FILTERING AND ESTIMATION, NONLINEAR

To have a historical perspective of the advent of nonlinear filtering and estimation, initially the development of linear filtering and estimation is described. The first studies of linear filtering or linear estimation for stochastic processes were made by Kolmogorov (1,2), Krein (3,4) and Wiener (5). The research of Kolmogorov and Krein and the research of Wiener were done independently. Kolmogorov, who was motivated by Wold (6), gave a solution to the prediction problem for discrete-time stochastic processes. Since Kolmogorov and Krein were not motivated for their work by any specific applications, the formulae for the optimum predictor did not play a special role. However, Wiener was motivated for his work during World War II by the analysis of anti-aircraft fire-control problems from ships. He solved the continuous-time linear prediction problem and derived an explicit formula for the optimum predictor. He also solved the filtering problem of estimating a stochastic signal process that is corrupted by an additive noise process. In this latter case Wiener expressed the solution in terms of an integral equation, the Wiener–Hopf equation (7). Wiener had obtained this equation in his work on potential theory a number of years earlier. This relation alludes to the probabilistic interpretation of potential theory using Brownian motion (8). Wiener's book (5) contains a number of elementary, explicitly solvable examples.

The sum of the signal process and the additive noise process is called the *observation process.* The prediction problem is to estimate the signal process at some future time usually on an ongoing basis from the observation process. The filtering problem is to estimate the signal process at the same time. The smoothing problem has a couple of variants: (1) Given the observation process in a fixed time interval, estimate the signal process at each element in the time interval, and (2) estimate the signal process at a time that is a fixed lag behind the observation process. The approach of Kolmogorov, Krein, and Wiener to these problems assumed that the stochastic processes are (wide-sense) stationary and that the infinite past of the observation process is available. Both of these assumptions are not physically reasonable, so there was a need to relax these assumptions.

In the late 1950s, control and system theory were undergoing a significant change from the frequency-domain approach to the state-space approach. Transfer function descriptions of linear systems were replaced by ordinary differential equation descriptions of linear systems. This state-space approach provided an impetus to reexamine the linear filtering problem. Using this approach the signal process is modeled as the solution of a linear differential equation with a Gaussian white noise input so the signal process is Gauss–Markov. The differential of the observations process is a linear transformation of the signal process plus Gaussian white noise. This filtering model does not require the infinite past of the observations. The signal and the observation processes can evolve from some fixed time with a Gaussian random variable as the initial condition for the differential equation that describes the signal process. The processes are not required to be stationary; and, in fact, the coefficients of the differential equation for the signal process and the linear transformation for the signal in the observation equation can be time-varying. While it is not required that the ordinary differential equation for the signal process be stable, which is implicit in the description for stationary processes, it is necessary to be able to model the signal process as the solution of an ordinary differential equation with a white noise input. In general a stationary Gaussian process may not have such a model.

With the success of these linear filtering results that were developed particularly by Kalman (9) for discrete-time processes and by Kalman-Bucy (10) for continuous-time processes, an interest developed in trying to solve a filtering problem where the signal is a solution to a nonlinear differential equation with a white noise input. It is natural to call such a problem *a nonlinear filtering problem.* The precise description of such a problem required the introduction of a significant amount of the modern theory of stochastic processes. The major technique for describing the signal process is the theory of stochastic differential equations that was initiated by K. Itô (11).

The Gaussian white noise processes that appear as inputs in the nonlinear differential equations require more sophisticated mathematical methods than do the inputs to linear differential equations. This occurs because the linear transformations of white noise have one natural interpretation but the nonlinear transformations of white noise have no single natural interpretation.

Interestingly, it was Wiener (12) who first constructed the basic sample path property of the integral of Gaussian white noise that is called the *Wiener process* or *Brownian motion* and which provided the basis for the interpretations of nonlinear transformations of white noise. Many important properties of Brownian motion were determined by P. Lévy (13). The solution of a stochastic differential equation (a nonlinear differential equation with a white noise input) required the theory of stochastic integrals (14), which depends on some martingale theory (15) associated with Brownian motion.

No one definition of stochastic integrals arises naturally from the Riemann sum approximations to the stochastic integral. This phenomenon occurs because Brownian motion does not have bounded variation. The definition of K. Itô (14) is the most satisfying probabilistically because it preserves the martingale property of Brownian motion and Wiener integrals (stochastic integrals with deterministic integrands). However, the calculus associated with the Itô definition of stochastic integral is somewhat unusual. The Fisk–Stratonovich definition of stochastic integral (16,17) preserves the usual calculus properties, but the family of integrable functions is significantly smaller. An uncountable family of distinct definitions of stochastic integrals can be easily exhibited (18). This choice or ambiguity in the definition of a stochastic integral has played an important role in nonlinear filtering because initially some nonlinear filtering solutions were given without specifying the interpretation or the definition of the stochastic integrals. This ambiguity often arose by a formal passage to the limit from discrete time.

In general, to compute conditional statistics of the state process given the observation process, it is necessary to compute the conditional density of the state process given the observation process. For linear filtering the signal and the observation processes are Gaussian, so the conditional density is determined by the conditional mean and the conditional covariance. The conditional covariance is not random, so it does not depend on the observation process. The conditional mean can be shown to satisfy a stochastic differential equation that models the signal process and that has the observations as the input. These two conditional statistics (i.e., function of the observations are called *sufficient conditional statistics* (19) because the conditional density can be recovered from them. For nonlinear filtering the solution does not simplify so easily. In general there is no finite family of sufficient conditional statistics for a nonlinear filtering problem.

The conditional density can be shown to satisfy a nonlinear stochastic partial differential equation (20,21). This equation is especially difficult to solve because it is a stochastic partial differential equaton and it is nonlinear. Even approximations are difficult to obtain. The conditional density can be expressed using Bayes formula (22,23), so that it has the same form as the Bayes formula in elementary probability though it requires function space integrals. The numerator in the Bayes formula expression for the conditional density is called the *unnormalized conditional density*. This unnormalized conditional density satisfies a stochastic partial differential equation that is linear. It is called the Duncan–Mortensen–Zakai (DMZ) equation of nonlinear filtering (24–26).

In nonlinear filtering, the question of finite dimensional filters describes the problem of finding finite dimensional solutions to the DMZ equation or to finite families of conditional statistics. A basic approach to this question on the existence or the nonexistence of finite-dimensional filters is the estimation algebra (27,28,28a), which is a Lie algebra of differential operators that is generated by the differential operators in the DMZ equation. Some families of nonlinear filtering problems have been given that exhibit finite-dimensional filters (e.g., see Ref. 29).

Some methods from algebraic geometry have been used to give necessary and sufficient conditions on the coefficients of the stochastic equations for small state-space dimension, so that the nonlinear filtering problem has finite-dimensional filters.

Since the results for the existence of finite-dimensional filters for nonlinear filtering problems are generally negative, many approximation methods have been developed for the numerical solution of the DMZ equation or the equations for some associated conditional statistics. In their study of Wiener space (the Banach space of continuous functions with the uniform topology and the Wiener measure), Cameron and Martin (30) showed that any square integrable functional on Wiener space could be represented as an infinite series of products of Hermite polynomials (Wick polynomials). K. Itô (31) refined this expression by using an infinite series of multiple Wiener integrals. The relation between these two representations is associated with including or excluding the diagonal in multiple integration. This relation carries over to Stratonovich integrals, and the explicit relation between these two stochastic integrals in this case is given by the Hu–Meyer formula (32).

For the solution of the linear filtering problem it was well known from the early work that the observation process in the optimal filter appears with a linear transformation of the estimate as a difference and that this difference is a process that is white with respect to the observation process. Since a family of square integrable zero mean random variables generates a vector space with an inner product that is the expectation of a product of two of the random variables, the (linear) filtering problem can be posed as a projection problem in a vector space (Hilbert space). The occurrence of a process that is white with respect to the observations is natural from a Gram–Schmidt orthogonalization procedure and projections. This process has been historically called the *innovation process* (6). For Wiener filtering, this innovations approach was introduced in the engineering literature by Bode and Shannon (6a). For linear filtering it is straightforward to verify that the observation process and the innovation process are "equivalent" (that is, there is a bijection between them) by showing that a linear operator is invertible.

For nonlinear filtering there is still an innovation process. It is more subtle to verify that the observation process and the innovation process are equivalent (33). Thus the nonlinear filtering solution has a vector space interpretation via orthogonalization and projections as for the linear filtering solution. However, this is not surprising because in both cases there is a family of (square integrable) random variables and the conditional expectation is a projection operator. This occurrence of the innovation process can be obtained by an absolute continuity of measures (34). In information theory, the mutual information for a signal and a signal plus noise can be computed similarly (35).

The expression for the conditional probability or the conditional density, given the past of the observations as a ratio of expectations, has a natural interpretation as a Bayes formula (22,23) that naturally generalizes the well-known Bayes formula of elementary probability.

The stochastic partial differential equations for the conditional probability density or the unnormalized conditional probability density are obtained by the change of variables formula of K. Itô (36).

The fact that the sample paths of Brownian motion (the formal integral of Gaussian white noise) do not have bounded variation has important implications concerning "robustness" questions. Wong and Zakai (37) showed that if Brownian motion in a stochastic differential equation is replaced by a sequence of piecewise smooth processes that converge (uniformly) to Brownian motion, then the corresponding sequence of solutions of the ordinary differential equations obtained from the stochastic differential equation by replacing the Brownian motion by the piecewise smooth processes do not converge to the solution of the stochastic differential equation for many nonlinear stochastic differential equations. This result of Wong and Zakai has important implications for nonlinear filtering. If nonlinear filters are constructed from time discretizations of the processes and a formal passage to the limit is made, then it may not be clear about the interpretation of the resulting solution. This question is closely related to the choice of the definition of stochastic integrals and the unusual calculus that is associated with K. Itô's definition. In the early development of nonlinear filtering theory, solutions were given that did not address the question of the definition of stochastic integrals. Generally speaking, formal passages to the limit from time discretizations require the Stratonovich definition of stochastic integrals because these integrals satisfy the usual properties of calculus.

One classical example of the use of nonlinear filtering in communication theory is the analysis of the phase-lock loop problem. This problem arises in the extraction of the signal in frequency modulation (FM) transmission. The process that is received by the demodulator is a sum of a frequency modulated sinusoid and white Gaussian noise. The phase-lock demodulator is a suboptimal nonlinear filter whose performance is often quite good and which is used extensively in FM radio demodulation circuits.

If the state of an nth-order linear stochastic system is reconstructed from samples of the output by well-known numerical differentiation schemes, then even in the limit as the sampling becomes arbitrarily fine, well-known computations such as quadratic variation do not converge to the desired results (8). This phenomenon did not occur for linear stochastic differential equations in the approach in Ref. 37.

NONLINEAR FILTERING PROBLEM FORMULATION AND MAIN RESULTS

In this section a nonlinear filtering problem is formulated mathematically and many of the main results of nonlinear filtering are described.

A basic nonlinear filtering problem is described by two stochastic processes: $(X(t),\ t \geqslant 0)$, which is called the *signal* or *state process;* and $(Y(t),\ t \geqslant 0)$, which is called the *observation process.* These two processes satisfy the following stochastic differential equations:

$$dX(t) = a(t, X(t))\,dt + b(t, X(t))\,dB(t) \tag{1}$$

$$dY(T) = h(t, X(t), Y(t))\,dt + g(t, Y(t))\,d\tilde{B}(t) \tag{2}$$

where $t \geqslant 0$, $X(0) = x_0$, $Y(0) = 0$, $X(t) \in \mathbb{R}^n$, $Y(t) \in \mathbb{R}^m$, a: $\mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$, b: $\mathbb{R}_+ \times \mathbb{R}^n \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$, h: $\mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, g: $\mathbb{R}_+ \times \mathbb{R}^m \to \mathscr{L}(\mathbb{R}^m, \mathbb{R}^m)$, $(B(t),\ t \geqslant 0)$, and $(\tilde{B}(t),\ t \geqslant 0)$ are independent, standard Brownian motions in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The following assumptions are made on the coefficients of the stochastic differential equations (1) and (2) that describe the signal or state process and the observation process, respectively.

The stochastic processes are defined on a fixed probability space $(\Omega, \mathscr{F}, P)$ with a filtration $(\mathscr{F}_t,\ t \geqslant 0)$. Often the space Ω can be realized as a family of continuous functions. The σ-algebras are assumed to be complete with respect to the probability measure P.

A1. The drift vector $a(t, x)$ in Eq. (1) is continuous in t and globally Lipschitz continuous in x. The vector $\nabla a(t, x)$ is continuous in t and globally Lipschitz continuous in x.

A2. The diffusion matrix $b(t, x)$ in Eq. (1) is Hölder continuous in t, globally Lipschitz continuous in x, and globally bounded. The symmetric matrix $c = b^{\mathrm{T}}b$ is strictly positive definite uniformly in (t, x). The terms

$$\frac{\partial c_{ij}(t,x)}{\partial x_i}, \quad \frac{\partial^2 c_{ij}(t,x)}{\partial x_i \partial x_j}, \qquad i, j \in \{1, \ldots, n\}$$

are continuous in t, globally Lipschitz continuous in x, and globally bounded.

A3. The drift vector $h(t, x, y)$ and the diffusion matrix $g(t, y)$ in Eq. (2) are continuous in x. The symmetric matrix $f = g^{\mathrm{T}}g$ is strictly positive definite uniformly in (t, x); that is, $\langle fx, x\rangle \geqslant c|x|^2$, where $c > 0$.

The global Lipschitz conditions on the coefficients of the stochastic differential equations in the "space" variables x and y ensure the existence and the uniqueness of the strong (i.e., sample path) solutions of these equations. The additional smoothness properties are used to verify properties of the transition density of the Markov process $(X(t),\ Y(t),\ t \geqslant 0)$. Since the soluton of Eq. (1) is "generated" by x_0 and $(B(t),\ t \geqslant 0)$, the process $(X(t),\ t \geqslant 0)$ is independent of the Brownian motion $(\tilde{B}(t),\ t \geqslant 0)$ and the process $(Z(t),\ t \geqslant 0)$ where

$$dZ(t) = g(t, Z(t))d\tilde{B}(t) \tag{3}$$

A transition probability measure or a transition probability function for a Markov process is a function $P(s, x; t, \Gamma)$ for $s \in [0, t)$, $x \in \mathbb{R}^d$, and $\Gamma \in B_{\mathbb{R}^d}$ the Borel σ-algebra on $\mathbb{R}^d$ that satisfies the following:

1. $P(s, x; t, \cdot)$ is a probability measure on $(\mathbb{R}^d, B_{\mathbb{R}^d})$ for all $s \in [0, t)$.
2. $P(s, \cdot; t, \Gamma)$ is $B_{\mathbb{R}^d}$-measurable for all $s \in [0, t)$ and $\Gamma \in B\mathbb{R}_d$.
3. If $s \in [0, t)$, $u > t$ and $\Gamma \in B_{\mathbb{R}^d}$, then

$$P(s, x; u, \Gamma) = \int P(t, y; u, \Gamma) P(s, x; t, dy) \tag{4}$$

With the notion of transition probability measure (function), a Markov process can be defined.

Definition. Let $P(s, x; t, \cdot)$ be a transition probability measure and μ be a probability measure on $(\mathbb{R}^d, B_{\mathbb{R}^d})$. A probability measure P on $((\mathbb{R}^d)^{\mathbb{R}_+}, B_{(\mathbb{R}^d)^{\mathbb{R}_+}})$ is called *a Markov process* with transition function $P(s, x; t, \cdot)$ and initial distribution μ if

$$P(X(0) \in \Gamma) = \mu(\Gamma), \qquad \Gamma \in B_{\mathbb{R}^d}$$

and for each $s \in [0, t)$ and $\Gamma \in B_{\mathbb{R}^d}$

$$P(x(t) \in \Gamma | \sigma(X(u), 0 \le u \le s)) = P(s, X(s); t, \Gamma) \tag{5}$$

The random variable $X(t)$ is the evaluation of an element on $(\mathbb{R}^d)^{\mathbb{R}_+}$ at $t \in \mathbb{R}_+$. Usually the Markov process is identified as $(X(t), t \geqslant 0)$ and the Markov property (5) is described as

$$P(X(t) \in \Gamma | X(u), 0 \le u \le s) = P(s, X(s); t, \Gamma) \tag{6}$$

If $P(s, x; t, \cdot) = P(t - s, x, \cdot)$, then the Markov process is said to be (time) homogeneous.

If $(X(t), t \geqslant 0)$ is a homogeneous Markov process, then there is a semigroup of operators $(P_t, t \geqslant 0)$ acting on the bounded Borel measurable functions (39), which is given by

$$(P_t \psi)(x) = E_x[\psi(X(t))] = \int \psi(y) P(0, x; t, dy) \tag{7}$$

Consider the restriction of $(P_t, t \geqslant 0)$ to the bounded, continuous functions that vanish at infinity which is a Banach space in the uniform topology. If $(X(t), t \geqslant 0)$ is a Markov process that is the solution of the stochastic differential equation

$$dX(t) = a(X(t))\,dt + b(X(t))\,dB(t)$$

where $a(\cdot)$ and $b(\cdot)$ satisfy a global Lipschitz condition, then the semigroup $(P_t, t \geqslant 0)$ has an infinitesimal generator that is easily computed from Itô's formula (36); that is,

$$Lf = \lim_{t \downarrow 0} \frac{P_t f - f}{t} \tag{8}$$

for

$$f \in D(L) = \left\{ f : \lim_{t \downarrow 0} \frac{P_t f - f}{t} \text{ exists} \right\}$$

It is straightforward to verify that

$$L = \sum_{i=1}^{d} a_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum c_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \tag{9}$$

where $c = b^{\mathrm{T}} b$.

An analogous result holds if the Markov process is not homogeneous, so that

$$dX(t) = a(t, X(t))\,dt + b(t, X(t))\,dB(t)$$

where the theory of two-parameter semigroups is used so that

$$P_{s,t} f(x) = E_{X(s)=x}[f(X(t))]$$

and

$$\frac{dP_{s,t} f}{dt} = Lf$$

where

$$L = \sum a_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum c_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \tag{10}$$

In the filtering solution the formal adjoint of L, L^*, appears; that is,

$$L^* = \sum \frac{\partial}{\partial x_i}(a_i \cdot) + \frac{1}{2} \sum \frac{\partial^2}{\partial x_i \partial x_j}(c_{ij} \cdot) \tag{11}$$

The operator L is often called *backward operator,* and L^* is called the *forward operator.*

The stochastic integrals that occur in the solution of Eq. (1) are interpreted using the definition of K. Itô (14). For a smooth integrand the integral is the limit of Riemann sums where the integrand is evaluated at the left endpoint assuming that the integrand is suitably measurable. This definition preserves the martingale property of Brownian motion.

Let $(M(t), t \geqslant 0)$ be an Itô process or a semimartingale, that is

$$dM(t) = \alpha(t)\,dt + \beta(t)\,dB(t) \tag{12}$$

or

$$dM(t) = dA(t) + dN(t) \tag{13}$$

where $(A(t), t \geqslant)$ is a process of bounded variation and $(N(t), t \geqslant 0)$ is a martingale. The Fisk–Stratonovich or Stratonovich integral (16,17) of a suitable stochastic process $(\gamma(t), t \geqslant 0)$ is denoted as

$$\int_0^t \gamma(s) \circ dM(s) \tag{14}$$

This integral is defined from the limit of finite sums that are formed from partitions as in Riemann–Stieltjes integration where the function γ is evaluated at the midpoint of each interval formed from the partition. Recall that the Itô integral is formed by evaluating the integrand at the left endpoint of each of the subintervals formed from the partition (40).

For the linear filtering problem of Gauss–Markov processes it is elementary to show that the conditional probability density is Gaussian so that only the conditional mean and the conditional covariance have to be determined. Furthermore, the estimate of the state, given the observations that minimizes the variance, is the conditional mean. Thus one approach to the nonlinear filtering problem is to obtain a stochastic equation for the conditional mean or for other conditional statistics. The difficulty with this approach is that typically no finite family of equations for the conditional statistics is closed; that is, any finite family of equations depends on other conditional statistics.

The conditional probability density of the state $X(t)$, given the observations $(Y(u), 0 \leqslant u \leqslant t)$, is the density for the conditional probability that represents all of the probabilistic information about $X(t)$ from the observations $(Y(u), 0 \leqslant u \leqslant t)$. A conditional statistic can be computed by integrating the conditional density with a suitable function of the state.

To obtain a useful expression for the conditional probability measure, it is necessary to use a result for the absolute continuity of measures on the Borel σ-algebra of the space of continuous functions with the uniform topology. These results center around the absolute continuity for Wiener measure, the measure for Brownian motion. The first systematic investigation of Wiener measure in this context was done by Cameron and Martin (30), who initiated a calculus for Wiener measure. Subsequently, much work was done on general Gaussian measures (e.g., see Ref. 41).

For Wiener measure and some related measures, a more general probabilistic approach was given by Skorokhod (42,43) and Girsanov (44). The following result is a version of Girsanov's result.

Theorem. Let $(\Omega, \mathscr{F}, P)$ be a probability space with a filtration $(\mathscr{F}_t,\ t \in [0,\ T])$. Let $(\varphi(t),\ t \in [0,\ T])$ be an $\mathbb{R}^n$-valued process that is adapted to $(\mathscr{F}_t,\ t \in [0,\ T])$ and let $(B(t),\ t \in [0,\ T])$ be an $\mathbb{R}^n$-valued standard Brownian motion. Assume that

$$E[M(T)] = 1 \tag{15}$$

where

$$M(t) = \exp\left[\int_0^t \langle \varphi(s), dB(s)\rangle - \frac{1}{2}\int_0^t |\varphi(s)|^2\, ds\right] \tag{16}$$

Then the process $(Y(t),\ t \in [0,\ T])$ given by

$$Y(t) = B(t) - \int_0^t \varphi(s)\, ds \tag{17}$$

is a standard Brownian motion for the probability $\tilde{P}$, where $d\tilde{P} = M(T)\, dP$.

Let μ_Z be the probability measure on the Borel σ-algebra of $\mathbb{R}^m$-valued continuous functions for the process $(Z(t),\ t \geqslant 0)$ that is the solution of Eq. (3). Let $(V(t),\ t \geqslant 0)$ be the process that is the solution of

$$\begin{aligned} dV(t) &= b(t, V(t))\, dB(t) \\ V(0) &= x_0 \end{aligned} \tag{18}$$

Let μ_{XY} be the measure on the Borel σ-algebra of $\mathbb{R}^{n+m}$-valued continuous functions for the process $(X(t),\ Y(t),\ t \geqslant 0)$ that satisfy Eqs. (1) and (2). It follows from Girsanov's Theorem above that $\mu_{XY} \ll \mu_V \otimes \mu_Z$. The Radon–Nikodym derivative $\eta(t) = \varphi(t)\psi(t) = E[d\mu_{XY/d(\mu_V \otimes \mu_Z)} \mid \mathscr{F}_t]$ is

$$\begin{aligned}\varphi(t) = \exp\Bigg[&\int_0^t \langle c^{-1}(s, X(s))a(s, X(s)),\ dX(s)\rangle \\ &-\frac{1}{2}\int_0^t \langle c^{-1}(s, X(s))a(s, X(s)), a(s, X(s))\rangle\, ds\Bigg]\end{aligned} \tag{19}$$

$$\begin{aligned}\psi(t) = \exp\Bigg[&\int_0^t \langle f^{-1}(s, Y(s))g(s, X(s), Y(s)),\ dY(s)\rangle \\ &-\frac{1}{2}\int_0^t \langle f^{-1}(s, Y(s))g(s, X(s), Y(s)), g(s, X(s), Y(s))\rangle\, ds\Bigg]\end{aligned} \tag{20}$$

To indicate the expectation with respect to one of the function space measures, E is subscripted by the measure—for example, $E\mu_X$.

A result for the absolute continuity of measures is given that follows from the result of Girsanov (44). For convenience it is stated that $\mu_X \ll \mu_V$, though it easily follows that $\mu_{XY} \ll \mu_V\mu_Z$ and in fact there is mutual absolute continuity.

Theorem. Let μ_V and μ_X be the probability measures on $(\Omega, \mathscr{F})$ for the process $(V(t),\ t \in [0,\ T])$ and $(X(t),\ t \in [0,\ T])$, respectively, that are solutions of Eqs. (18) and (1). Then μ_X is absolutely continuous with respect to μ_V, denoted $\mu_X \ll \mu_V$, and

$$\frac{d\mu_X}{d\mu_V} = \varphi(T)$$

where φ is given by Eq. (19).

Corollary. Let $(V(t),\ t \in [0,\ T])$ satisfy (18) on $(\Omega, \mathscr{F}, \mu_V)$. Then

$$\hat{B}(t) = B(t) - \int_s^t b^{-1}(s, B(s))a(s, B(s))\, ds$$

is a Brownian motion on $(\Omega, \mathscr{F}, \mu_X)$.

It can be shown that the linear growth of the coefficients ensures that there is absolute continuity, so that Eq. (15) is satisfied (45).

The following result gives the conditional probability measure in function space (22,23).

Proposition. For $t > 0$ the conditional probability measure of $X(t)$, given $(Y(u),\ 0 \leqslant u \leqslant t)$, is given by

$$P(\Lambda, t \mid x_0, Y_u, 0 \le u \le t) = \frac{E_{\mu_Z}[1_\Lambda \varphi_t \psi_t]}{E_{\mu_Z}[\varphi_t \psi_t]} \tag{21}$$

for $\Lambda \in \sigma(X(t))$, the σ-algebra generated by $X(t)$.

The absolute continuity of measures and the associated Radon–Nikodym derivatives are important objects even in elementary probability and statistics. In this latter context there is usually a finite family of random variables that have a joint density with respect to Lebesgue measure. The likelihood function in statistical tests is an example of a computation of a Radon–Nikodym derivative.

The conditional probability measure Eq. (21) is specialized to the conditional probability distribution and a density is given for this function. The conditional density is shown to satisfy a stochastic partial differential equation (17, 20, 46).

Theorem. Let $(X(t),\ Y(t),\ t \geqslant 0)$ be the processes that satisfy Eqs. (1) and (2). If A1–A5 are satisfied, then

$$dp(t) = L^* p + \langle f^{-1}(t, Y(t))(g(t) - \hat{g}(t)), dY(t) - \hat{g}(t)\rangle p(t) \tag{22}$$

where

$$p(t) = p(X(t), t \mid x_0, Y(u), 0 \le u \le t) \tag{23}$$

$$g(t) = g(t, X(t), Y(t)) \tag{24}$$

$$\hat{g}(t) = \frac{E_{\mu_X}[\psi(t)g(t)]}{E_{\mu_X}[\psi(t)]} \tag{25}$$

Equation (22) is a nonlinear stochastic partial differential equation. The nonlinearity occurs from the terms $\hat{g}(t)p(t)$, and the partial differential operator L^* is the forward differential operator for the Markov process $(X(t), t \geqslant 0)$.

Often only some conditional statistics of the state given the observations are desired for the nonlinear filtering problem solution. However, such equations are usually coupled to an infinite family of conditional statistics. The following theorem describes a result for conditional statistics (47,48).

Theorem. Let $(X(t), Y(t), t \geqslant 0)$ satisfy Eqs. (1) and (2). Assume that $a(t, x)$ and $b(t, x)$ in Eq. (1) are continuous in t and globally Lipschitz continuous in x, $h(t, x, y)$ is continuous in t and globally Lipschitz continuous in x and y, and $g(t, y)$ is continuous in t and globally Lipschitz continuous in y and $f = g^{\mathrm{T}}g$ is strictly positive definite uniformly in (t, y). If $\gamma \in C^2(\mathbb{R}^n, \mathbb{R})$ such that

$$\int_0^T E|\gamma(X(t))|^2\,dt < \infty \tag{26}$$

$$E\int_0^T |\langle D\gamma(X(t)), X(t)\rangle|^2\,dt < \infty \tag{27}$$

$$E\int_0^T |\langle D^2\gamma(X(t))X(t), X(t)\rangle|^2\,dt < \infty \tag{28}$$

then the conditional expectation of $\gamma(X(t))$, given the observations $(Y(u), u \leqslant t)$,

$$\hat{\gamma}(t) = E[\gamma(X(t))\,|\,x_0, Y(u), 0 \leq u \leq t]$$

satisfies the stochastic equation

$$\begin{aligned} d\hat{\gamma}(t) = {} & L\gamma(X(t))\,dt + \langle f^{-1}(t, Y(t))\hat{\gamma}g(t, X(t), Y(t)) \\ & - \hat{\gamma}(X(t))\hat{g}(t, X(t), Y(t)), dY(t) - \hat{g}(t, X(t), Y(t))\,dt\rangle \end{aligned} \tag{29}$$

where L is the backward differential operator for the Markov process $(X(t), t \geqslant 0)$ and $\hat{}$ is conditional expectation— for example,

$$\hat{\gamma}(X(t)) = \frac{E_{\mu_X}[\gamma(X(t))\psi(t)]}{E_{\mu_X}[\psi(t)]} \tag{30}$$

The stochastic equation for the condition probability density is a nonlinear stochastic partial differential equation. The stochastic equation for a conditional statistic is typically coupled to an infinite famiily of such equations. The conditional density is more useful because it represents all of the probabilistic information about the state given the observations, but it is a nonlinear equation. If the so-called unnormalized conditional density is used, then the stochastic partial differential equation is linear. This unnormalized conditional density was given by Duncan (24), Mortensen (25), and Zakai (26). The equation is usually called the Duncan–Mortensen–Zakai (DMZ) equation.

Theorem. Let $(X(t), Y(t), t \geqslant 0)$ be the processes that are the solutions to Eqs. (1) and (2). Let r be given by

$$r(x, t\,|\,x_0, Y(u), 0 \leq u \leq t) = E_{\mu_X}[\psi(t)\,|\,X(t) = x]p_X(0, x_0; t, x) \tag{31}$$

where p_X is the transition density for the Markov process $(X(t), t \geqslant 0)$.

Assume that A1–A3 are satisfied. Then r satisfies the following linear stochastic partial differential equations:

$$\begin{aligned} & dr(X(t), t\,|\,x_0, Y(u), 0 \leq u \leq t) \\ & \quad = L^*r + \langle f^{-1}(t, Y(t))g(t, X(t), Y(t)), dY(t)\rangle r \end{aligned} \tag{32}$$

The normalization factor for r is

$$q(t) = E_{\mu_X}[\psi(t)] \tag{33}$$

so that

$$p(x, t\,|\,x_0, Y(u), 0 \leq u \leq t) = r(t)q^{-1}(t)$$

It is elementary to obtain a stochastic equation for $q^{-1}(t)$ using Itô's formula; that is,

$$\begin{aligned} dq^{-1}(t) = {} & -q^{-2}(t)\,dq(t) \\ & + q^{-3}(t)\langle f^{-1}(t)E_{\mu_X}[\psi(t)g(t)], E_{\mu_X}[\psi(t)g(t)]\rangle\,dt \end{aligned} \tag{34}$$

where

$$dq(t) = \langle f^{-1}(t)\psi(t)g(t), dY(t)\rangle \tag{35}$$

To apply algebro-geometric methods to the nonlinear filtering problem the following form of the DMZ equation is used

$$dr_t = [L^* - (1/2)\langle g_t, g_t\rangle]r_t\,dt + r_t g_t^{\mathrm{T}} \circ dY(t) \tag{36}$$

Recall that the symbol $\circ$ in Eq. (36) indicates the Stratonovich integral. The reason that this form of the DMZ equation is sometimes more useful is that it satisfies the usual rules of calculus. Thus the Lie algebras can be computed in the same way that would be used for smooth vector fields.

A more elementary nonlinear filtering problem than the one for diffusion processes that is important for applications is the case where the signal or state process is a finite-state Markov process (in continuous time). The finite-state space for the process significantly reduces the mathematical difficulties. Let $\mathscr{S} = \{s_1, \ldots, s_n\}$ be the state space for the finite state Markov process and $\tilde{p}^i(t) = P(X(t) = s_i)$ and $\tilde{p}(t) = [\tilde{p}^1(t), \ldots, \tilde{p}^n(t)]^{\mathrm{T}}$. It follows that

$$\frac{d}{dt}\tilde{p}(t) = A\tilde{p}(t) \tag{37}$$

where A is the intensity matrix or the transpose of the generator of the Markov process $(X(t), t \geqslant 0)$. The dependence of $\tilde{p}$ on the initial value $X(0)$ has been suppressed for notational convenience. By analogy with the DMZ equations (31) and (36) in this case for the finite-state Markov process (49) it follows that the unnormalized conditional density $\rho(t)$ satisfies

$$\rho(t) = \rho(0) + \int_0^s A\rho(s)\,ds + \int_0^t B\rho(s)\,dY(s) \tag{38}$$

or

$$\rho(t) = \rho(0) + \int_0^s (A - (1/2)B^2)\rho(s)\,ds + \int_0^t B\rho(s) \circ dY(s) \tag{39}$$

where $B = \mathrm{diag}(s_1, \ldots, s_n)$ and $\rho(t) = [\rho^1(t), \cdot, \rho^n(t)]^T$. These equations are a finite family of bilinear stochastic differential equations for the unnormalized conditional probabilities. The conditional expectation of the statistic $\varphi{:}\mathcal{S} \to \mathbb{R}$, denoted $\pi_t(\varphi)$, is

$$\pi_t(\varphi) = \frac{\sum_{i=1}^n \varphi(s_i)\rho^i(t)}{\sum_{i=1}^n \rho^i(t)} \tag{40}$$

A Lie algebra associated with the DMZ equation (36) plays a basic role in determining the existence or the nonexistence of finite-dimensional filters for conditional statistics of the signal (or state) process. To introduce Lie algebras, its definition is given.

Definition. A Lie algebra V over a field k is a vector space over k with a bilinear form $[\cdot,\cdot]{:}V \times V \to V$ (the Lie bracket) that satisfies for $v_1, v_2, v_3 \in V$ the following:

1. $[v_1, v_2] = -[v_2, v_1]$,
2. $[v_1, [v_2, v_3]] + [v_2, [v_3, v_1]] + [v_3, [v_1, v_2]] = 0$.

A Lie subalgebra of a Lie algebra V is a linear subspace of V that is a Lie algebra. If I, a subalgebra of V, is an ideal of V, then the quotient algebra is V/I, a vector space with the induced Lie bracket. A Lie algebra homomorphism $\varphi{:}V_1 \to V_2$ of the Lie algebras V_1 and V_2 is a linear map that commutes with the bracket operations, $\varphi([u, v]) = [\varphi(u), \varphi(v)]$.

The algebro-geometric methods for the nonlinear filtering problem arose from the system theory for finite-dimensional, nonlinear, affine, deterministic control systems. Consider a deterministic control system of the form

$$\frac{dx}{dt} = f(x(t)) + \sum_{i=1}^m u_i(t) g_i(x(t)) \tag{41}$$

where $x(t) \in M$, a smooth d-dimensional manifold. A controllability property has local significance in analogy to its global significance for linear control systems.

Definition. The controllability Lie algebra of Eq. (41) is the Lie algebra $\mathcal{L}$ generated by $\{f, g_1, \ldots, g_m\}$.$\mathcal{L}(x)$ is the linear space of vectors in T_xM, the tangent space of M at x, spanned by the vector fields of $\mathcal{L}$ at x. The dimension of $\mathcal{L}(x)$ has implication for the local reachable set starting at $x \in M$.

Another basic notion in system theory is observability. This condition implies that different "states"—that is, different points in M—can be distinguished using an appropriate control.

Definition. Consider the control system (41). Let $h \in C^\infty(M, \mathbb{R})$ give the observation as

$$y(t) = h(x(t)) \tag{42}$$

The system (41) and (42) is said to be *observable* under the following condition: If $x_1, x_2 \in M$ and $x_1 \neq x_2$, then there is an input (control) function u such that the outputs associated with x_1 and x_2 are not identical.

By analogy to the structure theory of linear systems, there is a "state-space" isomorphism theorem; that is, given two systems of the form (41) and (42) on two analytic manifolds such that the coefficients are complete analytic vector fields, the systems are observable and dim $\mathcal{L}(x)$ is minimal, and the two systems realize the same input–output map, then there is an analytic map between the manifolds that preserves trajectories (50). Associated with the equivalence of systems of the form (41) and (42) there is a realization of such systems that is observable and dim $\mathcal{L}(x) = n$; that is, the Lie algebra of vector fields evaluated at each point in the manifold has maximal dimension.

In general, the DMZ equation (36) can be viewed as the state equation of an infinite-dimensional system for $\rho(t)$ with the "input" function from the observation Y and the output $\pi_t(\varphi)$ (27,51). An investigation of the existence or the nonexistence of a finite dimensional realization of the input–output map is the investigation of the existence of finite-dimensional filters.

A finite-dimensional (recursive) filter for $\pi_t(\varphi)$ is a stochastic equation

$$d\eta(t) = a(\eta(t))\,dt + b(\eta(t)) \circ dY(s) \tag{43}$$

$$\pi_t(\varphi) = \gamma(\eta(t)) \tag{44}$$

where $\eta(t) \in \mathbb{R}^n$.

The application of the Lie algebraic methods described above and the use of nonlinear system theory presupposed a finite-dimensional manifold. For the DMZ equation (36) the solution evolves in an infinite-dimensional manifold. Thus it is necessary to be precise when translating these finite-dimensional algebro-geometric results to the DMZ equation. If this approach can be applied to the DMZ equation, then the questions of the existence or the nonexistence of finite-dimensional stochastic equations for conditional statistics and the equivalence of two nonlinear filtering problems can be resolved. Even for finite-state Markov processes it can be determined if some conditional statistics are the solutions of stochastic equations whose dimension is significantly smaller than the number of states of the Markov process (52).

For the DMZ equation (36) by analogy with the finite-dimensional input–output systems (41) and (42), the Lie algebra generated by the operators $L^* - (1/2)\langle h, h\rangle$ and $\langle h, \cdot\rangle$ acting on smooth (C^∞) functions is called the *estimation algebra* associated with Eqs. (41) and (42) (53,54,54a).

To identify equivalent filtering problems it is important to investigate transformations that induce isomorphic estimation algebras. A simple, important transformation is a change of scale of the unnormalized conditional probability density $r(\cdot)$. Let $\eta{:}\mathbb{R}^n \to \mathbb{R}$ be a strictly positive, smooth function and let $\tilde{r}(t) = \eta r(t)$. This transformation acts on the generators of the estimation algebra as

$$\eta L^* \eta^{-1} - \frac{1}{2}\langle h, h\rangle$$

$$\langle \eta g \eta^{-1}, \cdot\rangle$$

Thus the estimation algebras are formally isomorphic. Furthermore, a smooth homeomorphism of the state space induces an estimation algebra that is formally isomorphic to the initial estimation algebra. The above two operations on the estimation algebra have been called the *estimation formal equivalence group* (55).

If for some distribution of $X(0)$ a conditional statistic, $\pi_t(\varphi)$, can be described by a minimal finite-dimensional (recursive) filter of the form (43) and (44), then the Lie algebra of this system should be a homomorphism of the estimation algebra for this filtering problem. This property has been called the homomorphism principle for the filtering problem (56). This homomorphism principle can be a guide in the investigation of the existence of finite-dimensional filters.

A specific example of this homomorphism property occurs when the estimation algebra is one of the Weyl algebras. The Weyl algebra W_n is the algebra of polynomial differential operators over $\mathbb{R}$ with operators $x_1, \ldots, x_n, \partial/\partial x_1, \ldots, \partial/\partial x_n$. The Lie bracket is the usual commutator for differential operators. This Lie algebra has a one-dimensional center and the quotient $W_n/\mathbb{R}$ is simple; that is, it contains no nontrivial ideals. For the estimation algebra these two properties imply that if W_n is the estimation algebra for a filtering problem, then either the unnormalized conditional density can be computed by a finite-dimensional filter or no conditional statistic can be computed by a finite-dimensional filter of the form (43) and (44). More specifically, for $n \neq 0$ there are no nonconstant homomorphisms from W_n or $W_n/\mathbb{R}$ to the Lie algebra of smooth vector fields on a smooth manifold (57).

As an example of a Weyl algebra occurring as an estimation algebra, consider

$$dX(t) = dB(t) \tag{45}$$

$$dY(t) = X^3(t)\,dt + d\tilde{B}(t) \tag{46}$$

It is straightforward to verify that the estimation algebra for this filtering problem is the Weyl algebra W_1. The homomorphism principle that has been described can be verified in principle for this estimation algebra to show that there are no nontrivial conditional statistics that can be computed with finite-dimensional filters of the form (43) and (44) (58).

It is natural to consider the linear filtering problem using the estimation algebra method. Consider the following scalar model:

$$dX(t) = dB(t) \tag{47}$$

$$dY(t) = X(t)\,dt + d\tilde{B}(t) \tag{48}$$

The estimation algebra is a four-dimensional Lie algebra with the basis

$$\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2, x, \frac{\partial}{\partial x}, 1$$

This algebra is called the *oscillator algebra* in physics (27,28). The oscillator algebra is the semidirect product of $\mathbb{R}\cdot 1$ and the Heisenberg algebra that is generated by

$$\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2, x, \text{ and } \frac{\partial}{\partial x}$$

It can be verified that the Lie algebra of the linear filtering equations of Kalman and Bucy is isomorphic to the oscillator algebra.

A well-known example of a filtering problem given by Beneš (29) has a finite-dimensional filter and it is closely related to the linear filtering problem. Consider the scalar filtering problem

$$dX(t) = f(X(t))\,dt + dB(t) \tag{49}$$

$$dY(t) = X(t)\,dt + d\tilde{B}(t) \tag{50}$$

where f satisfies the differential equation

$$\frac{df}{dx} + f^2(x) = ax^2 + bx + c$$

for some $a, b, c \in \mathbb{R}$. It is assumed that this Riccati equation has a global solution, so that either $a > 0$ or $a = b = 0$ and $c > 0$. The unnormalized conditional density can be computed to verify that there is a ten-dimensional sufficient conditional statistic. However, Beneš (29) showed that a two-dimensional filter provides a sufficient conditional statistic. The estimation algebra $\tilde{\mathcal{L}}$ for (49)–(50) is generated by

$$\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x}f - \frac{1}{2}x^2, x$$

and is four-dimensional and solvable. The estimation algebra $\hat{\mathcal{L}}$ for (47)–(48) arises for the algebra $\tilde{\mathcal{L}}$ by letting $f \equiv 0$. To associate $\tilde{\mathcal{L}}$ with the estimation algebra $\hat{\mathcal{L}}$ let $F(x) = \int^x f$, $\psi(x) = \exp(-F(x))$ and $\tilde{r}(t, x) = \psi(x)r(t, x)$. Then the DMZ equation for (47) and (48) is transformed by the gauge transformation ψ as

$$d\tilde{r} = \left(\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}[(a+1)x^2 + bx + c]\right)\tilde{r} + x\tilde{r} \circ dY \tag{51}$$

This has the same form as the DMZ equation for (49) and (50). Thus the nonlinear filtering problem (49) and (50) is obtained from the linear filtering problem (47) and (48) by a gauge transformation of the conditional density. Various other examples of finite-dimensional filters are available (e.g., see Refs. 59–61).

Ocone (62) showed that for a scalar filtering problem with the observation equation of the form (50) the two examples (47)–(48) and (49)–(50) are the only ones that give a finite-dimensional estimation algebra. This result is given in the following theorem.

Theorem. Let $n = m = 1$ in (49)–(50) and let $g \equiv 1$ in the observation equation (2). Then the dimension of the estimation algebra is finite only if:

1. $h(x) = ax$ and

$$\frac{df}{dx} + f^2 = ax^2 + bx + c$$

or

2. $h(x) = \alpha x^2 + \beta x$, $\alpha \neq 0$ and

$$\frac{df}{dx} + f^2(x) = -h^2(x) + a(2\alpha x + \beta)^2 + b + c(2\alpha x + \beta)^{-1}$$

or

$$\frac{df}{dx} + f^2(x) = -h^2(x) + ax^2 + bx + c$$

Some multidimensional results are discussed in the section entitled "Some Recent Areas of Nonlinear Filtering."

Another family of nonlinear filtering problems given by Liptser and Shiryayev (63,64) that can be solved by the Gaussian methods is the conditional linear models. Let ($X(t)$, $Y(t)$, $t \geqslant 0$) satisfy

$$dX(t) = [A(t,Y)X(t) + a(t,Y)]\,dt + B(t,Y)\,dB(t) + \sum_{j=1}^{d}[G^j(t,Y)X(t) + g^j(t,Y)]\,dY^j(t) \tag{52}$$

$$dY(t) = [H(t,Y)X(t) + h(t,Y)]\,dt + d\tilde{B}(t) \tag{53}$$

where $X(0)$ is a Gaussian random variable, $Y(0) \equiv 0$. The random variable $X(t)$, given $\mathscr{Y}(t) = \sigma(Y(u), u \leqslant t)$, is conditionally Gaussian. More precisely, it is assumed that ($X(t)$, $t \geqslant 0$) is an $\mathbb{R}^n$-valued $(\mathscr{F}_t)$-adapted process, ($Y(t)$, $t \geqslant 0$) is an $\mathbb{R}^m$-valued $(\mathscr{F}_t)$-adapted process, ($B(t)$, $t \geqslant 0$) and ($\tilde{B}(t)$, $t \geqslant 0$) are independent standard $\mathbb{R}^n$- and $\mathbb{R}^m$-valued Brownian motions, respectively, in the filtered probability space (Ω, $\mathscr{F}$, $(\mathscr{F}_t)$, P), and $X(0)$ is $N(m_0, R_0)$. The functions A, a, B, G^j, g^j, H, and h are defined on $\mathbb{R}_+ \times C(\mathbb{R}_+, \mathbb{R}^m)$ with values in a suitable Euclidean space, and they are progressively measurable. The functions $|A|^2$, $|a|$, $|B|^2$, $|G^j|^2$, $|H|$, and $|h|$ are in $L^1_{\text{loc}}(\mathbb{R}_+)$ for each $y \in C(\mathbb{R}_+, \mathbb{R}^m)$. For each $T > 0$, $E\Lambda^{-1}(T) = 1$ where

$$\Lambda(T) = \exp\left[\int_0^T \langle H(s,Y)X(t) + h(s,Y), dY(s)\rangle - \tfrac{1}{2}\int_0^T |H(s,Y)X(s) + h(s,Y)|^2\,ds\right] \tag{54}$$

Haussmann and Pardoux (65) proved the following result.

Theorem. Consider the filtering problem (52) and (53). For each $T > 0$ the conditional distribution of $X(t)$, given $\mathscr{Y}(T) = \sigma(Y(u), u \leqslant T)$, is Gaussian.

Furthermore, if it is assumed that $|a|$, $|g^j|$, $|h|$, $|h|\,|g^j|$, $|A|$, $|B|$, $|G^j|^2$, $|H^2|$, $|h|\,|G^j|$, $|g^j|\,|H|$, and $|H(t, Y)X(t) + h(t, Y)|$ are in $L^2([0, T] \times \Omega)$ for all $T \in \mathbb{R}_+$, then the following result for the conditional mean $\tilde{m}(t)$ and the conditional covariance $R(t)$ can be verified (65)

Theorem. Consider the filtering problem (52) and (53). The conditional mean $\tilde{m}(t)$ and the conditional covariance $\tilde{R}(t)$ satisfy the following equations:

$$d\tilde{m}(t) = \Big[A(t,Y)\tilde{m}(t) + a(t,Y) - \tilde{R}(t)H(t,Y)^*[H(t,Y)\tilde{m}(t) + h(t,Y)] + \sum_j G^j(t,Y)\tilde{R}(t)H^j(t,Y)\Big]\,dt + \sum_j [G^j(t,Y)\tilde{m}(t) + g^j(t,Y) + \tilde{R}(t)H^j(t,Y)^* dY^j(t)] \tag{55}$$

and

$$d\tilde{R}(t) = \Big[B(t,Y)B(t,Y)^* + A(t,Y)\tilde{R}(t) + \tilde{R}(t)A(t,Y)^* + \sum_j G^j(t,Y)\tilde{R}(t)G^j(t,Y)^* - \tilde{R}(t)H(t,Y)^*H(t,Y)\tilde{R}(t)\Big]\,dt + \sum_j [G^j(t,Y)\tilde{R}(t) + \tilde{R}(t)G^j(t,Y)^*]\,dY^j(t) \tag{56}$$

where $\tilde{m}(0) = m_0$ and $\tilde{R}(0) = R_0$.

SOME RECENT AREAS OF NONLINEAR FILTERING

A generalization of the filtering problem occurs when some or all of the processes take values in an infinite-dimensional space such as a Hilbert, Banach, or Fréchet space. A basic question is the existence of a probability measure on one of these infinite-dimensional spaces. For example, for the existence of a zero mean Gaussian measure in a Hilbert space it is necessary and sufficient that the covariance is nuclear (trace class). The usual Daniell–Kolmogorov construction of a probability measure from a projective family of measures on finite-dimensional subspaces (finite dimensional distributions) does not guarantee a measurable space with a "nice" topology.

However, in some cases in infinite-dimensional spaces it is possible to use a cylindrical noise (e.g., the covariance of the Gaussian process is the identify) and have it "regularized" by the system so that the stochastic integral in the variation of parameters formula is a nice process. To describe this approach consider a semilinear stochastic differential equation

$$sX(t) = -AX(t)\,dt + f(X(t))\,dt + Q^{1/2}\,dW(t) \tag{57}$$

where $X(0)$, $X(t) \in H$, a separable, infinite-dimensional Hilbert space, and ($W(t)$, $t \geqslant 0$) is a standard cylindrical Wiener process. A standard cylindrical Wiener process means that if ℓ_1, $\ell_2 \in H = H^*$, $\langle \ell_1, \ell_2\rangle = 0$, and $\langle \ell_1, \ell_1\rangle = \langle \ell_2, \ell_2\rangle = 1$ where $\langle\cdot, \cdot\rangle$ is the inner product in H, then ($\langle \ell_1, W(t)\rangle$, $t \geqslant 0$) and ($\langle \ell_2, W(t)\rangle$, $t \geqslant 0$) are independent standard Wiener processes. If $-A$ is the generator of an analytic semigroup ($S(t)$, $t \geqslant 0$) and $S(r)Q^{1/2}$ is Hilbert–Schmidt for each $r > 0$ and

$$\int_0^t |S(r)Q^{1/2}|^2_{L_2(H)}\,dr < \infty \tag{58}$$

where $|\cdot|_{L_2(H)}$ is the norm for the Hilbert–Schmidt operators, then the process ($Z(t)$, $t \geqslant 0$) where

$$Z(t) = \int_0^t S(t-r)Q^{1/2}\,dW(r) \tag{59}$$

is an H-valued process that has a version with continuous sample paths. Thus, the solution of Eq. (57) with some suitable assumptions on f (66) can be given by the mild solution (67)

$$X(t) = S(t)X(0) + \int_0^t S(t-r)f(X(r))\,dr + \int_0^t S(t-r)Q^{1/2}\,dW(r) \tag{60}$$

This semigroup approach can be used to model stochastic partial differential equations arising from elliptic operators and delay-time ordinary differential equations. Some problems of ergodic control and stochastic adaptive control are described in Refs. 66 and 68.

For the stochastic partial differential equations it is natural to consider noise on the boundary of the domain or at discrete points in the domain. Furthermore, signal processes can be considered to be on the boundary or at discrete points in the domain. Some of the descriptions of these noise processes can be found in Refs. 66 and 68.

For the nonlinear filtering problem it can be assumed that the signal process is infinite-dimensional and that the observation process is finite-dimensional; or perhaps more interestingly it can be assumed that the signal process is finite-dimensional, occurring at distinct points of the domain or the boundary, and that the observation process is infinite-dimensional in the domain.

Another nonlinear filtering formulation occurs when the processes evolve on manifolds. This approach requires the theory of stochastic integration in manifolds (69). Many well-known manifolds arise naturally in the modeling of physical systems such as spheres and positive definite matrices. To justify the conditional mean as the minimum variance estimate and to compare the estimate and the signal, it is useful to model the signal process as evolving in a linear space or a family of linear spaces. The observation process can evolve in a manifold and have the drift vector field depend on the signal process, or the observation process can be the process in the base of a vector bundle; for example, the tangent bundle and the signal can evolve in the fibers of the vector bundle (70,71). These formulations allow for some methods similar to filtering problems in linear spaces. An estimation problem in Lie groups is solved in Ref. 72. The DMZ equation for a nonlinear filtering problem in a manifold is given in Ref. 73. A description of the stochastic calculus on manifolds with applications is given in Ref. 74.

The study of estimation algebras for nonlinear filtering problems has been a recent active area for nonlinear filtering. A number of questions naturally arise for estimation algebras. A fundamental question is the classification of finite-dimensional estimation algebras. This classification would clearly provide some important insight into the nonlinear filtering problem. This classification has been done for finite-dimensional algebras of maximal rank that correspond to state-space dimensions less than or equal to four (60,75–77). The result is described in the following theorem.

Theorem. Consider the filtering problem described by the following stochastic differential equations:

$$\begin{aligned} dX(t) &= f(X(t))\,dt + g(X(t))\,dV(t) \\ X(0) &= x_0 \end{aligned} \tag{61}$$

$$\begin{aligned} dY(t) &= h(X(t))\,dt + dW(t) \\ Y(0) &= 0 \end{aligned} \tag{62}$$

Assume that $n \leqslant 4$, where $X(t) \in \mathbb{R}^n$ and $Y(t) \in \mathbb{R}$. If $\mathscr{E}$ is the finite-dimensional estimation algebra of maximal rank, then the drift term f must be a linear vector field plus a gradient vector field and $\mathscr{E}$ is a real vector space of dimension $2n + 2$.

Another basic question is to find necessary conditions for finite-dimensional estimation algebras. It was conjectured by Mitter (28) that the observation terms are polynomials of degree at most one. An important result related to this conjecture is the following result of Ocone (62,78) that describes polynomials in the estimation algebra.

Theorem. Let $\mathscr{E}$ be a finite-dimensional estimation algebra. If φ is a function in $\mathscr{E}$, then φ is a polynomial of degree at most two.

The following result of Chen and Yau (79) verifies the Mitter conjecture for a large family of estimation algebras.

Theorem. If $\mathscr{E}$ is a finite-dimensional estimation algebra of maximal rank, then the polynomials in the drift of the observation equation (6) are degree-one polynomials.

Two basic approaches to the problem of finite-dimensional filters for nonlinear filtering problems are the Wei–Norman approach and the symmetry (or invariance) group approach (e.g., see Ref. 80). Wei and Norman (81) provided a global representation of a solution of a linear differential equation as a product of exponentials. The Wei–Norman approach requires an extension of the Wei–Norman results to semigroups. This has been done by introducing some function spaces or using some results for the solutions of partial differential equations (82,83). This result is important for the construction of finite-dimensional filters from finite dimensional estimation algebras (e.g., see Refs. 82 and 83).

Recall the problem of the existence of finite-dimensional filters for a linear filtering problem with a non-Gaussian initial condition. The question of finite-dimensional filters for nonlinear filtering problems can be formulated in different ways. In one formulation the probability law of the initial condition is fixed. It has been shown (82) that a necessary condition for a finite-dimensional filter is the existence of a nontrivial homomorphism from the estimation algebra into the Lie algebra of vector fields on a manifold.

Another formulation of the finite-dimensional filter problem is the requirement that a filter exist for all Dirac measures of the initial condition. It has been shown (82) that if a finite-dimensional filter has a regularity property with respect to initial conditions and dynamics, then the estimation algebra is finite-dimensional.

For linear filtering it is elementary to verify that minimizing a quadratic form which is the negative of the formal exponent in a likelihood function gives the solution of the linear filtering problem. By analogy, an approach to the nonlinear filtering problem based on minimizing a formal likelihood function in function space was introduced in the 1960s (84,85). This approach has been generalized and made rigorous by Fleming and Mitter (86) by relating a filtering problem to a stochastic control problem. This method uses a logarithmic transformation.

For an example of this method of logarithmic transformation consider the following linear parabolic partial differential equation:

$$\begin{aligned} p_t &= \frac{1}{2}\mathrm{tr}\, a(x)p_{xx} + \langle g(x,t), p_x\rangle + V(x,t)p \\ p(x,0) &= p^0(x) \end{aligned} \tag{63}$$

It is assumed that there is a $C^{2,1}$ solution. If this solution is positive, then $S = -\log p$ satisfies the nonlinear parabolic equation:

$$S_t = \frac{1}{2}\mathrm{tr}\, a(x)S_{xx} + H(x,t,S_x) \tag{64}$$

$$S(x,0) = -\log p^0(x) = S^0(x) \tag{65}$$

$$H(x,t,S_x) = \langle g(x,t), S_x\rangle - \frac{1}{2}\langle a(x), S_x, S_x\rangle - V(x,t) \tag{66}$$

This type of transformation is well known. For example, it transforms the heat equation ($g = V = 0$) into Burger's equation.

The nonlinear PDE (64) is the dynamic programming (Hamilton–Jacobi–Bellman) equation for a stochastic control problem. For example, let ($X(t)$, $t \geqslant 0$) satisfy

$$\begin{aligned} dX(t) &= (g(X(t),t) + u(X(t),t))\,dt + \sigma(X(t))\,dB(t) \\ X(0) &= x \end{aligned} \tag{67}$$

and let the cost functional be

$$J(x,t,u) = E_x\left[\int_0^t L(X(s), t-s, u(s))\,ds + S^0(X(t))\right] \tag{68}$$

where

$$L(x,t,u) = \frac{1}{2}\langle a^{-1}(x)u, u\rangle - V(x,t) \tag{69}$$

With suitable assumptions on the family of admissible controls and conditions on the terms in the model it can be shown from the Verification Theorem (87) that Eq. (63) is the dynamic programming equation for this stochastic control problem. This approach can provide a rigorous basis for the formal maximization of a likelihood function in function space. See Ref. 87a.

An approach to the robustness of the nonlinear filter (87b,87c) is to obtain a so-called pathwise solution to the Duncan–Mortensen–Zakai (DMZ) equation by expressing the solution as an (observation) path dependent semigroup. The infinitesimal generator of this semigroup is the conjugation of the generator of the signal process by the observation path multiplied by the drift in the observation where the Stratonovich form of the DMZ equation is used. The fact that the observation path appears explicitly rather than its differential implies the robustness of the solution of the DMZ equation.

It is important to obtain estimation methods that are applicable to both stochastic disturbances (noise) and deterministic disturbances. For Brownian motion, a Hilbert (or Sobolev space) of functions that are functions that are absolutely continuous and whose derivatives are square integrable having probability zero often plays a more important role than the Banach space of continuous functions that has probability one. This Hilbert space (or Sobolev space) alludes to the fact that there are some natural relations between stochastic and deterministic disturbances. In recent years the study of risk sensitive control problems has occupied an important place in stochastic control. Risk sensitive control problems (i.e., control problems with an exponential cost) have been used with the maximum likelihood methods in (85,87d) to obtain robust nonlinear filters, that is, filters that are effective for square integrable disturbances as well as Gaussian white noise (87e). These ideas are related to the approach of H control as a robust control approach. The robust nonlinear filter can be naturally related to a robust nonlinear observer. For a number of problems there is a natural relation between estimation for deterministic and stochastic systems. For example, a weighted least squares algorithm can be used for the identification of parameters for both deterministic and stochastic systems.

Since it is usually not feasible to solve explicitly the stochastic equation for the conditional mean or the conditional covariance for a nonlinear filtering problem it is important to obtain lower and upper bounds on the filtering error. These bounds enable an effective comparison of the suboptimal filters with the optimal filter. The bounds have typically been obtained as the noise approaches zero (87f).

An important theoretical and practical problem in nonlinear filtering is the infinite time stability or continuity of the filter with respect to the initial conditions and the parameters of the filter. The problem of stability of the optimal nonlinear filter with respect to initial conditions is investigated in (87g) for two different cases. Stability of the Riccati equation for linear filtering is used to obtain almost sure asymptotic stability for linear filters with possible non-Gaussian initial conditions. For signals that are ergodic diffusions it is shown that the optimal filter is asymptotically stable in the sense of weak convergence of measures for incorrect initial conditions. Another stability property that is important for the optimal filter is asymptotic stability with respect to the parameters of the filter.

Another important question in nonlinear filtering is to develop numerical methods for the DMZ equation. One numerical approach to the solution of the DMZ equation is to consider that it is a stochastic partial differential equation of a special form and use numerical methods from PDE for the numerical discretization of the problem (e.g., finite-difference schemes). There has been some success with this approach (88–90), but it is limited to small space dimension and also often to the small intervals of time.

Another approach is to use the Wiener chaos expansion that is based on an orthogonal expansion of a square integrable functional on Wiener space (30,31,91). The solution, r, of the DMZ equation is expressed in the following expansion (92):

$$r(t,x) = \sum \frac{1}{\sqrt{\alpha!}}\varphi_\alpha(t,x)\psi_\alpha(y) \tag{70}$$

where ψ_α are Wick polynomials (special products of Hermite polynomials) formed from Wiener integrals and φ_α are Hermite–Fourier coefficients in the orthogonal expansion. The separation of x and y in the expansion (70) implies a splitting in the computations, one that is associated with the Markov process ($X(t)$, $t \geqslant 0$) and the other one that depends on the observations ($Y(t)$, $t \geqslant 0$). The family of functions (φ_α) can be computed recursively from a family of equations of Kolmogorov type (92,93). The Wick polynomials (ψ_α) depend on the observations and there are numerical methods to compute them. For the numerical success, the Wick polynomials have to be computed in special ways. A direct approach to the evaluation of the Wick polynomials and thereby the expansion (70) is limited to short time intervals because the errors of truncation of the infinite series increase rapidly, probabily exponentially.

The numerical methods for nonlinear filtering are still under active development. It seems that all of the methods are limited to at most two or three space dimensions, and many methods have significant restrictions on the time intervals that are allowable for computations. However, these methods have been demonstrated to perform significantly better than more elementary methods such as the extended linear filter which is commonly used. Many stochastic problems require nonlinear filtering, so the filtering problems have to be addressed.

It should be clear that the area of nonlinear filtering is still an active area for research. This research includes mathematical investigations using methods from probability and geometry and implementation investigations that use numerical schemes for solving stochastic problems.

BIBLIOGRAPHY

1. A. N. Kolmogorov, Sur l'interpolation et extrapolation des suites stationnaires, *C. R. Acad. Sci.,* **208**: 2043, 1939.
2. A. N. Kolmogorov, Stationary sequences in Hilbert space (in Russian), *Bull. Math. Univ. Moscow,* **2**: 6, 1941.
3. M. G. Krein, On a generalization of some investigations of G. Szegö, W. M. Smirnov, and A. N. Kolmogorov, *Dokl. Adad. Nauk SSSR,* **46**: 91–94, 1945.
4. M. G. Krein, On a problem of extrapolation of A. N. Kolmogorov, *Dokl. Akad. Nauk SSSR,* **46**: 306–309, 1945.
5. N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications,* New York: Technology Press and Wiley, 1949. (Originally issued in February 1942, as a classified National Defense Research Council Report)
6. H. Wold, *A Study in the Analysis of Stationary Time Series,* 2nd ed., Uppsala, Sweden: Almqvist and Wiksell, 1938.

6a. H. W. Bode and C. E. Shannon, A simplified derivation of linear least squares smoothing and prediction theory, *Proc. IRE,* **38**: 417–425, 1950.

7. N. Wiener and E. Hopf, On a class of singular integral equations, *Proc. Prussian Acad., Math.-Phys. Ser.,* 696, 1931.
8. J. L. Doob, Wiener's work in probability theory, *Bull. Am. Math. Soc.,* **72**: 69–72, 1966.
9. R. E. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME, Ser. D, J. Basic Eng.,* **82**: 34–45, 1960.
10. R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, *Trans. ASME, Ser. D, J. Basic Eng.,* **83**: 95–107, 1961.
11. K. Itô, On stochastic differential equations, *Mem. Am. Math. Soc.,* **4**: 1951.
12. N. Wiener, Differential space, *J. Math. Phys.,* **2**: 131–174, 1923.
13. P. Lévy, *Processus Stochastiques et Mouvement Brownien,* 2nd ed., Paris: Gauthier-Villars, 1964.
14. K. Itô, Stochastic integral, *Proc. Imp. Acad. Tokyo,* **20**: 519–524, 1944.
15. P. A. Meyer, *Probability and Potentials,* Waltham, MA: Blaisdell, 1966.
16. D. L. Fisk, Quasi-martingales and stochastic integrals, Technical Report 1, Dept. of Mathematics, Michigan State University, 1963.
17. R. L. Stratonovich, Conditional Markov process theory, *Theory Probability Appl. (USSR),* **5**: 156–178, 1960.
18. E. J. McShane, Stochastic differential equations and models of random processes, In *Proceedings of the 6th Symposium on Mathematical, Statistics, and Probability,* Berkeley, CA: University of California Press, 1973.
19. S. S. Wilks, *Mathematical Statistics,* New York: Wiley, 1962.
20. H. Kushner, On the differential equations satisfied by conditional densities of Markov processes, with applications, *SIAM J. Control,* **2**: 106–119, 1964.
21. R. L. Stratonovich, Application of the theory of Markov processes for optimum filtration of signals, *Radio Eng. Electron. Phys. (URRS),* **1**: 1–19, 1960.
22. R. S. Bucy, Nonlinear filtering theory, *IEEE Trans. Autom. Control,* **AC-10**: 198, 1965.
23. G. Kallianpur and C. Striebel, Estimation of stochastic systems: Arbitrary system process with additive white observation errors, *Ann. Math. Statist.,* **39**: 785–801, 1969.
24. T. E. Duncan, Probability Densities for Diffusion Processes with Applications to Nonlinear Filtering Theory and Detection Theory, Ph.D. Dissertation, Stanford University, Stanford, CA, June 1967.
25. R. E. Mortensen, Optimal Control of Continuous-Time Stochastic Systems, Ph.D. dissertation, University of California, Berkeley, CA, 1966.
26. M. Zakai, On the optimal filtering of diffusion processes, *Z. Wahrsch, Verw. Geb.* **11**: 230–243, 1969.
27. R. Brockett, Remarks on finite dimensional nonlinear estimation, *Astérisque,* **764**: 47–55, 1980.
28. S. K. Mitter, On the analogy between mathematical problems on nonlinear filtering and quantum physics, *Ric. Autom.,* **10** (2): 163–216, 1980.

28a. S. K. Mitter, Filtering theory and quantum fields, *Astérisque,* **75–76**: 199–205, 1980.

29. V. E. Beneš, Exact finite dimensional filters for certain diffusions with nonlinear drift, *Stochastics,* **5**: 65–92, 1981.
30. R. H. Cameron and W. T. Martin, The orthogonal development of non-linear functionals in a series of Fourier–Hermite functions, *Ann. Math.,* **48**: 385–392, 1947.
31. K. Itô, Multiple Wiener integral, *J. Math. Soc. Japan,* **3**: 157–169, 1951.
32. Y. Hu and P. A. Meyer, *Sur les intégrales multiples de Stratonovich,* Lecture Notes in Mathematics, No 1321, Berlin: Springer-Verlag, pp. 72–81, 1988.
33. D. F. Allinger and S. K. Mitter, New results on the innovations problem for nonlinear filtering, *Stochastics,* **4**: 339–348, 1981.
34. T. E. Duncan, Evaluation of likelihood functions, *Inform. Control* **13**: 62–74, 1968.
35. T. E. Duncan, On the calculation of mutual information, *SIAM J. Math.,* **19**: 215–220, 1970.
36. K. Itô, On a formula concerning stochastic differentials, *Nagoya Math. J.,* **3**: 55–65, 1951.
37. E. Wong and M. Zakai, On the relation between ordinary and stochastic differential equations, *Int. J. Eng. Sci.,* **3**: 213–229, 1965.
38. T. E. Duncan, P. Mandl, and B. Pasik-Duncan, Numerical differentiation and parameter estimation in higher order stochastic systems, *IEEE Trans. Autom. Control,* **41**: 522–532, 1996.

39. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations,* New York: Springer-Verlag, 1983.
40. I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus,* 2nd ed., New York: Springer-Verlag, 1991.
41. I. E. Segal, Tensor algebras over Hilbert spaces. I, *Trans. Am. Math. Soc.,* **81**: 106–134, 1956.
42. A. V. Skorokhod, On the differentiability of measures which correspond to stochastic processes. I: Processes with independent increments, *Theory Probability Appl.,* **2**: 407–432, 1957.
43. A. V. Skorokhod, *Studies in the Theory of Random Processes,* Reading, MA: Addison-Wesley, 1965.
44. I. V. Girsanov, On transforming a certain class of stochastic processes by absolutely continuous substitution of measures, *Theory Probability Appl.,* **5**: 285–301, 1960.
45. T. E. Duncan and P. Varaiya, On the solutions of a stochastic control system. II, *SIAM J. Control,* **13**: 1077–1092, 1975.
46. A. N. Shiryayev, On stochastic equations in the theory of conditional Markov processes, *Theor. Probability Appl.,* **11**: 179–184, 1966.
47. M. Fujisaki, G. Kallianpur, and H. Kunita, Stochastic differential equations for non-linear filtering, *Osaka J. Math.,* **9**: 19–40, 1972.
48. G. Kallianpur, *Stochastic Filtering Theory,* New York: Springer-Verlag, 1980.
49. W. M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, *SIAM J. Control,* **2**: 347–369, 1965.
50. H. Sussmann, Existence and uniqueness of minimal realizations of nonlinear systems, *Math. Syst. Theory,* **10**: 263–284, 1977.
51. R. Brockett and J. M. C. Clark, The geometry of the conditional density equation, In O. Jacobs et al. (eds.), *Analysis and Optimization of Stochastic Systems,* New York: Academic Press, 1980, pp. 299–310.
52. S. I. Marcus, Low dimensional filters for a class of finite state estimation problems with Poisson observations, *Syst. Control Lett.,* **1**: 237–241, 1982.
53. M. Davis and S. Marcus, An introduction to nonlinear filtering, In M. Hazewinkel and J. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications,* Dordrecht, The Netherlands: D. Riedel, 1981, pp. 53–75.
54. S. I. Marcus, Algebraic and geometric methods in nonlinear filtering, *SIAM J. Control,* **22**: 817–844, 1984.
54a. S. K. Mitter, Geometric theory of nonlinear filtering, *Outils et Modèles Mathematiques pour l'Automatiques,* **3**: CNRS, 37–60, 1983.
55. R. W. Brockett, Classification and equivalence in estimation theory, In *Proceedings of the 18th IEEE Conference on Decision and Control,* Ft. Lauderdale, FL, 1979, pp. 172–175.
56. R. Brockett, Nonlinear systems and nonlinear estimation theory, In M. Hazewinkel and J. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications,* Dordrecht, The Netherlands: D. Riedel, pp. 442–477, 1981.
57. M. Hazewinkel and S. Marcus, Some results and speculations on the role of Lie algebras in filtering, In M. Hazewinkel and J. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications,* Dordrecht, The Netherlands: D. Riedel, pp. 591–604, 1981.
58. M. Hazewinkel, S. Marcus, and H. Sussmann, Nonexistence of exact finite dimensional filters for conditional statistics of the cubic sensor problem, *Systems Control Lett.,* **3**: 331–340, 1983.
59. D. Ocone, J. S. Baras, and S. I. Marcus, Explicit filters for diffusions with certain nonlinear drifts, *Stochastics,* **8**: 1–16, 1982.
60. W. Wong, Theorems on the structure of finite dimensional estimation algebras, *Systems Control Lett.,* **9**: 117–124, 1987.
61. S. S.-T. Yau, Finite dimensional filters with nonlinear drift. I: A class of filters including both Kalman–Bucy filters and Beneš filters, *J. Math. Systems, Estim. Control* **4**: 181–203, 1994.
62. D. Ocone, Topics in Nonlinear Filtering Theory, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1980.
63. R. Sh. Liptser and A. N. Shiryayev, *Statistics of Random Processes. I,* New York: Springer-Verlag, 1977.
64. R. Sh. Liptser and A. N. Shiryayev, *Statistics of Random Processes. II,* New York: Springer-Verlag, 1978.
65. U. Haussmann and E. Pardoux, A conditionally almost linear filtering problem with nongaussian initial condition, *Stochastics,* **23**: 241–275, 1988.
66. T. E. Duncan, B. Maslowski, and B. Pasik-Duncan, Ergodic boundary/point control of stochastic semilinear systems, *SIAM J. Control Optim.,* **36**: 1998.
67. G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions,* New York: Cambridge University Press, 1992.
68. T. E. Duncan, B. Maslowski, and B. Pasik-Duncan, Adaptive boundary control of linear distributed parameter systems described by analytic semigroups, *Appl. Math. Optim.,* **33**: 107–138, 1996.
69. T. E. Duncan, Stochastic integrals in Riemann manifolds, *J. Multivariate Anal.,* **6**: 397–413, 1976.
70. T. E. Duncan, Estimation for jump processes in the tangent bundle of a Riemann manifold, *Appl. Math. Optim.,* **4**: 265–274, 1978.
71. T. E. Duncan, Some filtering results on Riemann manifolds, *Inform. Control,* **35**: 182–195, 1977.
72. T. E. Duncan, An estimation problem in compact Lie groups, *Syst. Control Lett.,* **10**: 257–263, 1988.
73. T. E. Duncan, Stochastic filtering in manifolds, In *Proceedings of the International Federation on Autonomic Control* World Congress, Elmsford, NY: Pergamon Press, 1981.
74. T. E. Duncan, An introduction to stochastic calculus in manifolds with applications, In J. S. Baras and V. Mirelli (eds.), *Recent Advances in Stochastic Calculus,* New York: Springer-Verlag, pp. 105–140, 1990.
75. J. Chen, S. S. Yau, and C. W. Leung, Finite dimensional filters with nonlinear drift. VIII: Classification of finite dimensional estimation algebras of maximal rank with state space dimension 4, preprint.
76. R. Dong, L. Tam, W. Wong, and S. Yau, Structure and classification theorem of finite-dimensional exact estimation algebras, *SIAM J. Control Optim.,* **29**: 866–877, 1991.
77. L. Tam, W. Wong, and S. Yau, On a necessary and sufficient condition for finite dimensionality of estimation algebras, *SIAM J. Control Optim.,* **28**: 173–185, 1990.
78. D. Ocone, Finite dimensional Lie algebras in nonlinear filtering, In M. Hazewinkel and J. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications,* Dordrecht, The Netherlands: D. Riedel, pp. 629–636, 1981.
79. J. Chen and S. S. Yau, Finite dimensional filters with nonlinear drift. VII: Mitter conjecture and structure of η, preprint.
80. J. Baras, Group invariance methods in nonlinear filtering of diffusion processes, In M. Hazewinkel and J. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications,* Dordrecht, The Netherlands: D. Riedel, pp. 565–572, 1981.
81. J. Wei and E. Norman, On the global representation of the solutions of linear differential equations as a product of exponentials, *Proc. Am. Math. Soc.,* **15**: 327–334, 1964.
82. M. Cohen de Lara, Finite-dimensional filters. Part I: The Wei–Norman techniques, *SIAM J. Control Optim.,* **35**: 980–1001, 1997.
83. M. Cohen de Lara, Finite-dimensional filters. Part II: Invariance group techniques, *SIAM J. Control Optim.,* **35**: 1002–1029, 1997.

84. A. E. Bryson and M. Frazier, Smoothing for linear and nonlinear dynamic systems, in *Proceedings of the Optimum System Synthesis Conference,* Wright-Patterson Air Force Base, Ohio, September, 1962, AST-TDR-63-119.

85. R. E. Mortensen, Maximum likelihood recursive nonlinear filtering, *J. Optim. Theory Appl.,* **2**: 386–394), 1968.

86. W. H. Fleming and S. K. Mitter, Optimal control and nonlinear filtering for nondegenerate diffusion processes, *Stochastics,* **8**: 63–77, 1982.

87. W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control,* New York: Springer-Verlag, 1975.

87a. S. K. Mitter, Lectures on nonlinear filtering and stochastic control, *Lecture Notes in Math,* **972**: 170–207, 1983.

87b. M. H. A. Davis, On a multiplicative functional transformation arising in nonlinear filtering, *Z. Wahr. Verw. Geb.,* **54**: 125–139, 1980.

87c. M. H. A. Davis, A pathwise solution of the equations of nonlinear filtering, *Theory Prob. Appl.,* **27**: 167–175, 1982.

87d. O. Hijab, *Minimum energy estimation,* Ph.D. dissertation, Univ. California, Berkeley, CA, 1980.

87e. W. H. Fleming and W. M. McEneaney, Risk sensitive and robust nonlinear filtering, *Proc. 36th IEEE Conf. Decision Cont.,* San Diego, 1088–1093, 1997.

87f. B. Z. Bobrovsky and M. Zakai, *Asymptotic bounds on the minimal error of nonlinear filtering, Stochastic Systems: the mathematics of filtering and identification and applications,* M. Hazeainkel and J. Willems (eds.), Dordrecht: Reidel, 1981, 573–582.

87g. D. Ocone and E. Pardoux, Asymptotic stability of the optimal filter with respect to the initial conditions, *SIAM J. Con. Optim.,* **34**: 226–243, 1996.

88. A. Bensoussan, R. Glowinski, and R. Rascanu, Approximations of the Zakai equation by splitting up method, *SIAM J. Control Optim.,* **28**: 1420–1431, 1990.

89. R. J. Elliott and R. Glowinski, Approximations to solutions of the Zakai filtering equation, *Stochastic Anal. Appl.,* **7**: 145–168, 1989.

90. K. Itô, Approximation of the Zakai equation for nonlinear filtering, *SIAM J. Control Optim.,* **34**: 620–634, 1996.

91. H. P. McKean, Geometry of differential space, *Ann. Prob.,* **1**: 197–206, 1973.

92. S. Lototsky, R. Mikulevicius, and B. L. Rozovskii, Nonlinear filtering revisited: A spectral approach, *SIAM J. Control Optim.,* **35**: 435–461, 1997.

93. A. Budhiraja and G. Kallianpur, Approximation to the solution of the Zakai equation using multiple Wiener and Stratonovich integral expansions, *Stochastics and Stoch. Repts.,* **56**: 271–315, 1996.

94. W.-L. Chiou and S.-T. Yau, Finite-dimensional filters with nonlinear drift. II: Brockett's problem on classification of finite-dimensional estimation algebras, *SIAM J. Control Optim.,* **32**: 297–310, 1994.

TYRONE E. DUNCAN
University of Kansas

FILTERING, LINEAR. See CONVOLUTION.