\mathbb{P}(\mathbb{P}(B|\mathscr{A})|\mathscr{C}) = \mathbb{P}(B|\mathscr{C})$ if $\mathscr{C} \subset \mathscr{A}$, the module property $\mathbb{P}(AB|\mathscr{C}) = B\,\mathbb{P}(A|\mathscr{C})$ for $B \in \mathscr{C}$, etc. As an example, let us prove linearity. It suffices to show that $Z = \alpha\,\mathbb{P}(A|\mathscr{C}) + \beta\,\mathbb{P}(B|\mathscr{C})$ satisfies $\mathbb{P}(ZC) = \mathbb{P}((\alpha A+\beta B)C)$ for all $C \in \mathscr{C}$. But this is immediate from the linearity of $\mathbb{P}$ and Definition 3.13. ∎

3.4. The Bayes formula. In section 2 we were able to calculate conditional expectations explicitly as all algebras were finite-dimensional. In most physical situations, however, at least the probe (and often the system as well) admits continuous observables and therefore we must deal with infinite-dimensional algebras. In this case it is usually not so simple to calculate the conditional expectations directly; however, the following Bayes-type formula will be of considerable assistance.

LEMMA 3.18 (Bayes formula [18]). *Let $\mathscr{C}$ be a commutative von Neumann subalgebra and let $\mathscr{C}'$ be equipped with a normal state $\mathbb{P}$. Choose $V \in \mathscr{C}'$ such that $V^*V > 0$ and $\mathbb{P}(V^*V) = 1$. Then we can define a new state on $\mathscr{C}'$ by $\mathbb{Q}(A) = \mathbb{P}(V^*AV)$ and*

$$\mathbb{Q}(X|\mathscr{C}) = \frac{\mathbb{P}(V^*XV|\mathscr{C})}{\mathbb{P}(V^*V|\mathscr{C})}, \qquad X \in \mathscr{C}'.$$

Proof. Let K be an element of $\mathscr{C}$. For all $X \in \mathscr{C}'$, we can write

$$\begin{aligned}\mathbb{P}(\mathbb{P}(V^*XV|\mathscr{C})K) &= \mathbb{P}(V^*XKV) = \mathbb{Q}(XK) = \mathbb{Q}(\mathbb{Q}(X|\mathscr{C})K)\\ &= \mathbb{P}(V^*V\mathbb{Q}(X|\mathscr{C})K) = \mathbb{P}(\mathbb{P}(V^*V\mathbb{Q}(X|\mathscr{C})K|\mathscr{C})) = \mathbb{P}(\mathbb{P}(V^*V|\mathscr{C})\mathbb{Q}(X|\mathscr{C})K).\end{aligned}$$

As this holds for all $K \in \mathscr{C}$, and as by construction the conditional expectations are elements of $\mathscr{C}$, we conclude that $\|\mathbb{P}(V^*XV|\mathscr{C}) - \mathbb{P}(V^*V|\mathscr{C})\mathbb{Q}(X|\mathscr{C})\|_{\mathbb{P}} = 0$, or equivalently, $\mathbb{P}(V^*XV|\mathscr{C}) = \mathbb{P}(V^*V|\mathscr{C})\mathbb{Q}(X|\mathscr{C})$ $\mathbb{P}$-a.s. ∎

We now have sufficient tools to deal with the Stern–Gerlach experiment described in section 2. Though the following example is not of much practical importance, it demonstrates the use of the Bayes theorem in a concrete setting. We will use a very similar "reference probability method" to obtain filtering equations later on.

Example 3.19. *Stern–Gerlach experiment.* Consider an atom with two degrees of freedom: a spin degree of freedom $\mathscr{N}_\mu = \mathscr{B}(\mathbf{C}^2)$ carrying the observables σ_x, σ_z etc., and a single spatial degree of freedom $\mathscr{N}_x = \mathscr{B}(\ell^2(\mathbf{N}))$ with the affiliated position q and momentum p observables defined[6] in Example 3.11 (we use the notations of that example). The total algebra describing the atom is then $\mathscr{N} = \mathscr{N}_\mu \otimes \mathscr{N}_x$. Initially the spin and position/momentum of the atom are uncorrelated; hence we work with the state $\mathbb{P} = \mathbb{P}_\mu \otimes \mathbb{P}_0$, where $\mathbb{P}_\mu$ is an arbitrary spin state and $\mathbb{P}_0(X) = \langle\psi_0, X\psi_0\rangle = \langle e(0), Xe(0)\rangle$. The latter implies that initially $I\otimes q$ and $I\otimes p$ (which we will interpret as position and momentum in the z direction) have zero mean and unit variance.

[6] We saw in Remark 3.12 that this description is $*$-isomorphic to the usual definition of position and momentum up to some numerical constants. These are not of essence, however, as they just correspond to a change of units in which we measure position and momentum. A little more care must be taken if we wish to make quantitative predictions on the outcomes of actual experiments; we will not worry about this, however, and work in arbitrary units.

To measure the spin, we apply a magnetic field gradient that is linear in q for some fixed period of time. The resulting force on the particle will cause its momentum to change; an observation of the momentum of the particle after the interaction should thus provide a measurement of its spin σ_z. In other words, the atomic spatial degree of freedom acts as a probe for the atomic spin degree of freedom. The action of the magnetic field is described by the unitary[7]

$$U = \exp(i\kappa\,\sigma_z \otimes q) = P_{z,1} \otimes e^{i\kappa q} + P_{z,-1} \otimes e^{-i\kappa q} = P_{z,1} \otimes W_{i\kappa} + P_{z,-1} \otimes W_{-i\kappa},$$

where $\kappa \in \mathbf{R}$ is the field gradient. Let us thus begin by calculating the characteristic function of $U^*(I \otimes p)U$, the momentum of the atom after the interaction:

$$\begin{aligned}\mathbb{P}(e^{ik\,U^*(I\otimes p)U}) &= \mathbb{P}(U^*(I \otimes W_{-k})U) = \mathbb{P}_\mu(P_{z,1})\,\mathbb{P}_0(W_{-i\kappa}W_{-k}W_{i\kappa}) \\ &\quad + \mathbb{P}_\mu(P_{z,-1})\,\mathbb{P}_0(W_{i\kappa}W_{-k}W_{-i\kappa}) = \mathbb{P}_\mu(P_{z,1})\,e^{2i\kappa k - k^2/2} + \mathbb{P}_\mu(P_{z,-1})\,e^{-2i\kappa k - k^2/2}.\end{aligned}$$

Hence the momentum of the atom after the interaction is distributed as a sum of two Gaussians of unit variance and means 2κ and -2κ, which are weighted, respectively, by $\mathbb{P}_\mu(P_{z,1})$ and $\mathbb{P}_\mu(P_{z,-1})$. Note that we cannot perfectly resolve the spin-up and down states using a Stern–Gerlach measurement; as the tails of the two Gaussians overlap, there is always a nonzero probability that we assign the wrong spin to the atom by looking, e.g., at the sign of the observed momentum. However, the error probability becomes very small when the gradient κ is large.

After the interaction, we may want to measure a spin observable $\sigma \in \mathscr{N}_\mu$ that does not necessarily commute with σ_z (e.g., σ_x). To describe this, let us calculate $\mathbb{P}(U^*(\sigma \otimes I)U|\mathrm{vN}(U^*(I \otimes p)U))$, the conditional expectation of the spin observable σ after the interaction given our observation of the momentum of the atom.

We begin by using the following elementary property: if U is a unitary operator and we define the state $\mathbb{Q}(X) = \mathbb{P}(U^*XU)$, then $\mathbb{P}(U^*XU|U^*\mathscr{C}U) = U^*\mathbb{Q}(X|\mathscr{C})U$ (this can be verified directly using Definition 3.13). Thus we obtain

$$\mathbb{P}(U^*(\sigma \otimes I)U|\mathrm{vN}(U^*(I \otimes p)U)) = U^*\mathbb{Q}(\sigma \otimes I|\mathrm{vN}(I \otimes p))U.$$

We would like to apply the Bayes rule to $\mathbb{Q}(\sigma \otimes I|\mathrm{vN}(I \otimes p))$. As U does not commute with $I \otimes p$, however, the Bayes rule does not apply in this form.

Fortunately, we can circumvent this problem using the following trick. Using the Baker–Campbell–Hausdorff formula, we can rewrite $e^{i\kappa q}$ as

$$e^{i\kappa q} = e^{i\kappa(a+a^*)} = e^{-\kappa^2/2}e^{i\kappa a^*}e^{i\kappa a}.$$

Beware that the Baker–Campbell–Hausdorff formula technically only holds for exponentials of bounded operators; thus here and below there will be domain issues, but these can be resolved with suitable care. As $a\,\psi_0 = 0$, we can write

$$e^{i\kappa q}\psi_0 = e^{-\kappa^2/2}e^{i\kappa a^*}e^{i\kappa a}\psi_0 = e^{-\kappa^2/2}e^{i\kappa a^*}\psi_0 = e^{-\kappa^2/2}e^{i\kappa a^*}e^{-i\kappa a}\psi_0 = e^{-\kappa^2}e^{\kappa p}\psi_0.$$

We obtain

$$\mathbb{P}_0(e^{-i\kappa q}Xe^{i\kappa q}) = \langle e^{i\kappa q}\psi_0, Xe^{i\kappa q}\psi_0\rangle = e^{-2\kappa^2}\langle e^{\kappa p}\psi_0, Xe^{\kappa p}\psi_0\rangle = e^{-2\kappa^2}\mathbb{P}_0(e^{\kappa p}Xe^{\kappa p}).$$

[7]This is the solution of (2.6) at some fixed time t for a suitable interaction Hamiltonian H.

It follows that we can equivalently replace U by V:

$$\mathbb{Q}(X) = \mathbb{P}(U^*XU) = \mathbb{P}(V^*XV), \quad V = e^{-\kappa^2} e^{\kappa\,\sigma_z \otimes p} = e^{-\kappa^2}(P_{z,1} \otimes e^{\kappa p} + P_{z,-1} \otimes e^{-\kappa p}).$$

V is not unitary, but it does commute with $I \otimes p$. Hence the Bayes rule gives

$$\mathbb{P}(U^*(\sigma \otimes I)U|\mathrm{vN}(U^*(I \otimes p)U)) = \frac{U^*\mathbb{P}(V^*(\sigma \otimes I)V|\mathrm{vN}(I \otimes p))U}{U^*\mathbb{P}(V^*V|\mathrm{vN}(I \otimes p))U}.$$

We can now use the module property and independence of $\sigma \otimes I$ and $I \otimes p$ under $\mathbb{P}$ to calculate explicitly the numerator and denominator; elementary manipulations give

$$\begin{aligned}&\mathbb{P}[U^*(\sigma \otimes I)U|\mathrm{vN}(U^*(I \otimes p)U)] \\ &= \frac{\mathbb{P}_\mu(P_{z,1}\sigma P_{z,1})e^{2\kappa\,U^*(I\otimes p)U} + \mathbb{P}_\mu(P_{z,-1}\sigma P_{z,-1})e^{-2\kappa\,U^*(I\otimes p)U} + 2\,\mathrm{Re}\,\mathbb{P}_\mu(P_{z,-1}\sigma P_{z,1})}{\mathbb{P}_\mu(P_{z,1})e^{2\kappa\,U^*(I\otimes p)U} + \mathbb{P}_\mu(P_{z,-1})e^{-2\kappa\,U^*(I\otimes p)U}}.\end{aligned}$$

By definition, $\mathbb{P}(U^*(\sigma \otimes I)U|\mathrm{vN}(U^*(I \otimes p)U))$ is affiliated to $\mathrm{vN}(U^*(I \otimes p)U)$, and indeed the expression above is simply a function of $U^*(I \otimes p)U$. If we observe $U^*(I \otimes p)U$ and obtain the value $\tilde{p}$, then the spectral theorem tells us that the conditional expectation takes the value given by the expression above if we simply substitute $\tilde{p}$ for $U^*(I \otimes p)U$. Note that the formula is not equivalent to the one given by the projection postulate for a measurement of σ_z. For large κ, however, we obtain approximately the projection postulate expression, and this becomes exact as $\kappa \to \infty$. ∎

4. Stochastic processes and quantum Itô calculus. After a general introduction to quantum probability, we now turn to one particular quantum probability space which we will use throughout the remainder of the article. In section 5 we shall argue that this model appropriately describes the quantum electromagnetic field and its interaction with matter. In the laboratory, the electromagnetic field can be measured by devices like photodetectors which can produce an electric current or even a discrete photocount. The statistics of data records from such experiments are well approximated by the model considered here. The model is rich and we will discover that it contains many interesting classical stochastic processes, i.e., a whole family of Poisson and Wiener processes. However, these processes do not commute with each other. An extension of the Itô calculus, due to Hudson and Parthasarathy [42], unites all these processes in one noncommutative stochastic calculus.

4.1. Poisson processes on Fock space. The theory we are about to discuss can be approached from many sides; here we have chosen to get started by finding a quantum probability space that naturally admits a Poisson process, and build the theory from there. As we have a particular classical process in mind, the general theory gives a hint as to how we could proceed. First, we define the process on a classical space $(\Omega, \mathcal{F}, \mathbf{P})$; equivalently, we can form the algebra $\mathscr{A} = L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ acting on $\mathsf{H} = L^2(\Omega, \mathcal{F}, \mathbf{P})$ by pointwise multiplication, with a suitable state $\mathbb{P}$, and represent the process as a family of observables affiliated to $\mathscr{A}$. To create a noncommutative model, we could now broaden our horizon and consider $\mathscr{N} = (\mathscr{B}(\mathsf{H}), \mathbb{P})$ rather than just $\mathscr{A}$. Obviously such a construction does not necessarily carry a physical interpretation; this must be considered separately; see section 5. For the time being, however, we will use this convenient construction to provide us with a rich quantum stochastic model. The following discussion is heavily inspired by the work of Maassen [50].

Consider a classical Poisson process on a finite time interval $[0, T]$. We wish to describe the space of paths Ω. This is not difficult; a Poisson process on a finite time

interval has (a.s.) finitely many jumps n. Hence we can specify every relevant path by specifying its jump times. Let us thus introduce

$$\Omega = \bigcup_{n=0}^{\infty} \Omega_n, \qquad \Omega_0 = \{\varnothing\}, \quad \Omega_n = \{\{t_1, \ldots, t_n\} : t_1 < t_2 < \cdots < t_n \in [0,T]\}. \tag{4.1}$$

In other words, Ω is the set of ordered sequences in $[0,T]$ with a finite number of elements. We still need to introduce a σ-algebra $\mathcal{F}$ and a measure $\mathbf{P}$. To this end, consider Ω_n as a subset of the cube $([0,T]^n, e^{-T}\mu_n)$ where μ_n is the Lebesgue measure, so that Ω_n inherits a σ-algebra $\mathcal{F}_n$ and a measure $\mathbf{P}_n$ from the cube. Under $\mathbf{P}_n$ the jump times $t_1, \ldots, t_n$ are uniformly distributed (as must be the case for a Poisson process with fixed rate) and $\mathbf{P}_n(\Omega_n) = T^n e^{-T}/n!$. The measure $\mathbf{P}$ induced on Ω is precisely the probability measure of a Poisson process with unit rate.

We now introduce the Hilbert space $\mathsf{F} = L^2(\Omega, \mathcal{F}, \mathbf{P})$. It is called the *symmetric* or *Boson Fock space* and plays a central role in the following. We will also need the spaces $\mathsf{F}_{t]}$, $\mathsf{F}_{[t}$, and $\mathsf{F}_{[s,t]}$, defined identically to F except that the interval $[0,T]$ is replaced by $[0,t]$, $[t,T]$, and $[s,t]$, respectively. It is not difficult to see that for any $0 < s < t < T$ we have[8] $\Omega = \Omega_{s]} \times \Omega_{[s,t]} \times \Omega_{[t}$, and as the Poisson process has independent increments the measure splits up similarly. It follows that

$$\mathsf{F} = \mathsf{F}_{s]} \otimes \mathsf{F}_{[s,t]} \otimes \mathsf{F}_{[t} \qquad \forall\, 0 < s < t < T. \tag{4.2}$$

This important property is known as a continuous tensor product structure; it will play a key role in the definition of quantum stochastic integrals, as it gives a natural notion of adaptedness. Indeed, the algebra $\mathscr{W} = \mathscr{B}(\mathsf{F})$ splits up accordingly,

$$\mathscr{W} = \mathscr{W}_{s]} \otimes \mathscr{W}_{[s,t]} \otimes \mathscr{W}_{[t} = \mathscr{B}(\mathsf{F}_{s]}) \otimes \mathscr{B}(\mathsf{F}_{[s,t]}) \otimes \mathscr{B}(\mathsf{F}_{[t}). \tag{4.3}$$

A process of operators $\{X_t\}$ affiliated to $\mathscr{W}$ is said to be *adapted* if X_t is affiliated to $\mathscr{W}_{t]}$ for every t; equivalently, X_t is of the form $X_{t]} \otimes I$ as an operator on $\mathsf{F}_{t]} \otimes \mathsf{F}_{[t}$.

Next, let us introduce a set of interesting vectors. The reader should keep in mind Example 3.11 which is conceptually quite similar. Let $f \in L^\infty([0,T])$ be a complex Lebesgue measurable function. Then we can define the *exponential vector*

$$e(f)(\varnothing) = 1, \quad e(f)(\tau) = \prod_{t \in \tau} f(t), \qquad f \in L^\infty([0,T]),\ \tau \in \Omega. \tag{4.4}$$

It is not difficult to verify that $e(f) \in \mathsf{F}$, as

$$\langle e(g), e(f)\rangle = \sum_{n=0}^{\infty} \frac{e^{-T}}{n!} \left(\int_0^T g^*(t) f(t)\, dt \right)^n = \exp\left[\int_0^T (g^*(t) f(t) - 1)\, dt \right],$$

hence $\langle e(f), e(f)\rangle = e^{\|f\|_2^2 - T} < \infty$ for any $f \in L^\infty([0,T])$. We define D, the exponential domain, as the linear span of all $e(f)$, $f \in L^\infty([0,T])$, and we note that D is dense in F. The exponential vectors have the important property that they factorize over the continuous tensor product structure (4.2): indeed, it is evident from (4.4) that $e(f) = e(f_{s]}) \otimes e(f_{[s,t]}) \otimes e(f_{[t})$ where $f_{t]}$ is the restriction of f to $[0,t]$, etc.

[8] A more precise statement would be something like $\Omega = \Omega_{s]} \times \Omega_{(s,t]} \times \Omega_{(t}$; however, the only paths for which this makes a difference are those that have jumps exactly at times s or t, which is a set of $\mathbf{P}$-measure zero. For notational simplicity, we are free to always use closed time intervals $[s,t]$.

We are now ready to define a Poisson process. Let us first define it as a random variable on Ω; we simply write $N_t(\tau) = |\tau \cap [0,t]|$, where $|\tau|$ denotes the number of elements in the set $\tau \in \Omega$. The random variable N_t counts the number of jumps up to time t, and hence $\{N_t\}$ is by construction a Poisson process with unit rate under the measure $\mathbf{P}$. We now turn this into an operator process by pointwise multiplication:

$$(\Lambda_t\psi)(\tau) = N_t(\tau)\,\psi(\tau) = |\tau \cap [0,t]|\,\psi(\tau), \qquad \psi \in \mathsf{F},\ \tau \in \Omega,\ t \in [0,T]. \tag{4.5}$$

$\{\Lambda_t\}$ is called the *gauge process*; it is not difficult to see that though Λ_t is an unbounded operator,[9] it is affiliated to $\mathscr{W}_{t]}$ and hence the gauge process is adapted; in fact, the increments $\Lambda_t - \Lambda_s$ are even affiliated to $\mathscr{W}_{[s,t]}$. Furthermore, Λ_s and Λ_t commute for all $s,t \in [0,T]$, and indeed $\mathrm{vN}(\Lambda_t,\ t \in [0,T]) = L^\infty(\Omega,\mathcal{F},\mathbf{P}) \subset \mathscr{W}$ is commutative. Hence we could use the spectral theorem to map Λ_t back to a classical stochastic process. It is somewhat futile to diagonalize the operators using the spectral theorem, however, as we have already constructed them in diagonal form.

We have yet to introduce a state; a particularly interesting class of states are the *coherent states* $\mathbb{P}_f(X) = \langle e(f), X\,e(f)\rangle\, e^{T-\|f\|_2^2}$. Because of the continuous tensor product property, the coherent states split up as follows:

$$X = X_{s]} \otimes X_{[s,t]} \otimes X_{[t}, \qquad \mathbb{P}_f(X) = \mathbb{P}_{f_{s]}}(X_{s]})\ \mathbb{P}_{f_{[s,t]}}(X_{[s,t]})\ \mathbb{P}_{f_{[t}}(X_{[t}). \tag{4.6}$$

But as $\Lambda_t - \Lambda_s$ is affiliated to $\mathscr{W}_{[s,t]}$, it follows that under the state $\mathbb{P}_f$ the gauge process has independent increments. Furthermore, if we denote by $P_{\Lambda_t-\Lambda_s}(B)$ the spectral measure of $\Lambda_t - \Lambda_s$, then we have

$$\mathbb{P}_f(P_{\Lambda_t-\Lambda_s}(B)) = \mathbb{P}_{f_{[s,t]}}(\chi_B(|\tau \cap [s,t]|)) = \sum_{n\in B} \frac{e^{-\int_s^t |f(r)|^2\,dr}}{n!} \left(\int_s^t |f(r)|^2\,dr\right)^n.$$

Evidently, Λ_t is an inhomogeneous Poisson process with rate $|f(t)|^2$ under the state $\mathbb{P}_f$. Note in particular that as $e(1)(\tau) = 1$, we have for any $X \in L^\infty(\Omega,\mathcal{F},\mathbf{P})$ the relation $\mathbb{P}_1(X) = \langle 1, X\,1\rangle = E_{\mathbf{P}}(X)$; hence the fact that under $\mathbb{P}_1$ the gauge process is a Poisson process with unit rate is exactly what we expect from the definition of $\mathbf{P}$. Under $\mathbb{P}_0$, on the other hand, the gauge process does not register any counts; $\mathbb{P}_0 = \phi$ is called the *vacuum state*, and $e(0) = \Phi$ is called the *vacuum vector*.

4.2. Weyl operators and Wiener processes. We have now exhausted the diagonal observables affiliated to the space $(L^\infty(\Omega,\mathcal{F},\mathbf{P}),\mathbb{P}_f)$: every such observable is some functional of the Poisson process Λ_t with rate $|f|^2$. Let us thus explore whether we can find interesting observables affiliated to $\mathscr{W}$ that do not commute with Λ_t. To this end, we follow again essentially Example 3.11. Given $f,g \in L^\infty([0,T])$ we look for a unitary operator $W(f)$ that implements the translation group $W(f)e(g) \propto e(f+g)$. A calculation identical to the one in Example 3.11 shows that we should define

$$W(f)e(g) = e^{-\int_0^T \left(f^*(t)g(t) + \frac{1}{2}f^*(t)f(t)\right)dt} e(f+g) = e^{-\langle f,g\rangle_2 - \|f\|_2^2/2}\,e(f+g). \tag{4.7}$$

The unitary operator $W(f)$ is called a *Weyl operator* and provides a projective unitary representation in the sense that $W(f)W(g) = W(f+g)\,e^{i\,\mathrm{Im}\langle g,f\rangle_2}$. Note that it is

[9] As can be verified by explicit computation, the domain of Λ_t contains at least D, the exponential domain. The reader may ask himself why we have only defined exponential vectors $e(f)$ for $f \in L^\infty([0,T])$ rather than $f \in L^2([0,T])$: this is because the latter may not be in the domain of Λ_t. Our domain D is sometimes called the *restricted* exponential domain in the literature.

sufficient to define the action of $W(f)$ only on exponential vectors; we can extend to D by linearity, and as D is dense and $W(f)$ is bounded the Weyl operators are uniquely extended to all of F. An important property, which follows immediately from the definition of $W(f)$ and the continuous tensor product property, is that

$$W(f)e(g) = W(f_{s]})e(g_{s]}) \otimes W(f_{[s,t]})e(g_{[s,t]}) \otimes W(f_{[t})e(g_{[t}). \tag{4.8}$$

In particular, we see that $W(f\chi_{[0,t]})$ is an adapted operator process.

Now fix $f \in L^\infty([0,T])$ and consider the unitary group $\{W(kf)\}_{k\in\mathbf{R}}$; this group is in fact continuous [55], and hence by Stone's theorem (Theorem 3.10) there exists a self-adjoint $B(f)$ such that $W(kf) = e^{ikB(f)}$. The operators $B(f)$, $f \in L^\infty([0,T])$, are called *field operators*. Finding the distribution of the observable $B(f)$ is straightforward, as the characteristic function of $B(f)$ (under the coherent state $\mathbb{P}_g$) is given by

$$b_f(k) = \mathbb{P}_g(W(kf)) = \langle e(g), e(g+kf)\rangle e^{T-\|g\|_2^2 - k\langle f,g\rangle_2 - k^2\|f\|_2^2/2} = e^{2ik\,\mathrm{Im}\langle g,f\rangle_2 - k^2\|f\|_2^2/2}.$$

Hence $B(f)$ is a Gaussian random variable with mean $2\,\mathrm{Im}\langle g, f\rangle_2$ and variance $\|f\|_2^2$. In the vacuum, i.e., $g = 0$, the mean vanishes; for simplicity, we will restrict ourselves to the vacuum case in the following.

Consider the operator process $\{B_t^\varphi = B(e^{i\varphi}\chi_{[0,t]}) : t \in [0,T]\}$ for some fixed, real function $\varphi \in L^\infty([0,T])$. B_t^φ is adapted, as we have already established that $W(f\chi_{[0,t]})$ is adapted for any f; moreover, $B(e^{i\varphi}\chi_{[s,t]}) = B_t^\varphi - B_s^\varphi$ is affiliated to $\mathscr{W}_{[s,t]}$ due to (4.8). This immediately tells us two important things. First, B_t^φ and B_s^φ commute for all $s,t \in [0,T]$; indeed, $B_t^\varphi - B_s^\varphi$ must commute with $B_s^\varphi - B_0^\varphi$, and commutativity follows from $B_0^\varphi = 0$. This means that $\mathrm{vN}(B_t^\varphi,\ t \in [0,T])$ is a commutative algebra and hence we can represent B_t^φ for every t as a classical random variable on the same probability space $(\Omega^\varphi, \mathcal{F}^\varphi, \mathbf{P}^\varphi)$; in particular, $\iota(B_t^\varphi)$ is a classical stochastic process. Second, (4.6) implies that the process B_t^φ has independent increments. But we have established $B_t^\varphi - B_s^\varphi$ is (in the vacuum) a mean zero Gaussian random variable with variance $t-s$, and as B_t^φ has independent increments we have established that $\iota(B_t^\varphi)$ is precisely a Wiener process on $(\Omega^\varphi, \mathcal{F}^\varphi, \mathbf{P}^\varphi)$.

Let us introduce the following notation. Define $Q_t = B(i\chi_{[0,t]})$, $P_t = B(-\chi_{[0,t]})$, and $A_t = (Q_t + iP_t)/2$. Note that Q_t and P_t are self-adjoint by Stone's theorem, whereas A_t has the adjoint $A_t^* = (Q_t - iP_t)/2$. We now compute

$$B(f)e(g) = \frac{1}{i}\frac{d}{dk}W(kf)e(g)\bigg|_{k=0} = i\langle f,g\rangle_2\, e(g) - i\,\frac{d}{dk}e(g+kf)\bigg|_{k=0}.$$

Evidently $A_t e(g) = \langle \chi_{[0,t]}, g\rangle_2\, e(g) = \int_0^t g(s)ds\, e(g)$. But then we can write

$$(A_t e(g))(\tau) = \int_0^t g(s)\,ds \prod_{r\in\tau} g(r) = \int_0^t g(s) \prod_{r\in\tau} g(r)\,ds = \int_0^t e(g)(\tau\cup\{s\})\,ds.$$

In particular, this formula extends to any $\psi \in \mathsf{F}$ for which the integral on the right-hand side (with $e(g)$ replaced by ψ) defines a normalizable vector. A_t is called the Fock space *annihilation operator*, as it generalizes the corresponding notion introduced in Example 3.11. The reader should verify that its adjoint can be expressed as

$$(A_t^*\psi)(\tau) = \sum_{s\in\tau\cap[0,t]} \psi(\tau\backslash\{s\})$$

on a sufficiently large domain. Not surprisingly, A_t^* is called the *creation operator*. It is conventional in quantum stochastic calculus to use A_t and its adjoint rather than Q_t and P_t; we shall conform to this standard.

In summary, we have constructed a quantum probability space $(\mathscr{W}, \phi)$ that admits an entire family (indexed by φ) of Wiener processes. Note, however, that these processes do not necessarily commute for different φ; in fact, it is not difficult to establish that $[B(f), B(g)]\psi = 2i\,\mathrm{Im}\langle f, g\rangle_2\,\psi$ on a suitably large domain (e.g., $\psi \in \mathsf{D}$). Therefore, even though every B_t^φ defines a Wiener process, these cannot be represented on the same classical probability space for different $\varphi_{1,2}$ unless $\mathrm{Im}(e^{i(\varphi_1-\varphi_2)}) = 0$.

We have also defined a Poisson process Λ_t, but unfortunately it vanishes in the vacuum. Consider, however, the process $\Lambda_t(f) = W(f)^*\Lambda_t W(f)$; for any Borel function b we can write $\phi(b(\Lambda_{t_1}(f), \ldots, \Lambda_{t_n}(f))) = \mathbb{P}_f(b(\Lambda_{t_1}, \ldots, \Lambda_{t_n}))$. Evidently $\Lambda_t(f)$ has the same statistics in the vacuum as does Λ_t under the coherent state $\mathbb{P}_f$. This shows that we can define even a whole family of Poisson processes in the vacuum. We do not lose much by restricting ourselves to the vacuum as an underlying state (as we will do in the remainder of the article), as we can always transform to a coherent state by "sandwiching" with Weyl operators. Note that like the family B_t^φ, the processes $\Lambda_t(f)$ do not commute amongst each other. We see that the quantum probability space $(\mathscr{W}, \phi)$ gives rise to a rich family of incompatible stochastic processes.

4.3. Quantum stochastic calculus. Now that we have obtained Wiener and Poisson processes, we can try to develop stochastic integrals with respect to these processes and an associated stochastic calculus. Note that if we were only interested in, e.g., integrating with respect to Q_t an adapted process which commutes with Q_t, then we could simply use the classical Itô integral definition through the spectral theorem. This will not suffice for our purposes, however, as we will want to consider stochastic differential equations that are driven simultaneously by the noncommuting noises Q_t and P_t (and even Λ_t). Moreover, we would like to have an Itô rule that tells us how to multiply stochastic integrals with respect to Q_t and P_t.

Our motivation for developing generalized quantum stochastic calculus is that this allows us to rigorously define and manipulate Schrödinger equations, as in (2.6), with a white-noise Hamiltonian formally defined by $H(t) = H_0 + H_1\,\dot{Q}_t + H_2\,\dot{P}_t$. In section 5 we will see that such models emerge naturally in applications. In this section we sketch the development of quantum stochastic calculus as it was introduced in a seminal paper by Hudson and Parthasarathy [42]. For a full development of this calculus we refer the reader to [42, 41, 55]. The Hudson–Parthasarathy approach has some technical issues, not surprisingly involving the unboundedness of operators, the full extent of which is still being explored. Though we cannot go into detail here, we will attempt to sketch some of the issues and give references to recent literature.

We work in the following setting. We wish to integrate processes against the three noises A_t, A_t^*, and Λ_t (the *fundamental noises*); i.e., we want to define $\int_0^t L_s\,dM_s$ where M_t is one of the fundamental noises. The noises are defined on the quantum probability space $(\mathscr{W}, \phi)$, but we will want to couple these noises to an external quantum system, the *initial system*,[10] with which they interact. To this end, let us introduce the initial Hilbert space h, $\mathscr{B} = \mathscr{B}(\mathsf{h})$, and the associated initial quantum

[10]This name has the following origin. Recall from section 2 that observables X evolve in time as $X_t = U_t^* X U_t$ (we will define a unitary evolution U_t in section 5). We would like to think of $X \otimes I \in \mathscr{B} \otimes \mathscr{W}$ as describing the external system; however, $U_t^*(X \otimes I)U_t$ will not be of the form $Y \otimes I$ except at $t = 0$. Hence the initial system observable $X \otimes I$ describes the external system at the initial time $t = 0$.

probability space $(\mathscr{B},\rho)$. We will choose our integrands L_t to be adapted processes on $(\mathscr{B}\otimes\mathscr{W},\rho\otimes\phi)$; i.e., each L_t is affiliated to $\mathscr{B}\otimes\mathscr{W}_{t]}$ and acts as I on $\mathscr{W}_{[t}$.

As usual, we begin with simple processes. Given $s<t$, recall that for the fundamental processes M_t-M_s is affiliated to $\mathscr{W}_{[s,t]}$, whereas for adapted processes L_s is affiliated to $\mathscr{B}\otimes\mathscr{W}_{s]}$; hence we can naturally write $L_s(M_t-M_s)=L_s\otimes(M_t-M_s)$. In particular, the increment M_t-M_s commutes with L_s, and we have no problems with operator multiplication of these unbounded operators. Let $\{t_i : i=0,\dots,n,\ t_i<t_{i+1}\}$ be a sequence of times with $t_0=0$ and $t_n=T$. By definition, we set

$$L_t=\sum_{i=0}^{n-1}L_{t_i}\chi_{[t_i,t_{i+1})}(t)\quad\Longrightarrow\quad\int_0^t L_s\,dM_s=\sum_{i=0}^{n-1}L_{t_i}\otimes(M_{t_{i+1}\wedge t}-M_{t_i\wedge t}).$$

This definition makes sense as long as the operators L_t and M_t have a sufficiently large common dense domain that the sum is well defined. To enforce this, we will require that the domain of every L_t contains at least the exponential domain D.

Now comes the hard part in any integration theory: given a quadruple of suitably restricted adapted processes (E,F,G,H), such that these admit simple approximations (E^n,F^n,G^n,H^n), we wish to define the integral

$$I_t=\int_0^t(E_t\,d\Lambda_t+F_t\,dA_t+G_t\,dA_t^*+H_t\,dt)\tag{4.9}$$

as a limit, in some sense, of the corresponding integrals I_t^n over the simple processes. Recall that in the classical theory, the Itô isometry allows us to define the stochastic integral as a mean square limit of simple processes, and a little more work shows that every square-integrable process admits a mean square approximation by simple processes. Things are not quite so "simple" in the noncommutative case, however.

To see what goes wrong, consider for simplicity the case $\mathsf{h}=\mathbf{C}$ so that we can forget about the initial state ρ. We already encountered the noncommutative L^2 (semi)norm $\|X\|_\phi^2=\phi(X^*X)$ when we discussed conditional expectations. We are thus looking for a suitable unbounded operator I_t such that we have mean square convergence, $\|I_t-I_t^n\|_\phi^2=\langle(I_t-I_t^n)\Phi,(I_t-I_t^n)\Phi\rangle\to0$ as $n\to\infty$. But this is a very ill-defined problem, as it only depends on the action of I_t on the vacuum vector Φ; in particular, what do we choose as the domain of I_t, and how do we define I_t on vectors orthogonal to Φ? There could be a large number of inequivalent ways of doing this, giving rise to limiting operators with very different properties.[11]

The solution of Hudson and Parthasarathy works as follows. First of all, we fix the domain of I_t at the outset: every stochastic integral will have $\mathsf{h}\otimes\mathsf{D}$ as its domain (one could choose a dense domain in h as well; we will not worry about this). To specify I_t as a limit of simple integrals I_t^n, we choose I_t as the unique operator on $\mathsf{h}\otimes\mathsf{D}$ such that $\langle(I_t-I_t^n)\,v\otimes\psi,(I_t-I_t^n)\,v\otimes\psi\rangle\to0$ for every $\psi\in\mathsf{D}$, $v\in\mathsf{h}$ (it is sufficient to verify this for $\psi=e(f)$, $f\in L^\infty([0,T])$). In essence this is like a mean square limit, but simultaneously for every coherent state. A suitable estimate replaces the Itô isometry [42, Corollary 1] and shows that this limit exists as long as $\int_0^T\|(E_s-E_s^n)\,v\otimes\psi\|^2ds\to0$ as $n\to\infty$ for every $\psi\in\mathsf{D}$, $v\in\mathsf{h}$ (and similarly for F,G,H), independent of the approximation. Finally, [42, Proposition 3.2] shows

[11] This was not a problem for the definition of conditional expectations; as all versions of the conditional expectation are affiliated to a single commutative algebra, they are a.s. equivalent by the spectral theorem. On the other hand, various "versions" of I_t that satisfy $\|I_t-I_t^n\|_\phi\to0$ need not even commute, and such operators are fundamentally inequivalent.

that every square-integrable process, i.e., $\int_0^T \|E_s\, v \otimes \psi\|^2 ds < \infty$ for all $\psi \in \mathsf{D}$, $v \in \mathsf{h}$, admits a suitable approximation by simple processes. We thus arrive at the following.

DEFINITION 4.1 (quantum Itô integral). *An operator process $\{X_t\}$ is stochastically integrable if it is adapted and square integrable. Given a quadruple (E, F, G, H) of such processes, the stochastic integral* (4.9) *is uniquely defined as the limit of simple approximations on the domain $\mathsf{h} \otimes \mathsf{D}$.*

Remark 4.2. It is often convenient to denote an expression of the form

$$X_t = X + \int_0^t (E_s\, d\Lambda_s + F_s\, dA_s + G_s\, dA_s^* + H_s\, ds),$$

symbolically as

$$dX_t = E_t\, d\Lambda_t + F_t\, dA_t + G_t\, dA_t^* + H_t\, dt, \qquad X_0 = X.$$

Both notations are used interchangeably in the literature.

A property that we will exploit in future is $\Lambda_t\Phi = A_t\Phi = 0$. It is immediate from the definition that stochastic integrals with respect to A_t and Λ_t acting on Φ vanish. Hence the vacuum expectations of stochastic integrals with respect to A_t and Λ_t vanish as well. Furthermore, as $\langle \Phi, A_t^*\Phi\rangle = \langle A_t\Phi, \Phi\rangle = 0$, we see that at least for simple processes (and indeed this holds for any integrand) the vacuum expectation of stochastic integrals with respect to A_t^* vanish. Note, however, that $A_t^*\Phi \neq 0$.

Our next task is to develop a stochastic calculus; the integrals defined above are not of much use, unless we have an Itô product rule with which they can be manipulated. Once again we run into unpleasant problems. If I_t and J_t are integrals of the form (4.9), there is no reason to expect that their product I_tJ_t is a well-defined operator on the domain $\mathsf{h} \otimes \mathsf{D}$. The idea of Hudson and Parthasarathy is inspired by the identity $\langle \psi', X^*Y\psi\rangle = \langle X\psi', Y\psi\rangle$ for bounded operators; rather than finding an expression for I_tJ_t, they calculate $\langle I_t\, v' \otimes \psi', J_t\, v \otimes \psi\rangle$ for every $v, v' \in \mathsf{h}$, $\psi, \psi' \in \mathsf{D}$, which is always well defined. One finds explicitly a lengthy expression [42, Theorems 4.3 and 4.4], which is essentially the quantum Itô rule expressed in terms of $\mathsf{h} \otimes \mathsf{D}$-matrix elements.

In practice, however, we are mostly interested in calculating actual operator products I_tJ_t. We will need the concept of an *adjoint pair*; two operators X and $X^\dagger$ are said to be an adjoint pair if $\langle v' \otimes \psi', X\, v \otimes \psi\rangle = \langle X^\dagger\, v' \otimes \psi', v \otimes \psi\rangle$ for every $v, v' \in \mathsf{h}$, $\psi, \psi' \in \mathsf{D}$. It is not difficult to verify that if (E, F, G, H) and $(E^\dagger, F^\dagger, G^\dagger, H^\dagger)$ are adjoint pairs, then I_t and $I_t^\dagger$ form an adjoint pair, where

$$I_t^\dagger = \int_0^t (E_t^\dagger\, d\Lambda_t + F_t^\dagger\, dA_t^* + G_t^\dagger\, dA_t + H_t^\dagger\, dt). \tag{4.10}$$

In essence, the adjoint $\dagger$ replaces the Hilbert space adjoint $*$ on the domain $\mathsf{h} \otimes \mathsf{D}$. Now suppose that we can verify explicitly that the product I_tJ_t is well defined; then we can read off an expression for I_tJ_t from the matrix elements $\langle I_t^\dagger\, v' \otimes \psi', J_t\, v \otimes \psi\rangle$. This gives the following explicit form of the quantum Itô rule.

THEOREM 4.3 (quantum Itô rule [55, Proposition 25.26]). *Let (F, G, H, I), (B, C, D, E), and $(B^\dagger, C^\dagger, D^\dagger, E^\dagger)$ be quadruples of stochastically integrable processes such that the latter two quadruples are adjoint pairs. Define the stochastic integrals*

$$dX_t = B_t\, d\Lambda_t + C_t\, dA_t + D_t\, dA_t^* + E_t\, dt,$$
$$dY_t = F_t\, d\Lambda_t + G_t\, dA_t + H_t\, dA_t^* + I_t\, dt,$$

and suppose that we have verified that the product X_tY_t is well defined and that $X_tF_t, \ldots, X_tI_t$, $B_tY_t, \ldots, E_tY_t$, and $B_tF_t, B_tG_t, \ldots, E_tI_t$ are well defined and stochastically integrable. Then the process X_tY_t satisfies the relation

$$d(X_tY_t) = X_t\,dY_t + (dX_t)\,Y_t + dX_t\,dY_t,$$

where $X_t\,dY_t = X_tF_t\,d\Lambda_t + X_tG_t\,dA_t + X_tH_t\,dA_t^* + X_tI_t\,dt$, $(dX_t)\,Y_t = B_tY_t\,d\Lambda_t + C_tY_t\,dA_t + D_tY_t\,dA_t^* + E_tY_t\,dt$, and $dX_t\,dY_t = B_tF_t\,d\Lambda_t + C_tF_t\,dA_t + B_tH_t\,dA_t^* + C_tH_t\,dt$ are evaluated according to the following quantum Itô table.

$dX \setminus dY$	dA_t	$d\Lambda_t$	dA_t^*	dt
dA_t	0	dA_t	dt	0
$d\Lambda_t$	0	$d\Lambda_t$	dA_t^*	0
dA_t^*	0	0	0	0
dt	0	0	0	0

In particular, the theorem holds if B_t, C_t, D_t, E_t, and X_t are bounded processes [42]*, in which case the adjoints $B^\dagger$, $C^\dagger$, etc. are simply taken to be the Hilbert space adjoints B^*, C^*, etc., and X_t extends uniquely to a bounded operator in $\mathscr{W}_{t]}$.*

Remark 4.4. The choice to restrict attention to a fixed domain $\mathsf{h} \otimes \mathsf{D}$ allowed Hudson and Parthasarathy [42] to develop a viable quantum stochastic calculus. This choice, however, has quite a few drawbacks; we highlight one of the problems. Suppose X is self-adjoint; implicit in this statement is that the domains of X and X^* coincide. It can happen that if we restrict the domain of X, then the restricted operator admits many inequivalent self-adjoint extensions; see [56, pages 257–259] for an example. Hence the restriction to a fixed domain can become a real, physical problem, that prevents us from uniquely interpreting unbounded operators on $\mathsf{h} \otimes \mathsf{D}$ as observables.

Such problems have prompted the development of alternative approaches to quantum stochastic integration, and the topic is still under active investigation. In a significant recent achievement Attal and Lindsay [5], building on several earlier approaches (see, e.g., [52, 14] and the references therein), develop a theory in which the integrals achieve their maximal domains. Unfortunately, the theory is very technical and a little daunting for everyday use. A different approach that even precedes Hudson and Parthasarathy is that of Barnett, Streater, and Wilde [8]. Their theory is attractive as it is completely algebraic in nature (the Hilbert space and its domains do not play a fundamental role), but lacks a satisfactory Itô rule.

Despite these issues, the Hudson–Parthasarathy approach works quite well. In practice one usually works with a "noisy Schrödinger equation" (5.2), the solution of which is unitary and thus bounded. As long as the integrals and integrands are bounded, they are uniquely defined by their specification on a dense domain. In this article, in keeping with our attitude towards unbounded operators, we will not worry about such issues and assume that we can apply the quantum Itô rules. □

Example 4.5. In section 5 we will encounter quantum stochastic differential equations (QSDEs), the treatment of which proceeds along the same lines as the classical theory. We claim that the Weyl operator $W(f_{t]})$ is the solution of the QSDE

$$dW(f_{t]}) = \Big\{ f(t)\,dA_t^* - f(t)^*\,dA_t - \frac{1}{2}|f(t)|^2\,dt \Big\} W(f_{t]}). \tag{4.11}$$

In particular, one can verify the Weyl relation $W(f)W(g) = W(f+g)\,e^{i\,\mathrm{Im}\langle g,f\rangle_2}$

directly using the quantum Itô rule. From (4.11) and $W(kf) = e^{ikB(f)}$ we obtain

$$B(f) = \int_0^T (if(t)^*\, dA_t - if(t)\, dA_t^*).$$

Hence $dB_t^\varphi = ie^{-i\varphi(t)}\, dA_t - ie^{i\varphi(t)}\, dA_t^*$, and the quantum Itô rules reduce to the classical Itô rule $(dB_t^\varphi)^2 = dt$. Finally, recall that we defined Poisson processes $\Lambda_t(f) = W(f)^*\Lambda_t W(f) = W(f_{t]})^*\Lambda_t W(f_{t]})$ (the latter equality is due to $W(f) = W(f_{t]}) \otimes W(f_{[t})$ and the fact that $W(f_{[t}) \in \mathscr{W}_{[t}$ is unitary and commutes with the adapted process Λ_t). Using the quantum Itô rule we obtain the explicit representation

$$d\Lambda_t(f) = d\Lambda_t + f(t)^*\, dA_t + f(t)\, dA_t^* + |f(t)|^2\, dt, \tag{4.12}$$

for which the quantum Itô rules reduce to the classical product rule $(d\Lambda_t(f))^2 = d\Lambda_t(f)$ for a Poisson process. □

5. The filtering problem in quantum optics. Many realistic physical scenarios are very well described by quantum stochastic differential equations driven by the processes A_t, A_t^*, and Λ_t discussed in the previous section. Of course, as in the classical theory, white-noise systems are only an idealization of physical interactions; a Markov limit of wide-band noise in the spirit of Wong and Zakai (see [36] for details) gives stochastic models in the Itô form. For a large class of quantum systems, particularly those arising in the field of quantum optics, such approximations are extremely good and describe laboratory experiments essentially to experimental precision. Though a detailed discussion of the physics involved in the modelling of such systems is beyond the scope of this article, we here very briefly describe the physical origin of the equations that are widely used in the physics community [34], describe the measurements that are made, and set up the quantum filtering problem to be solved.

5.1. The quantum optics model. The basic model of quantum optics consists of some fixed physical system, e.g., a collection of atoms, in interaction with the electromagnetic field. The atomic observables are self-adjoint operators on a Hilbert space h. The description of the electromagnetic field and its interaction with the atoms follows from basic physical arguments (see the excellent monograph [23] for a thorough treatment of this theory, known as *quantum electrodynamics*). It turns out that the free electromagnetic field, i.e., an optical field in empty space, is described by a stationary Gaussian (noncommutative) wide-band noise $\tilde{a}(t, \mathbf{r})$ that propagates through space at the speed of light c; i.e., if we restrict ourselves to a single spatial dimension, $\tilde{a}(t + \tau, z) = \tilde{a}(t, z - c\tau)$. If we now place the atoms at the origin $z = 0$, then the quantum dynamics is given by a Schrödinger equation of the form

$$\frac{d}{dt}\tilde{U}(t) = [-iH + L\,\tilde{a}^*(t,0) - L^*\,\tilde{a}(t,0)]\,\tilde{U}(t), \qquad \tilde{U}(0) = I, \tag{5.1}$$

where $L \in \mathscr{B}$ is an atomic (dipole) operator and $H \in \mathscr{B}$ is an atomic Hamiltonian, H being self-adjoint. This equation, which follows directly from the physical model, has wide-band right-hand side. Note that we have set $\hbar = 1$ for convenience, a convention ubiquitous in physics (the only consequence is a change of units).

We now want to approximate the wide-band noise by white noise. This can be done in a rigorous way [1, 2, 36], but we will not detail the procedure here (a brief sketch can be found in [62]). Suffice it to say that one arrives at the following QSDE:

$$dU_t = \Big\{L\, dA_t^* - L^*\, dA_t - \frac{1}{2}L^*L\, dt - iH\, dt\Big\}U_t, \qquad U_0 = I, \tag{5.2}$$

which is driven by the noncommuting white-noise processes A_t and A_t^*. Note that this is almost precisely of the same form as (5.1), except that we have added the Itô correction term $-\frac{1}{2}L^*LU_t\,dt$. A Picard iteration argument [42, 55] ensures existence and uniqueness of the solution. The adjoint U_t^* satisfies

$$dU_t^* = U_t^*\Big\{L^*\,dA_t - L\,dA_t^* - \frac{1}{2}L^*L\,dt + iH\,dt\Big\}, \qquad U_0^* = I.$$

Using the quantum Itô rule we can calculate $d(U_t^*U_t) = d(U_tU_t^*) = 0$; i.e., the solution U_t is unitary for all t (as the solution of a Schrödinger equation should be).

Henceforth we will take (5.2) as our physical model. U_t defines the time evolution or *flow* $j_t : X \mapsto U_t^*(X \otimes I)U_t$ of every atomic observable $X \in \mathscr{B}$ (recall the time evolution in section 2.1); i.e., an observation of $X \in \mathscr{B}$ at time t is described by the observable $X_t = j_t(X)$. Using the Itô rules, we find an explicit dynamical equation

$$dj_t(X) = j_t(\mathcal{L}_{L,H}(X))\,dt + j_t([L^*, X])\,dA_t + j_t([X, L])\,dA_t^*, \quad X \in \mathscr{B}, \tag{5.3}$$

where the so-called Lindblad generator [48] is given by

$$\mathcal{L}_{L,H}(X) = i[H, X] + L^*XL - \frac{1}{2}(L^*LX + XL^*L), \qquad X \in \mathscr{B}.$$

In quantum probability, this object plays the same role as the infinitesimal generator of a Markov diffusion in classical probability theory.

Remark 5.1. Though it is unusual, one could use a very similar notation in classical stochastic models. Suppose some system is described by an underlying configuration x_t that obeys $dx_t = b(x_t)\,dt + \sigma(x_t)\,dW_t$. Then the "observables" in the theory, i.e., things we could try to measure, are functions f of the configuration of the system. The observable f at time t is described by the random variable $j_t(f) = f(x_t)$. Using the classical Itô rules, we get $dj_t(f) = j_t(\mathcal{L}f)\,dt + j_t(\Sigma f)\,dW_t$ where $\mathcal{L}f(x) = \sum_i b^i(x)\partial_i f(x) + \frac{1}{2}\sum_{ij}\sigma^i(x)\sigma^j(x)\partial_i\partial_j f(x)$ is the generator of the Markov diffusion x_t, and $\Sigma f(x) = \sum_i \sigma^i(x)\partial_i f(x)$. This expression is the classical analog of (5.3); the sample paths x_t do not have a quantum counterpart, however. ∎

5.2. Measurements. Having described the system and its interaction with the field, let us now turn to the observations that we can perform. Unlike in classical models, where one observes the system directly (with the addition of some corrupting noise), in quantum models an observation is generally performed in the field. From the system's perspective, the interaction with the field looks like an (albeit noncommutative) noisy driving force. Similarly, however, the field is perturbed by its interaction with the atoms and carries off information as it propagates away after the interaction. By performing a measurement in the field, then, we can attempt to perform statistical inference of the atomic observables. The entire setup is depicted in Figure 5.1.

To calculate the perturbation of the field by the atoms we once again calculate $U_t^*YU_t$, where now, however, Y is a field observable. The field observable of interest depends on the type of measurement we choose to perform. Without entering into the details, we mention two types of measurement that are extremely common in quantum optics: direct photodetection (photon counting), for which the observation at time t is given by $Y_t^\Lambda = U_t^*\Lambda_tU_t$, and homodyne detection, for which $Y_t^W = U_t^*(A_t + A_t^*)U_t$ (more generally $Y_t^W = U_t^*(e^{-i\varphi}A_t + e^{i\varphi}A_t^*)U_t$). We refer the reader to [6, 7] for a detailed treatment of quantum optical measurements. Using the Itô rules we obtain

$$dY_t^\Lambda = d\Lambda_t + j_t(L)\,dA_t^* + j_t(L^*)\,dA_t + j_t(L^*L)\,dt, \tag{5.4}$$

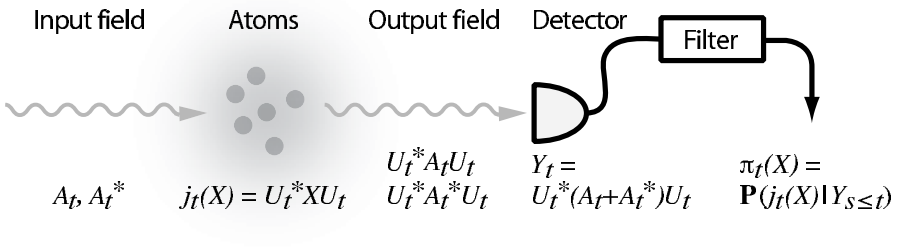

FIG. 5.1. *Diagram of the quantum filtering setup in quantum optics. An optical field, described by the field operators A_t, A_t^*, interacts with a system, e.g., a cloud of atoms. After the atom-field interaction the field operators, as well as system operator, X, are rotated by the unitary U_t. The field is detected, giving rise to the observation Y_t. Finally, the quantum filter (implemented on a classical signal processor) estimates atomic observables based on the field observations.*

$$dY_t^W = j_t(L + L^*)\,dt + dA_t + dA_t^*. \tag{5.5}$$

Intuitively, it would appear that Y_t^Λ is like a Poisson process whose intensity is controlled by $j_t(L^*L)$ (recall Example 4.5), whereas Y_t^W looks like a noisy observation of $j_t(L + L^*)$. One should be careful with this conclusion, however, as $j_t(L)$ need not commute with A_t or A_t^*, nor with itself at different times.

It is essential, however, that the observation process commutes with itself at different times and is hence equivalent to a classical stochastic process through the spectral theorem. An observation process that does not obey this property cannot be observed in a single realization of an experiment and is physically meaningless. Let us show that the observations processes we have defined above do obey this property, which is called the *self-nondemolition* property. Let Z be any operator of the form $I \otimes Z_{s]} \otimes I$ on $\mathsf{h} \otimes \mathsf{F}_{s]} \otimes \mathsf{F}_{[s}$ and let $t \ge s$. Then the Itô rules give directly

$$U_t^* Z U_t = U_s^* Z U_s + \int_s^t U_\tau^* \mathcal{L}_{L,H}(Z) U_\tau \, d\tau + \int_s^t U_\tau^* [L^*, Z] U_\tau \, dA_\tau + \int_s^t U_\tau^* [Z, L] U_\tau \, dA_\tau^*.$$

Now let $Z = A_s + A_s^*$ or $Z = \Lambda_s$. In both cases $\mathcal{L}_{L,H}(Z) = [Z, L] = 0$ as L and H are system observables and Z is a field observable. Hence $Y_s^W = U_t^*(A_s + A_s^*)U_t$ and $Y_s^\Lambda = U_t^* \Lambda_s U_t$ for all $t \ge s$. It is now easily verified, using the unitarity of U_t and the fact that $A_s + A_s^*$ and Λ_s are commutative processes, that $[Y_t^W, Y_s^W] = [Y_t^\Lambda, Y_s^\Lambda] = 0$ for all t, s. We denote by $\mathscr{Y}_t^W$ and $\mathscr{Y}_t^\Lambda$ the commutative von Neumann algebras generated by the observation processes Y_s^W and Y_s^Λ, $s \le t$, respectively. Do note, however, that Y_t^W and Y_t^Λ do not commute with each other; in any experiment, we can choose to perform only one of these measurements. Once we have made this choice, however, we can use the spectral theorem to represent the observations Y_t as a classical stochastic process $\iota(Y_t)$ on a probability space.

5.3. Statement of the filtering problem. Moving on to the next step in our program, we now wish to use the information gained from the measurement process to infer something about the system. To find a least mean square estimate of a system observable $X \in \mathscr{B}$ at time t, given the observations Y_t up to this time, we must calculate the conditional expectation

$$\pi_t(X) = \mathbb{P}(j_t(X) | \mathscr{Y}_t), \tag{5.6}$$

where $\mathscr{Y}_t = \mathrm{vN}(Y_s : 0 \le s \le t)$. The remainder of this article is devoted to finding a recursive equation for $\pi_t(X)$ (the *filtering equation*). Recall, however, that the conditional expectation is only defined if $j_t(X)$ is in the commutant of $\mathscr{Y}_t$, the interpretation being that statistical inference of an observable is only physically meaningful if the conditional statistics could possibly be tested through a compatible experiment. Through an entirely identical procedure to the one used to show the self-nondemolition property, we can show that $j_t(X)$ is in the commutant of $\mathscr{Y}_t$ for any $X \in \mathscr{B}$. This is known as the *nondemolition property*, which can be written as

$$[j_t(X), Y_s] = 0 \quad \forall s \le t,\ X \in \mathscr{B}. \tag{5.7}$$

We note that we have now obtained a system-theoretic model of our system and observations, defined on the quantum probability space $(\mathscr{B} \otimes \mathscr{W}, \mathbb{P} = \rho \otimes \phi)$ by

$$dj_t(X) = j_t(\mathcal{L}_{L,H}(X))\,dt + j_t([L^*, X])\,dA_t + j_t([X, L])\,dA_t^*, \tag{5.8}$$

$$dY_t = j_t(L + L^*)\,dt + dA_t + dA_t^* \tag{5.9}$$

in the case of homodyne detection, or by (5.8), and

$$dY_t = d\Lambda_t + j_t(L)\,dA_t^* + j_t(L^*)\,dA_t + j_t(L^*L)\,dt \tag{5.10}$$

in the case of counting observations. These equations define a system-observation model in direct analogy to such models used throughout classical nonlinear filtering and stochastic control theory.

Remark 5.2. Unlike in a classical filtering scenario, we have not added any independent corrupting noise to the observations. Nonetheless, the filtering problem does not reduce to a problem with complete observations because the system is driven by noise that does not commute with the observations. Hence the problem of partial observations is intrinsic to quantum measurement theory. The quantum filtering problem considered here is the simplest possible one; one could add additional corrupting noise as in the classical case, have the system interact with multiple fields (some of which are observed, others unobserved), etc. These are not essential complications, however, and filters for such models are obtained much in the same way. □

6. The reference probability method. The goal of this section is to derive the quantum filtering equation, a recursive equation for $\pi_t(X)$, using a method that is close to the classical reference probability method of Duncan [29], Mortensen [53], and Zakai [66]. We consider first the homodyne detection case, then the photon counting case. In section 7 we will rederive the filtering equation for the homodyne detection case using martingale methods; the chief advantage of the reference probability method is that it is somewhat simpler to apply. The following approach is based on [18].

6.1. Homodyne detection. Let us briefly recall the classical reference probability procedure; for an introduction see, e.g., [32]. In order to simplify the filtering problem, one starts by introducing a new probability measure, using a Girsanov transformation [49, section 6.3], under which the measurement record is a Wiener process. Then various (elementary) properties of the conditional expectation allow the filtering problem to be expressed, and solved, with respect to the new measure. We now apply this logic to the quantum filtering problem. Note that we have already applied the method in Example 3.19; the following is essentially a continuous time version of that example.

We consider the homodyne detection setup given by (5.8) and (5.9). We could try to find a new state under which Y_t is a Wiener process; however, it will be more convenient to work not in terms of Y_t but in terms of $Z_t = A_t + A_t^*$, as it is very easy to manipulate Z_t using the methods of section 4. Thus before we really start filtering, let us transform the problem in terms of Z_t. Introduce the state $\mathbb{Q}^t$ defined by

$$\mathbb{Q}^t(X) = \mathbb{P}(U_t^* X U_t), \tag{6.1}$$

with U_t as in section 5, and we fix from now on $\mathbb{P} = \rho \otimes \phi$. Now recall from Example 3.19 that $\mathbb{Q}(X) = \mathbb{P}(U^*XU)$ implies $\mathbb{P}(U^*XU|U^*\mathscr{C}U) = U^*\mathbb{Q}(X|\mathscr{C})U$ (this is easily checked using the definition of the conditional expectation). Thus we have

$$\mathbb{P}(j_t(X)|\mathscr{Y}_t) = U_t^* \mathbb{Q}^t(X|\mathscr{C}_t) U_t, \qquad X \in \mathscr{B}, \tag{6.2}$$

where $\mathscr{C}_t = \mathrm{vN}(Z_s : 0 \le s \le t)$. Note that $\mathscr{Y}_t = U_t^* \mathscr{C}_t U_t$ follows from the fact that $U_s^* Z_s U_s = U_t^* Z_s U_t$ for $t \ge s$, the property we used in section 5.2 to prove self-nondemolition of Y_t. The ease with which we will now be able to manipulate $\mathbb{Q}^t(X|\mathscr{C}_t)$ highlights the usefulness of the transformation (6.2).

Our strategy will be as follows. We wish to calculate $\mathbb{Q}^t(X|\mathscr{C}_t)$; however, the state $\mathbb{P}$ has the nice property that $Z_{s\le t}$, which generates $\mathscr{C}_t$, is a $\mathbb{P}$-Wiener process. We want to use the Bayes formula, Lemma 3.18, in order to express $\mathbb{Q}^t(X|\mathscr{C}_t)$ in terms of $\mathbb{P}$-conditional expectations. We run into a problem, however, as the "change of measure" operator U_t that relates $\mathbb{P}$ with $\mathbb{Q}^t$ does not satisfy the requirement of Lemma 3.18 that[12] $U_t \in \mathscr{C}_t'$. To solve this problem, we will replace U_t by a different operator V_t which is affiliated to $\mathscr{C}_t'$, but which still defines the same state in the sense that $\mathbb{P}(U_t^* X U_t) = \mathbb{P}(V_t^* X V_t)$ for every X. The following technique, to our knowledge, first appeared in [39]; it replaces Girsanov's theorem in the quantum context.

LEMMA 6.1. *Let V_t be the solution of the QSDE*

$$dV_t = \Big\{ L\,(dA_t^* + dA_t) - \frac{1}{2} L^* L\,dt - iH\,dt \Big\} V_t. \tag{6.3}$$

Then V_t is affiliated to $\mathscr{C}_t'$ and $\mathbb{Q}^t(X) = \mathbb{P}(V_t^* X V_t)$ for all $X \in \mathscr{B} \otimes \mathscr{W}$.

Proof. Let us assume for simplicity that the state ρ on $\mathscr{B}$ is pure; we can always obtain a mixed state later by taking convex combinations. Then $\mathbb{P}(X) = \langle \psi \otimes \Phi, X\,\psi \otimes \Phi \rangle$ for some vector $\psi \in \mathsf{h}$ (and $\Phi \in \mathsf{F}$ is the vacuum vector). To show that $\mathbb{P}(U_t^* X U_t) = \mathbb{P}(V_t^* X V_t)$, it is thus sufficient to show that

$$U_t\,\psi \otimes \Phi = V_t\,\psi \otimes \Phi. \tag{6.4}$$

Note that $\big\langle (U_t - V_t)\psi \otimes \Phi,\, (U_t - V_t)\psi \otimes \Phi \big\rangle = \big\langle \psi \otimes \Phi,\, (U_t - V_t)^*(U_t - V_t)\psi \otimes \Phi \big\rangle$. A simple application of the quantum Itô rule and the fact that vacuum expectations of stochastic integrals vanish show that $\big\langle \psi \otimes \Phi,\, (U_t - V_t)^*(U_t - V_t)\psi \otimes \Phi \big\rangle = 0$ and (6.4) holds. ∎

Note that the only difference between the equation for U_t, (5.2), and the equation for V_t, (6.3), is that we have modified the coefficient in front of dA_t. In principle, we could change the integrand of the A_t-integral arbitrarily without affecting how the QSDE acts on the vacuum; essentially this is due to the fact that any integral with respect to A_t vanishes when acting on the vacuum, as remarked after Remark 4.2.

[12] If this were the case, then we could calculate $Y_t = U_t^* Z_t U_t = Z_t U_t^* U_t = Z_t$; i.e., the observations would carry no information about the system and the filtering problem would be trivial.

In Lemma 6.1 we exploit this fact to modify U_t precisely so that it is in the commutant of $\mathscr{C}_t$; indeed, (6.3) is driven only by the noise $Z_t = A_t + A_t^*$ and its coefficients are in $\mathscr{B} \subset \mathscr{C}_t'$. We are now ready to apply the Bayes formula, Lemma 3.18. Together with Lemma 6.1 and (6.2), we immediately obtain the following result.

THEOREM 6.2 (noncommutative Kallianpur–Striebel). *Define for any system operator $X \in \mathscr{B}$ the unnormalized conditional expectation*

$$\sigma_t(X) = U_t^* \, \mathbb{P}(V_t^* X V_t | \mathscr{C}_t) \, U_t \in \mathscr{Y}_t. \tag{6.5}$$

Then the conditional expectation (5.6) *is given by*

$$\pi_t(X) = \frac{\sigma_t(X)}{\sigma_t(I)} \qquad \forall \, X \in \mathscr{B}. \tag{6.6}$$

We now obtain an explicit expression for $\sigma_t(X)$.

THEOREM 6.3 (unnormalized quantum filtering equation). *The unnormalized conditional expectation $\sigma_t(X)$ satisfies the following linear QSDE:*

$$d\sigma_t(X) = \sigma_t(\mathcal{L}_{L,H}(X)) \, dt + \sigma_t(L^* X + X L) \, dY_t. \tag{6.7}$$

To obtain (6.7) we will need to take conditional expectations of quantum Itô integrals. Let us briefly show how to do this. First, we claim that if K_t is an adapted process with K_s affiliated to $\mathscr{C}_s'$, then $\mathbb{P}(K_s|\mathscr{C}_t) = \mathbb{P}(K_s|\mathscr{C}_s)$ for $s \le t$. This follows from the fact that $\mathscr{C}_t = \mathscr{C}_s \otimes \mathscr{C}_{[s,t]}$ and that K_s is independent from $\mathscr{C}_{[s,t]}$ by adaptedness. Second, conditional expectations and integrals can be exchanged as follows:

$$\mathbb{P}\left(\int_0^t K_s \, ds \middle| \mathscr{C}_t \right) = \int_0^t \mathbb{P}(K_s|\mathscr{C}_s) \, ds, \qquad \mathbb{P}\left(\int_0^t K_s \, dZ_s \middle| \mathscr{C}_t \right) = \int_0^t \mathbb{P}(K_s|\mathscr{C}_s) \, dZ_s.$$

These properties are immediate if K_t is a simple process, and a proof of the general case is not difficult.

Proof. Using the quantum Itô rules we have

$$V_t^* X V_t = X + \int_0^t V_s^* \mathcal{L}_{L,H}(X) V_s \, ds + \int_0^t V_s^* (L^* X + X L) V_s \, d(A_s + A_s^*).$$

We next take conditional expectations of the terms in this expression; we obtain

$$\begin{aligned} \mathbb{P}(V_t^* X V_t | \mathscr{C}_t) = \mathbb{P}(X) &+ \int_0^t \mathbb{P}(V_s^* \mathcal{L}_{L,H}(X) V_s | \mathscr{C}_s) \, ds \\ &+ \int_0^t \mathbb{P}(V_s^* (L^* X + X L) V_s | \mathscr{C}_s) \, d(A_s + A_s^*). \end{aligned}$$

Another application of the quantum Itô rules now yields (6.7). ∎

By applying the Itô rules to the noncommutative Kallianpur–Striebel formula (6.6), we obtain an expression for the normalized conditional state

$$d\pi_t(X) = \pi_t(\mathcal{L}_{L,H}(X)) dt + \Big(\pi_t(L^* X + X L) - \pi_t(L^* + L) \, \pi_t(X) \Big) \Big(dY_t - \pi_t(L^* + L) \, dt \Big). \tag{6.8}$$

This (normalized) *quantum filtering equation* is a quantum analog of the classical Kushner–Stratonovich equation of nonlinear filtering. Note that this is a classical

stochastic differential equation by the spectral theorem: it is a recursive equation that is only driven by the (commutative) observations Y_t. Hence it can be implemented on a classical (digital) signal processor, as depicted in Figure 5.1.

Remark 6.4. Equation (6.8) is expressed in terms of the conditional state $\pi_t(X)$, where $X \in \mathscr{B}$. Now recall from section 2 that any state on a finite-dimensional Hilbert space can be expressed as $\mathrm{Tr}[\rho X]$ for some density matrix ρ. Similarly, if h (and hence $\mathscr{B}$) is finite dimensional, then we can always write $\pi_t(X) = \mathrm{Tr}[\rho_t X]$ where ρ_t, the conditional density matrix, is a (random) density matrix that is a function of the observations up to time t. From (6.8) we obtain explicitly

$$d\rho_t = -i[H,\rho_t]\,dt + (L\rho_t L^* - \tfrac{1}{2}L^*L\rho_t - \tfrac{1}{2}\rho_t L^*L)\,dt + (L\rho_t + \rho_t L^* - \mathrm{Tr}[(L+L^*)\rho_t]\rho_t)\,dW_t,$$

where $dW_t = dY_t - \mathrm{Tr}[(L+L^*)\rho_t]\,dt$. In section 7 we will see that W_t is a Wiener process. It is this representation that is usually found in the physics literature. ∎

6.2. Photon counting measurements. We now consider the photon counting setup given by (5.8) and (5.10). We would like to follow the same procedure as for homodyne detection. The following lemma, which replaces Lemma 6.1, suggests how to proceed. The proof is identical to that of Lemma 6.1.

LEMMA 6.5. *Let U'_t be the solution of the QSDE*

$$dU'_t = \Big\{L'\,dA_t^* - L'^*\,dA_t - \frac{1}{2}L'^*L'\,dt - iH'\,dt\Big\}U'_t$$

and let V'_t be the solution of

$$dV'_t = \Big\{L'(d\Lambda_t + dA_t^* + dA_t + dt) - \frac{1}{2}L'^*L'\,dt - L'\,dt - iH'\,dt\Big\}V'_t.$$

Then V'_t is affiliated to $\mathrm{vN}(\Lambda_s + A_s^* + A_s + s : s \le t)'$ *and* $\mathbb{P}(U_t'^* X U'_t) = \mathbb{P}(V_t'^* X V'_t)$.

Define $Z_t = \Lambda_t + A_t^* + A_t + t$ and $\mathscr{C}_t = \mathrm{vN}(Z_s : 0 \le s \le t)$. Lemma 6.5 directly provides us with a nondemolition change of measure, provided that we rotate our problem so that $\mathscr{Y}_t = U_t'^*\mathscr{C}_t U'_t$ using a suitable unitary operator U'_t. Then, defining $\sigma_t(X) = U_t'^*\,\mathbb{P}(V_t'^* X V'_t|\mathscr{C}_t)\,U'_t$, the Kallianpur–Striebel formula holds for $\sigma_t(X)$.

Define R_t as the solution of the QSDE

$$dR_t = (dA_t - dA_t^* - \tfrac{1}{2}dt)\,R_t.$$

Recall Example 4.5; evidently R_t is a Weyl operator, and in particular $\Lambda_t = R_t^* Z_t R_t$. But recall that $Y_t = U_t^*\Lambda_t U_t = U_t^* R_t^* Z_t R_t U_t$; thus $U'_t = R_t U_t$ is our rotation of choice. Using the quantum Itô rules we obtain

$$dU'_t = \Big\{(L-I)\,dA_t^* - (L^*-I)\,dA_t - \frac{1}{2}(L^*L + I - 2L + 2iH)\,dt\Big\}U'_t,$$

which corresponds to the nondemolition change of measure

$$dV'_t = \Big\{(L-I)\,dZ_t - \frac{1}{2}(L^*L - I + 2iH)\,dt\Big\}V'_t.$$

For $X \in \mathscr{B}$, using the quantum Itô rules we obtain

$$dV_t'^* X V'_t = V_t'^*\big(\mathcal{L}_{L,H}(X)\big)V'_t\,dt + V_t'^*(L^*XL - X)V'_t\,(dZ_t - dt).$$

Finally, using the definition of σ_t and the quantum Itô rules we obtain

$$d\sigma_t(X) = \sigma_t(\mathcal{L}_{L,H}(X))\,dt + \big(\sigma_t(L^*XL) - \sigma_t(X)\big)\big(dY_t - dt\big),$$

which is the unnormalized quantum filtering equation for counting observations.

Using the Kallianpur–Striebel formula $\pi_t(X) = \sigma_t(X)\,/\,\sigma_t(I)$ we can now obtain an expression for the normalized conditional state

$$d\pi_t(X) = \pi_t(\mathcal{L}_{L,H}(X))\,dt + \left(\frac{\pi_t(L^*XL)}{\pi_t(L^*L)} - \pi_t(X)\right)\big(dY_t - \pi_t(L^*L)\,dt\big),$$

which is the normalized quantum filtering equation for photon counting.

7. The innovations method. In this section we rederive the filtering equation for homodyne detection, (6.8), using martingale methods that are analogous to the classical case [13, 17]. We follow the classical treatment as in [33], [31, Chapter 18], and [65, Chapter 7]. Martingale methods have enjoyed wide and successful application in classical stochastic theory. The procedure is less straightforward than the reference probability method, however, and some familiarity with classical filtering theory would be helpful (see, e.g., [26] for an excellent introduction).

Let $\xi_t, \beta_t, \lambda_t, \mu_t$ be adapted processes affiliated to $\mathscr{Y}_t'$, where

$$\xi_t = \xi_0 + \int_0^t \beta_s\,ds + m_t = \xi_0 + \int_0^t \beta_s\,ds + \int_0^t (\lambda_s\,dA_s + \mu_s\,dA_s^*). \tag{7.1}$$

The measurement process Y_t is given by (5.9), and in what follows we write $h_t = j_t(L + L^*)$ and $Z_t = A_t + A_t^*$. Note that the conditional expectation $\hat{\xi}_t = \mathbb{P}(\xi_t|\mathscr{Y}_t)$ is well defined, and similarly for the coefficients β_t, λ_t, and μ_t.

The main filtering result for a process of the form (7.1) is the following.

THEOREM 7.1 (noncommutative Fujisaki–Kallianpur–Kunita). *Under the above assumptions, the filtered process $\hat{\xi}_t$ satisfies the QSDE*

$$d\hat{\xi}_t = \hat{\beta}_t\,dt + (\hat{\lambda}_t + \widehat{\xi_t h_t} - \hat{\xi}_t\hat{h}_t)\,dW_t, \tag{7.2}$$

where $\hat{r}_t \equiv \mathbb{P}(r_t|\mathscr{Y}_t)$ for any r_t affiliated to $\mathscr{Y}_t'$, and $dW_t = dY_t - \hat{h}_t\,dt$ defines the $\mathscr{Y}_t$-Wiener process (with respect to $\mathbb{P}$) W_t, called the innovations process.

The filtering expression (7.2) is formally identical to the classical case [31, Theorem 18.11] and [65, Proposition 3.2]. Before we prove Theorem 7.1, we will show how to obtain the quantum filtering equation (6.8) using this result.

COROLLARY 7.2. *The conditional state $\pi_t(X)$ is given by* (6.8).

Proof. We set $\lambda_t = -j_t([X, L^*])$, $\mu_t = j_t([X, L])$, $\beta_t = j_t(\mathcal{L}_{L,H}(X))$, and $\xi_t = j_t(X)$. Then $\widehat{\xi_t h_t} = \pi_t(X(L + L^*))$, $\hat{\xi}_t\hat{h}_t = \pi_t(X)\pi_t(L + L^*)$, $\hat{\lambda}_t = -\pi_t([X, L^*])$, and $\hat{\beta}_t = \pi_t(\mathcal{L}_{L,H}(X))$. Hence using (7.2), (6.8) follows. ∎

Proof of Theorem 7.1. *Step* 1. We first show that the process

$$M_t = \hat{\xi}_t - \hat{\xi}_0 - \int_0^t \hat{\beta}_s\,ds$$

is a $\mathscr{Y}_t$-martingale, i.e., $\mathbb{P}(M_t|\mathscr{Y}_s) = M_s$ for all $s \le t$. This property is equivalent to $\mathbb{P}((M_t - M_s)K) = 0$ for all $K \in \mathscr{Y}_s$, or equivalently

$$\mathbb{P}\left[\left(\hat{\xi}_t - \hat{\xi}_s - \int_s^t \hat{\beta}_r\,dr\right)K\right] = \mathbb{P}\left[\left(\xi_t - \xi_s - \int_s^t \beta_r\,dr\right)K\right] = \mathbb{P}[(m_t - m_s)K] = 0$$

for all $K \in \mathscr{Y}_s$, where we have used Definition 3.13 in the first step. But as $K \in \mathscr{Y}_s \subset \mathscr{B} \otimes \mathscr{W}_{s]}$,

$$\mathbb{P}[(m_t - m_s)K] = \mathbb{P}\left[K \int_s^t (\lambda_r \, dA_r + \mu_r \, dA_r^*)\right] = \mathbb{P}\left[\int_s^t (K\lambda_r \, dA_r + K\mu_r \, dA_r^*)\right] = 0,$$

where we have used that the vacuum expectation of quantum Itô integrals vanishes. Thus we have demonstrated that M_t is a $\mathscr{Y}_t$-martingale.

Step 2. We now show that W_t is a Wiener process under $\mathbb{P}$. We begin by verifying that the innovations process

$$W_t = Y_t - \int_0^t \hat{h}_s \, ds \tag{7.3}$$

is a $\mathscr{Y}_t$-martingale. We need to show that $\mathbb{P}[(W_t - W_s)K] = 0$ for any $s \le t$ and $K \in \mathscr{Y}_s$. This is equivalent to

$$\mathbb{P}\left[\left(Y_t - Y_s - \int_s^t \hat{h}_r \, dr\right) K\right] = \mathbb{P}\left[\left(Y_t - Y_s - \int_s^t h_r \, dr\right) K\right] = 0$$

for all $K \in \mathscr{Y}_s$, where the second expression follows from the definition of the conditional expectation. But from (5.9) we obtain

$$\mathbb{P}\left[\left(Y_t - Y_s - \int_s^t h_r \, dr\right) K\right] = \mathbb{P}[(Z_t - Z_s)K] = 0$$

as $K \in \mathscr{Y}_s \subset \mathscr{B} \otimes \mathscr{W}_{s]}$, $(Z_t - Z_s) \in \mathscr{W}_{[s,t]}$, and hence $\mathbb{P}[(Z_t - Z_s)K] = \mathbb{P}(K)\,\mathbb{P}(Z_t - Z_s) = 0$. Thus W_t is a $\mathscr{Y}_t$-martingale.

From (7.3) we read off the Itô rule $dW_t^2 = dt$; classically, a process that obeys this property and is a martingale must be a Wiener process by Lévy's theorem (e.g., [31, Lemma 18.7]). But we can simply apply the classical result, as W_t is a commutative process (note that $\hat{h}_t \in \mathscr{Y}_t$ for $s \le t$ by construction) and is hence equivalent to the corresponding classical process obtained through the spectral theorem.

Now that we have shown that W_t is a Wiener process, we can try to represent the martingale M_t as a stochastic integral with respect to W_t. As usual in filtering theory the ordinary martingale representation theorem does not suffice for this purpose, but the representation theorem of Fujisaki–Kallianpur–Kunita (e.g., [49, Theorem 5.20]) allows us to conclude nonetheless that

$$M_t = \int_0^t \gamma_s \, dW_s \quad \Longrightarrow \quad \hat{\xi}_t = \hat{\xi}_0 + \int_0^t \hat{\beta}_s \, ds + \int_0^t \gamma_s \, dW_s \tag{7.4}$$

for some adapted process $\gamma_t \in \mathscr{Y}_t$.

Step 3. We next obtain a first expression for $\widehat{\xi_t Y_t}$:

$$\widehat{\xi_t Y_t} = \int_0^t [\widehat{\beta_s Y_s} + \widehat{\xi_s h_s} + \hat{\lambda}_s] ds + M_1(t), \tag{7.5}$$

where $M_1(t)$ is a $\mathscr{Y}_t$-martingale. As before, it suffices to show that

$$\mathbb{P}((M_1(t) - M_1(s))K) = \mathbb{P}\left[\left(\xi_t Y_t - \xi_s Y_s - \int_s^t [\beta_s Y_s + \xi_s h_s + \lambda_s] ds\right) K\right] = 0$$

for all $K \in \mathscr{Y}_s$, where we have used the definition of the conditional expectation. But

$$\begin{aligned} d(\xi_t Y_t) &= (d\xi_t)Y_t + \xi_t dY_t + d\xi_t dY_t \\ &= (\beta_t dt + dm_t)Y_t + \xi_t(h dt + dZ_t) + dm_t dZ_t \\ &= (\beta_t Y_t + \xi_t h_t + \lambda_t)dt + (Y_t\lambda_t + \xi_t)dA_t + (Y_t\mu_t + \xi_t)dA_t^*. \end{aligned}$$

Hence exactly as before, it follows that $M_1(t)$ is a $\mathscr{Y}_t$-martingale.

Step 4. Next, we derive a second expression for $\widehat{\xi_t Y_t}$:

$$\widehat{\xi_t Y_t} = \int_0^t [\hat\beta_s Y_s + \hat\xi_s \hat h_s + \gamma_s]ds + M_2(t), \tag{7.6}$$

where $M_2(t)$ is a $\mathscr{Y}_t$-martingale. To show this, note that $\widehat{\xi_t Y_t} = \hat\xi_t Y_t$. By Itô's rules,

$$\begin{aligned} d(\hat\xi_t Y_t) &= (d\hat\xi_t)Y_t + \hat\xi_t dY_t + d\hat\xi_t dY_t \\ &= (\hat\beta_t dt + \gamma_t dW_t)Y_t + \hat\xi_t(\hat h_t dt + dW_t) + \gamma_t dW_t dW_t \\ &= (\hat\beta_t Y_t + \hat\xi_t\hat h_t + \gamma_t)dt + (\gamma_t Y_t + \hat\xi_t)dW_t \end{aligned}$$

which establishes (7.6).

Step 5. We can now identify γ_t. From (7.5) and (7.6) we have two representations for $\widehat{\xi_t Y_t}$. By uniqueness, it follows that the finite variation terms are equal, namely,

$$\widehat{\beta_s Y_s} + \widehat{\xi_s h_s} + \hat\lambda_s = \hat\beta_s Y_s + \hat\xi_s \hat h_s + \gamma_s.$$

Therefore $\gamma_s = \widehat{\xi_s h_s} + \hat\lambda_s - \hat\xi_s\hat h_s$ as required. □

8. Conclusion. In this article we have provided an introduction to quantum filtering. Our goal has been to emphasize the mathematical structures of quantum probability and to show their use in system-probe models from quantum optics. We have seen that the techniques employed in quantum filtering theory closely mirror their analogs in the classical theory of nonlinear filtering. As in the classical theory, an important role is played by the conditional expectation as the mean least square estimate of the system given the observations thus far.

The spectral theorem provides a one-to-one correspondence between commutative von Neumann algebras equipped with normal states and classical (Kolmogorov) probability spaces. This enabled us to represent commuting observables (self-adjoint operators acting on a Hilbert space) as random variables on a single classical probability space. For an observable X that commutes with all members of a commutative family of observables $\mathscr{Y}$, we can define the conditional expectation of X onto $\mathscr{Y}$ by pulling back the classical conditional expectation of X onto $\mathscr{Y}$, both represented on a classical probability space via the spectral theorem. For this procedure to work, it is crucial that the family $\mathscr{Y}$ is commutative, the *self-nondemolition* property, and that X commutes with $\mathscr{Y}$, the *nondemolition* property.

As a model for the quantum electromagnetic field we introduced the algebra $\mathscr{W}$ of bounded operators on the Boson Fock space equipped with the vacuum state ϕ. We studied families of commuting operators affiliated to $\mathscr{W}$, and by representing these commuting operators on a classical probability space, we found that $(\mathscr{W}, \phi)$ contains families of Wiener and Poisson processes, all of which can be written as linear combinations of the so-called fundamental noises Λ_t (gauge process), A_t^* (creation process), and A_t (annihilation process). These families of operators do not necessarily

commute with each other, and therefore the different Wiener and Poisson processes in $(\mathscr{W}, \phi)$ cannot be represented on the same classical probability space. Physically, these processes can all be observed with a suitable measurement setup (e.g., homodyne detection or photon counting); however, these observations cannot be made in a single realization of the experiment, as the different families do not commute.

Although the noncommuting processes affiliated to $(\mathscr{W}, \phi)$ cannot be represented on the same probability space, it is still possible to capture them within a single stochastic calculus. Stochastic integrals with respect to the fundamental noises can be defined as operators affiliated with $\mathscr{W}$, and a quantum Itô rule (integration by parts rule) based on a quantum Itô table for the fundamental noises can be shown to hold [42]. The interaction of some fixed system, e.g., a cloud of atoms, with the electromagnetic field in a Markov limit [1, 36] is given by a unitary U_t that satisfies a quantum stochastic differential equation, i.e., a stochastic differential equation given in terms of quantum stochastic integrals with respect to the fundamental noises. Note that the equation for U_t can therefore be driven by different noises that do not necessarily commute with each other.

Given the unitary U_t, the Heisenberg evolution of the observables of the system (e.g., a cloud of atoms) is given by $j_t(X) = U_t^* X U_t$. Instead of directly observing the system at time t, we only had access to field observables up to time t, e.g., $Y_s = U_s^*(A_s + A_s^*)U_s,\ 0 \le s \le t$ (homodyne detection) or $Y_s = U_s^* \Lambda_s U_s,\ 0 \le s \le t$ (photon counting). We showed that these system-observation pairs satisfy the nondemolition requirements that are necessary for the existence of the conditional expectation $\mathbb{P}(j_t(X)|\mathscr{Y}_t)$ of a system observable $j_t(X)$ at time t onto the observations thus far. The quantum filtering equation recursively propagates the conditional expectation, our best estimate of system observables, in time. In close analogy with the classical case, we derived the quantum filtering equation in two ways, once using a change of measure technique and once using martingales and martingale representation.

There are many points we did not touch upon in this introduction to quantum filtering. Some noteworthy omissions are the linear theory [30] and models in discrete time with discrete observables [19]. Another notable omission is how the filtering equations can be used when controlling a quantum system [10, 11, 28, 18]. As in the classical case, a separation theorem can be shown to hold [18]. This means that the optimal controller will depend on the observation history only through the filter. This separates the control problem into an estimation step (filtering) and a control step based on the estimates only.

REFERENCES

[1] L. Accardi, A. Frigerio, and Y. Lu, *The weak coupling limit as a quantum functional central limit*, Comm. Math. Phys., 131 (1990), pp. 537–570.

[2] L. Accardi, J. Gough, and Y. Lu, *On the stochastic limit for quantum theory*, Rep. Math. Phys., 36 (1995), pp. 155–187.

[3] M. A. Armen, J. K. Au, J. K. Stockton, A. C. Doherty, and H. Mabuchi, *Adaptive homodyne measurement of optical phase*, Phys. Rev. Lett., 89 (2002), 133602.

[4] E. Arthurs and J. L. Kelly, *On the simultaneous measurement of a pair of conjugate observables*, Bell Syst. Tech. J., 44 (1965), pp. 725–729.

[5] S. Attal and J. M. Lindsay, *Quantum stochastic calculus with maximal operator domains*, Ann. Probab., 32 (2004), pp. 488–529.

[6] A. Barchielli, *Continual measurements in quantum mechanics and quantum stochastic calculus*, in Open Quantum Systems III: Recent Developments, S. Attal, A. Joye, and C.-A. Pillet, eds., Springer, Berlin, Heidelberg, 2006, pp. 207–292.

[7] A. Barchielli, *Direct and heterodyne detection and other applications of quantum stochastic calculus to quantum optics*, Quantum Opt., 2 (1990), pp. 423–441.
[8] C. Barnett, R. F. Streater, and I. F. Wilde, *Quasi-free quantum stochastic integrals for the CAR and CCR*, J. Funct. Anal., 52 (1983), pp. 19–47.
[9] V. P. Belavkin, *Quantum filtering of Markov signals with white quantum noise*, Radiotechnika i Electronika, 25 (1980), pp. 1445–1453.
[10] V. P. Belavkin, *Theory of the control of observable quantum systems*, Autom. Rem. Control, 44 (1983), pp. 178–188.
[11] V. P. Belavkin, *Nondemolition measurement and control in quantum dynamical systems*, in Information Complexity and Control in Quantum Physics, CISM Courses and Lectures 294, S. Diner and G. Lochak, eds., Springer-Verlag, Vienna, 1987, pp. 331–336.
[12] V. P. Belavkin, *Quantum continual measurements and a posteriori collapse on CCR*, Comm. Math. Phys., 146 (1992), pp. 611–635.
[13] V. P. Belavkin, *Quantum stochastic calculus and quantum nonlinear filtering*, J. Multivariate Anal., 42 (1992), pp. 171–201.
[14] P. Biane, *Calcul stochastique noncommutatif*, in Lectures on Probability Theory (Saint-Flour, 1993), Lecture Notes in Math. 1608, P. Bernard, ed., Springer-Verlag, Berlin, 1995, pp. 1–96.
[15] L. M. Bouten, *Filtering and Control in Quantum Optics*, Ph.D. thesis, Radboud Universiteit Nijmegen, 2004.
[16] L. M. Bouten, S. C. Edwards, and V. P. Belavkin, *Bellman equations for optimal feedback control of qubit states*, J. Phys. B, At. Mol. Opt. Phys., 38 (2005), pp. 151–160.
[17] L. M. Bouten, M. I. Guţă, and H. Maassen, *Stochastic Schrödinger equations*, J. Phys. A, 37 (2004), pp. 3189–3209.
[18] L. M. Bouten and R. Van Handel, *On the separation principle of quantum control,* in Proceedings of the 2006 QPIC Symposium (Nottingham, UK), M. Guta, ed., World Scientific, Singapore, to appear.
[19] L. M. Bouten, R. Van Handel, and M. R. James, *A discrete invitation to quantum filtering and feedback control*, SIAM Rev., to appear.
[20] O. Bratteli and D. Robinson, *Operator Algebras and Quantum Statistical Mechanics* 1, 2nd ed., Springer-Verlag, Berlin, Heidelberg, 1987.
[21] P. Bushev, D. Rotter, A. Wilson, F. Dubin, C. Becher, J. Eschner, R. Blatt, V. Steixner, P. Rabl, and P. Zoller, *Feedback cooling of a single trapped ion*, Phys. Rev. Lett., 96 (2006), p. 043003.
[22] H. J. Carmichael, *An Open Systems Approach to Quantum Optics*, Springer-Verlag, Berlin, Heidelberg, New York, 1993.
[23] C. Cohen Tannoudji, J. Dupont Roc, and G. Grynberg, *Photons and Atoms: Introduction to Quantum Electrodynamics*, Wiley, New York, 1989.
[24] E. B. Davies, *Quantum stochastic processes*, Comm. Math. Phys., 15 (1969), pp. 277–304.
[25] E. B. Davies, *Quantum Theory of Open Systems*, Academic Press, London, New York, San Francisco, 1976.
[26] M. H. A. Davis and S. I. Marcus, *An introduction to nonlinear filtering*, in Stochastic Systems: The Mathematics of Filtering and Identification and Applications, M. Hazewinkel and J. C. Willems, eds., D. Reidel, Dordrecht, 1981, pp. 53–75.
[27] A. C. Doherty, S. Habib, K. Jacobs, H. Mabuchi, and S. M. Tan, *Quantum feedback and classical control theory*, Phys. Rev. A, 62 (2000), 012105.
[28] A. C. Doherty and K. Jacobs, *Feedback-control of quantum systems using continuous state-estimation*, Phys. Rev. A, 60 (1999), pp. 2700–2711.
[29] T. Duncan, *Evaluation of likelihood functions*, Inform. and Control, (1968), pp. 62–74.
[30] S. C. Edwards and V. P. Belavkin, *Optimal Quantum Feedback Control via Quantum Dynamic Programming*, preprint quant-ph/0506018v2, University of Nottingham, 2005.
[31] R. J. Elliott, *Stochastic Calculus and Applications*, Springer-Verlag, New York, 1982.
[32] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Springer-Verlag, New York, 1995.
[33] M. Fujisaki, G. Kallianpur, and H. Kunita, *Stochastic differential equations for the nonlinear filtering problem*, Osaka J. Math., 1 (1972), pp. 19–40.
[34] C. W. Gardiner and M. J. Collett, *Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation*, Phys. Rev. A, 31 (1985), pp. 3761–3774.
[35] J. M. Geremia, J. K. Stockton, A. C. Doherty, and H. Mabuchi, *Quantum Kalman filtering and the Heisenberg limit in atomic magnetometry*, Phys. Rev. Lett., 91 (2003), 250801.

[36] J. Gough, *Quantum flows as Markovian limit of emission, absorption and scattering interactions*, Comm. Math. Phys., 254 (2005), pp. 489–512.
[37] J. Gough, V. P. Belavkin, and O. G. Smolyanov, *Hamilton-Jacobi-Bellman equations for quantum filtering and control*, J. Opt. B Quantum Semiclass. Opt., 7 (2005), pp. S237–S244.
[38] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, North–Holland, Amsterdam, New York, Oxford, 1982.
[39] A. Holevo, *Quantum stochastic calculus*, J. Soviet Math., 56 (1991), pp. 2609–2624 (in English); Itogi Nauki i Tekhniki, Ser. Sovr. Prob. Mat. 36 (1990), pp. 3–28 (in Russian).
[40] A. Holevo, *Statistical Structure of Quantum Theory*, Springer-Verlag, Berlin, 2001.
[41] R. L. Hudson, *An introduction to quantum stochastic calculus and some of its applications*, in Quantum Probability Communications, vol. XI, S. Attal and J. Lindsay, eds., World Scientific, Singapore, 2003, pp. 221–271.
[42] R. L. Hudson and K. R. Parthasarathy, *Quantum Itô's formula and stochastic evolutions*, Comm. Math. Phys., 93 (1984), pp. 301–323.
[43] M. R. James, *Risk-sensitive optimal control of quantum systems*, Phys. Rev. A, 69 (2004), 032108.
[44] M. R. James, *A quantum Langevin formulation of risk-sensitive optimal control*, J. Opt. B Quantum Semiclass. Opt., 7 (2005), pp. S198–S207.
[45] R. V. Kadison and J. R. Ringrose, *Fundamentals of the Theory of Operator Algebras*, Vol. I, Academic Press, San Diego, 1983.
[46] R. V. Kadison and J. R. Ringrose, *Fundamentals of the Theory of Operator Algebras*, Vol. II, Academic Press, San Diego, 1986.
[47] B. Kümmerer, *Markov dilations on W^*-algebras*, J. Funct. Anal., 63 (1985), pp. 139–177.
[48] G. Lindblad, *On the generators of quantum dynamical semigroups*, Comm. Math. Phys., 48 (1976), pp. 119–130.
[49] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes* I: *General Theory*, Springer-Verlag, Berlin, 2001.
[50] H. Maassen, *Quantum Markov processes on Fock space described by integral kernels*, in QP and Applications II, Lecture Notes in Math. 1136, L. Accardi and W. von Waldenfels, eds., Springer-Verlag, Berlin, 1985, pp. 361–374.
[51] E. Merzbacher, *Quantum Mechanics*, 3rd ed., Wiley, New York, 1998.
[52] P.-A. Meyer, *Quantum Probability for Probabilists*, Springer-Verlag, Berlin, 1993.
[53] R. Mortensen, *Optimal Control of Continuous-Time Stochastic Systems*, Ph.D. thesis, University of California, Berkeley, 1966.
[54] E. Nelson, *Notes on non-commutative integration*, J. Funct. Anal., 15 (1974), pp. 103–116.
[55] K. R. Parthasarathy, *An Introduction to Quantum Stochastic Calculus*, Birkhäuser, Basel, 1992.
[56] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol.* 1: *Functional Analysis,* Academic Press, San Diego, London, 1980.
[57] S. Sakai, *C^*-algebras and W^*-algebras*, Springer-Verlag, Berlin, 1998.
[58] J. K. Stockton, *Continuous Quantum Measurement of Cold Alkali-Atom Spins*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2006.
[59] M. Takesaki, *Conditional expectations in von Neumann algebras*, J. Funct. Anal., 9 (1971), pp. 306–321.
[60] R. Van Handel and H. Mabuchi, *Quantum projection filter for a highly nonlinear model in cavity QED*, J. Opt. B Quantum Semiclass. Opt., 7 (2005), pp. S226–S236.
[61] R. Van Handel, J. K. Stockton, and H. Mabuchi, *Feedback control of quantum state reduction*, IEEE Trans. Automat. Control, 50 (2005), pp. 768–780.
[62] R. Van Handel, J. K. Stockton, and H. Mabuchi, *Modelling and feedback control design for quantum state preparation*, J. Opt. B Quantum Semiclass. Opt., 7 (2005), pp. S179–S197.
[63] D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, UK, 1991.
[64] H. Wiseman, *Quantum theory of continuous feedback*, Phys. Rev. A, 49 (1994), pp. 2133–2150.
[65] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, Springer-Verlag, New York, 1985.
[66] M. Zakai, *On the optimal filtering of diffusion processes*, Z. Wahrsch. Verw. Geb., 11 (1969), pp. 230–243.