March 13, 2006

WIENER CHAOS AND NONLINEAR FILTERING

S. V. LOTOTSKY

Abstract. The paper discusses two algorithms for solving the Zakai equation in the time-homogeneous diffusion filtering model with possible correlation between the state process and the observation noise. Both algorithms rely on the Cameron-Martin version of the Wiener chaos expansion, so that the approximate filter is a finite linear combination of the chaos elements generated by the observation process. The coefficients in the expansion depend only on the deterministic dynamics of the state and observation processes. For real-time applications, computing the coefficients in advance improves the performance of the algorithms in comparison with most other existing methods of nonlinear filtering. The paper summarizes the main existing results about these Wiener chaos algorithms and resolves some open questions concerning the convergence of the algorithms in the noise-correlated setting. The presentation includes the necessary background on the Wiener chaos and optimal nonlinear filtering.

To appear in *Applied Mathematics and Optimization, 2006.*

Contents

2000 *Mathematics Subject Classification.* Primary 60H15; Secondary 60G35, 62M20, 93E11.
Key words and phrases. Correlated Noise, Error Estimation, Recursive Algorithm, Zakai Equation.
The author gratefully acknowledges support from the Sloan Research Fellowship, the NSF CAREER award DMS-0237724, and the ARO Grant DAAD19-02-1-0374.

1. Introduction

Let us recall the general mathematical formulation of the filtering problem [27, 28, 44, 62]. Consider two random processes, the STATE $X = X(t)$ and OBSERVATIONS $Y = Y(t), t \geq 0$, both defined on a suitable stochastic basis $\mathbb{F} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Let $f = f(t, X_{0,t}, Y_{0,t})$ be a square integrable measurable functional, depending at time t on the trajectories $X_{0,t}$ and $Y_{0,t}$ of the processes X and Y up to time t. The filtering problem is to find the OPTIMAL FILTER estimate of f, that is, the best mean-square estimate $\widehat{f_t}$ of $f(t, X_{0,t}, Y_{0,t})$ on the basis of the observations $Y(s),\ 0 \leq s \leq t$. From the basic probability theory, this estimate is known to be the conditional expectation

$$\widehat{f_t} = \mathbb{E}\big(f(t, X_{0,t}, Y_{0,t})|\mathcal{F}_t^Y\big), \tag{1.1}$$

where $\mathcal{F}_t^Y$ is the σ-algebra generated by the random variables $Y(s),\ 0 \leq s \leq t$. The subject of the mathematical theory of filtering is finding suitable ways of computing this conditional expectation, either exactly or approximately.

If $f(t, X_{0,t}, Y_{0,t}) = f(X(t))$, then, for a large class of processes X, Y, the optimal filter (1.1) is computable by the formula

$$\widehat{f_t} = \frac{\int_S f(x)u(t,x)dx}{\int_S u(t,x)dx}, \tag{1.2}$$

where S is the phase space of the unobserved process X and $u = u(t, x)$ is a random field called unnormalized filtering density. Moreover, the function u is a unique solution of a linear stochastic parabolic equation driven by the observation process Y; the equation is known as the ZAKAI EQUATION. Below is a rough idea of the filtering algorithms discussed in this paper.

By definition, the unnormalized filtering density u is a function of the time t, space x, and the elementary outcome ω; by convention, the dependence on ω is usually not shown, but always implied. Assume that the filtering problem is considered on a fixed time interval $[0, T]$. Similar to the usual Fourier series expansion, we can write

$$u(t, x, \omega) = \sum_{k,l=1}^{\infty} \varphi_{kl}(t)e_k(x)\xi_l(\omega), \tag{1.3}$$

where $\varphi_{kl},\ k, l \geq 1$, are *deterministic* functions of time, $\{e_k,\ k \geq 1\}$ is an orthonormal basis in $L_2(S)$, and $\{\xi_l,\ l \geq 1\}$, is an orthonormal basis in the suitable space of random variables generated by the observation process $Y(s),\ 0 \leq s \leq T$.

Let $0 = t_0 < t_1 < \ldots < t_n = T$, with $t_k = k\Delta$, be a uniform partition of the interval $[0, T]$. We will see that the uniqueness of the unnormalized filtering density and linearity of the Zakai equation imply the following recursive version of (1.3):

$$u(t_i, x, \omega) = \sum_{k=1}^{\infty} \varphi_k^i(\omega)e_k(x), \qquad \varphi_k^i(\omega) = \sum_{l,n=1}^{\infty} q_k^{ln}\varphi_n^{i-1}(\omega)\xi_l^i(\omega), \tag{1.4}$$

where q_k^{ln} are real numbers and $\{\xi_l^i,\ l \geq 1\}$, is an orthonormal basis in the suitable space of random variables generated by the observation process $Y(s),\ t_{i-1} \leq s \leq t_i$. In most applications, the space S is all or part of $\mathbb{R}^d$, so that there is no difficulty in

selecting the functions e_k. Construction of the basis in the space of random variables generated by the observation process is less straightforward and relies on a theorem of Cameron and Martin [8].

A recursive approximation of u is obtained by keeping only finitely many terms in each of the sums in (1.4); if the filtering model is time-homogeneous, then the numbers q_k^{ln} depend only the time step Δ and the coefficients in equations describing the state and observation processes, and therefore can be computed in advance. We call the corresponding algorithm for computing the approximation of u a *spectral separating scheme*. We will see that there are two fundamentally different ways to truncate the expansions in (1.4), resulting in the spectral separating schemes of the first and second kind.

The structure of the paper is as follows. Sections 2 and 3 contain the background information about nonlinear filtering and Wiener chaos. Section 2 introduces the time-homogeneous diffusion filtering model and outlines the derivation of representation (1.2) for the optimal filter. Section 3 reviews the construction of the Cameron-Martin basis and presents the corresponding expansion for the solution of a linear stochastic evolution equation. Section 4 is the main part of the paper and is mostly based on papers [46] and [47]. In this section, we derive representation (1.4) for the solution of the Zakai equation in the time-homogeneous diffusion filtering model, study the spectral separating schemes, and preform the compare-and-contrast analysis of the results from [46] and [47]. Theorem 4.5, establishing the convergence of the spectral separating scheme of the first kind in the noise-correlated setting, is new. In Section 5 we discuss other filtering models and other approaches to constructing the optimal filter, as well as possible connections with the Wiener chaos method. It appears that the spectral separating schemes have both theoretical and practical interest due to a unique combination of features not found in other works on nonlinear filtering. In particular, (i) these algorithms have a potential for good real-time performance when the dimension of the state process is large, and (ii) the implementation and analysis of the algorithms are not affected by the presence of the observation noise in the state equation.

2. The Diffusion Filtering Model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with independent standard Wiener processes $V = V(t)$ and $W = W(t)$ of dimensions d_1 and r respectively. Let X_0 be a random variable independent of W and V. We consider a time-homogeneous diffusion filtering model, in which the unobserved d - dimensional state, or signal, process $X = X(t)$ and the r-dimensional observation process $Y = Y(t)$ satisfy the Itô stochastic ordinary differential equations

$$\begin{aligned} dX(t) &= b(X(t))dt + \sigma(X(t))dV(t) + \rho(X(t))dW(t), \\ dY(t) &= h(X(t))dt + dW(t),\ 0 < t \le T; \\ X(0) &= X_0,\ Y(0) = 0, \end{aligned} \tag{2.1}$$

where $b(x) \in \mathbb{R}^d$, $\sigma(x) \in \mathbb{R}^{d\times d_1}$, $\rho(x) \in \mathbb{R}^{d\times r}$, $h(x) \in \mathbb{R}^r$. We call (2.1) *time-homogeneous* because the functions b, σ, ρ, and h do not depend on time. A more general filtering model is discussed in Section 5.

Assumption R1. The components of the matrix functions σ and ρ are $\mathbf{C}_b^3(\mathbb{R}^d)$, that is, bounded and three times continuously differentiable on $\mathbb{R}^d$ so that all the derivatives are also bounded; the components of the functions b and h are $\mathbf{C}_b^2(\mathbb{R}^d)$, and the distribution of the random variable X_0 has a density p_0 with respect to the Lebesgue measure on $\mathbb{R}^d$.

Under Assumption **R1** system (2.1) has a unique strong solution [31, Theorems 5.2.5 and 5.2.9].

Let $f = f(x)$ be a scalar measurable function on $\mathbb{R}^d$ so that $\sup_{0\le t\le T}\mathbb{E}|f(X(t))|^2 < \infty$. In what follows, we show that the optimal filter

$$\widehat{f}_t = \mathbb{E}\left(f(X(t))|\mathcal{F}_t^Y\right)$$

has the representation (1.2) and derive the equation for the function u.

Define a new probability measure $\widetilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ as follows: for $A \in \mathcal{F}$,

$$\widetilde{\mathbb{P}}(A) = \int_A Z_T^{-1} d\mathbb{P},$$

where

$$Z_t = \exp\left(\int_0^t h^*(X(s))dY(s) - \frac{1}{2}\int_0^t |h(X(s))|^2 ds\right).$$

Here and below, if $\zeta \in \mathbb{R}^k$, then ζ is a *column* vector, $\zeta^* = (\zeta_1, \dots, \zeta_k)$, and $|\zeta|^2 = \zeta^*\zeta$); the adjoint of a matrix A is A^*.

The measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent because the function h is bounded. The expectation with respect to the measure $\widetilde{\mathbb{P}}$ will be denoted by $\widetilde{\mathbb{E}}$.

Theorem 2.1. *The measure $\widetilde{\mathbb{P}}$ has the following properties:*

- **P1.** *Under the measure $\widetilde{\mathbb{P}}$, the distributions of the Wiener process V and the random variable X_0 are unchanged, the observation process Y is a standard Wiener process, and the state process X satisfies*

$$\begin{aligned} dX(t) &= b(X(t))dt + \sigma(X(t))dV(t) \\ &\quad + \rho(X(t))\left(dY(t) - h(X(t))dt\right),\ 0 < t \le T; \\ X(0) &= X_0; \end{aligned} \tag{2.2}$$

- **P2.** *Under the measure $\widetilde{\mathbb{P}}$, the Wiener processes W and Y and the random variable X_0 are independent of one another;*
- **P3.** *The optimal filter $\widehat{f}_t$ satisfies*

$$\widehat{f}_t = \frac{\widetilde{\mathbb{E}}\left(f(X(t))Z_t|\mathcal{F}_t^Y\right)}{\widetilde{\mathbb{E}}(Z_t|\mathcal{F}_t^Y)}. \tag{2.3}$$

P4 *If $\xi \in L_2(\Omega, \widetilde{\mathbb{P}})$, then $\xi \in L_1(\Omega, \mathbb{P})$ and*

$$\mathbb{E}|\xi| \le C\sqrt{\widetilde{\mathbb{E}}\xi^2}, \tag{2.4}$$

where C depends only on T and $\sup_{x\in\mathbb{R}^d} |h(x)|$.

Proof. Properties **P1** and **P2** follow from the Girsanov Theorem; notice that equation (2.2) is equivalent to the first equation in (2.1) because $dW = dY - hdt$. Property **P3** is a particular case of the Bayes formula [59, Lemma 8.6.2]. Finally, **P4** follows from the Cauchy-Schwartz inequality:

$$\mathbb{E}|\xi| = \widetilde{\mathbb{E}}(Z_T\xi) \le \sqrt{\widetilde{\mathbb{E}}Z_T^2}\sqrt{\widetilde{\mathbb{E}}\xi^2} \le C\sqrt{\widetilde{\mathbb{E}}\xi^2}.$$

The details of the arguments can be found in [28, 44, 62]. □

Next, consider the partial differential operators

$$\mathcal{L}g(x) = \frac{1}{2}\sum_{i,j=1}^{d} \left((\sigma(x)\sigma^*(x))_{ij} + (\rho(x)\rho^*(x))_{ij}\right) \frac{\partial^2 g(x)}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x)\frac{\partial g(x)}{\partial x_i};$$

$$\mathcal{M}_l g(x) = h_l(x)g(x) + \sum_{i=1}^{d} \rho_{il}(x)\frac{\partial g(x)}{\partial x_i},\ l = 1, \dots, r;$$

and their adjoints

$$\mathcal{L}^* g(x) = \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}\left((\sigma(x)\sigma^*(x))_{ij} g(x) + (\rho(x)\rho^*(x))_{ij} g(x)\right) - \sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(b_i(x)g(x)\right);$$

$$\mathcal{M}_l^* g(x) = h_l(x)g(x) - \sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(\rho_{il}(x)g(x)\right),\ l = 1, \dots, r.$$

Let $\mathbf{H}^\gamma,\ \gamma \in \mathbb{R}$, be the Sobolev space with the norm

$$\|f\|_\gamma = \left(\int_{\mathbb{R}^d} (1 + |y|^2)^{\gamma/2} |\breve{f}(y)|^2 dy\right)^{1/2},$$

where $\breve{f} = \breve{f}(y)$ is the Fourier transform of f; $\mathbf{H}^0 = L_2(\mathbb{R}^d)$. Both the inner product in $L_2(\mathbb{R}^d)$ and the duality between $\mathbf{H}^1$ and $\mathbf{H}^{-1}$ relative to $L_2(\mathbb{R}^d)$ will be denoted by $(\cdot, \cdot)_0$.

Note that the operators $\mathcal{L}, \mathcal{L}^*$ are bounded from $\mathbf{H}^1$ to $\mathbf{H}^{-1}$ and operators $\mathcal{M}, \mathcal{M}^*$ are bounded from $\mathbf{H}^1$ to $L_2(\mathbb{R}^d)$. Moreover, direct calculations show that, under Assumption **R1**, there exists a number $C > 0$ so that, for every $g \in \mathbf{H}^1$,

$$2(\mathcal{L}^* g, g)_0 + \sum_{l=1}^{r} \|\mathcal{M}_l^* g\|_0^2 \le C\|g\|_0^2. \tag{2.5}$$

If the matrix $\sigma\sigma^*$ is uniformly positive definite, that is, there exists a $\delta > 0$ so that, for all $x, y \in \mathbb{R}^d$,

$$\sum_{i,j=1}^{d} \sum_{k=1}^{d_1} \sigma_{ik}(x)\sigma_{jk}(x) y_i y_j \ge \delta|y|^2 \tag{2.6}$$

then a stronger version of (2.5) holds:

$$2(\mathcal{L}^*g,g)_0+\sum_{l=1}^r\|\mathcal{M}_l^*g\|_0^2\le-\delta\|g\|_1^2+C\|g\|_0^2. \tag{2.7}$$

Theorem 2.2. *In addition to Assumption R1 suppose that the initial density p_0 belongs to the space $\mathbf{H}^1$. Then there exists a unique $\mathcal{F}_t^Y$-adapted random field $u=u(t,x),\ t\in[0,T],\ x\in\mathbb{R}^d$, with the following properties:*

U1 *The function $u(t,x)$ is an element of $L_2(\Omega,\widetilde{\mathbb{P}};\mathbf{C}([0,T],\mathbf{H}^1))$ and, for every $v\in\mathbf{H}^1$, satisfies the equality*

$$(u,v)_0(t)=(p_0,v)_0+\int_0^t(u,\mathcal{L}v)_0(s)ds+\int_0^t\sum_{l=1}^r(u,\mathcal{M}_lv)_0(s)dY_l(s)$$

on the same set of $\widetilde{\mathbb{P}}$-measure one for all $0\le t\le T$; see Section 4.2 in [62] *for details. In other words, $u=u(t,x)$ is the unique generalized solution of the stochastic partial differential equation*

$$\begin{aligned}du(t,x)&=\mathcal{L}^*u(t,x)dt+\sum_{l=1}^r\mathcal{M}_l^*u(t,x)dY_l(t),\ 0<t\le T,\ x\in\mathbb{R}^d;\\ u(0,x)&=p_0(x).\end{aligned} \tag{2.8}$$

U2 *The equality*

$$\widetilde{\mathbb{E}}\left(f(X(t))Z_t|\mathcal{F}_t^Y\right)=\int_{\mathbb{R}^d}f(x)u(t,x)dx \tag{2.9}$$

holds for all bounded measurable functions f.

Proof. Existence, uniqueness, and regularity of an $\mathcal{F}_t^Y$-adapted generalized solution of equation (2.8) follow from Theorem 4.2.1 in [62]. Representation (2.9) follows from Theorem 5.3.1 in [62]. □

The random field $u=u(t,x)$ is called the UNNORMALIZED FILTERING DENSITY and the random variable $\phi_t[f]=\widetilde{\mathbb{E}}\left[f(X(t))Z_t|\mathcal{F}_t^Y\right]$, the UNNORMALIZED OPTIMAL FILTER.

If condition (2.6) holds, then the unique solvability of (2.8) holds under weaker regularity conditions on the coefficients and initial condition; see Theorem 4.1.1 in [62].

3. The Wiener Chaos

In this section, we review the construction of the Cameron-Martin basis in the space of square integrable functionals of the Wiener process. For a linear *stochastic* evolution equation, we show that the coefficients of the Cameron-Martin expansion of the solution satisfy a lower-triangular system of linear *deterministic* evolution equations.

For a fixed $T>0$, let $\mathbb{F}=(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{0\le t\le T},\mathbb{P})$ be a stochastic basis with the usual assumptions and $W=(W_1(t),\dots,W_r(t)),\ 0\le t\le T$, an r-dimensional Wiener process on $\mathbb{F}$. Denote by $\mathcal{F}_t^W$ the sigma-algebra generated by the random variables

$W_k(s),\ k=1,\dots,r,\ s\le t$, and by $L_2(\mathbb{W})$, the Hilbert space of $\mathcal{F}_T^W$-measurable square integrable random variables. The Cameron-Martin basis is a special orthonormal basis in the space $L_2(\mathbb{W})$.

Let $\mathfrak{m}=\{m_k,\ k\ge 1\}$ be an orthonormal basis in $L_2((0,T))$ and define independent standard Gaussian random variables

$$\xi_{ik}=\int_0^T m_i(s)dW_k(s).$$

Consider the set of multi-indices

$$\mathcal{J}=\Big\{\alpha=(\alpha_i^k,\ i\ge 1,\ k=1,\dots,r),\ \ \alpha_i^k\in\{0,1,2,\dots\},\ \ \sum_{i,k}\alpha_i^k<\infty\Big\}. \tag{3.1}$$

The set $\mathcal{J}$ is countable, and, for every $\alpha\in\mathcal{J}$, only finitely many of α_i^k are not equal to zero. For $\alpha\in\mathcal{J}$, we write

$$|\alpha|=\sum_{i,k}\alpha_i^k,\ \ \alpha!=\prod_{i,k}\alpha_i^k!,$$

and define the collection $\Xi=\{\xi_\alpha,\ \alpha\in\mathcal{J}\}$ of random variables so that

$$\xi_\alpha=\frac{1}{\sqrt{\alpha!}}\prod_{i,k}H_{\alpha_i^k}(\xi_{ik}), \tag{3.2}$$

where

$$H_n(t)=(-1)^n e^{t^2/2}\frac{d^n}{dt^n}e^{-t^2/2} \tag{3.3}$$

is n-th Hermite polynomial. Recall that $H_0(t)=1$ and $H_1(t)=t$. The N-TH WIENER CHAOS is the linear span of the random variables $\xi_\alpha,\ |\alpha|=N$.

As an example, let $r=2$ and consider the multi-index

$$\alpha=\begin{pmatrix}0&1&0&3&0&0&\cdots\\2&0&0&0&4&0&\cdots\end{pmatrix}$$

with four non-zero entries $\alpha_2^1=1;\ \alpha_4^1=3;\ \alpha_1^2=2;\ \alpha_5^2=4$. Then the corresponding basis element

$$\xi_\alpha=\xi_{2,1}\cdot\frac{H_3(\xi_{4,1})}{\sqrt{3!}}\cdot\frac{H_2(\xi_{1,2})}{\sqrt{2!}}\cdot\frac{H_4(\xi_{5,2})}{\sqrt{4!}}.$$

Theorem 3.1. *The collection $\Xi=\{\xi_\alpha,\ \alpha\in\mathcal{J}\}$ is an orthonormal basis in $L_2(\mathbb{W})$.*

Proof. This is a version of the classical result due to Cameron and Martin [8]. □

If $\eta\in L_2(\mathbb{W})$, then, by Theorem 3.1,

$$\eta=\sum_{\alpha\in\mathcal{J}}\mathbb{E}\big(\eta\xi_\alpha\big)\xi_\alpha, \tag{3.4}$$

and

$$\mathbb{E}|\eta|^2=\sum_{\alpha\in\mathcal{J}}|\mathbb{E}\big(\eta\xi_\alpha\big)|^2. \tag{3.5}$$

Representation (3.4) is an example of the Wiener chaos expansion, and is known as the CAMERON-MARTIN EXPANSION OF η.

Recall that, beside constructing the Cameron-Martin basis Ξ, another goal of this section is to derive the corresponding expansion for the solution of a linear stochastic evolution equation. To achieve this goal, we start with the following technical result.

Lemma 3.2. *Define* $\xi_\alpha(t)$ *by*

$$\xi_\alpha(t) = \mathbb{E}(\xi_\alpha|\mathcal{F}_t^W). \tag{3.6}$$

Then

$$\xi_\alpha(t) = I(|\alpha| = 0) + \int_0^t \sum_{i,k} \sqrt{\alpha_i^k}\, \xi_{\alpha^-(i,k)}(s) m_i(s) dW_k(s), \tag{3.7}$$

where

$$I(|\alpha| = 0) = \begin{cases} 1, & |\alpha| = 0, \\ 0, & \text{otherwise}, \end{cases}$$

and $\alpha^-(i,k)$ *is the multi-index with components*

$$\Big(\alpha^-(i,k)\Big)_j^l = \begin{cases} \max(\alpha_i^k - 1, 0), & \text{if } i = j \text{ and } k = l, \\ \alpha_j^l, & \text{otherwise.} \end{cases} \tag{3.8}$$

Proof. Let $h = (h_1(t), \dots, h_r(t))$ be an r-vector so that each h_k is a finite linear combination of the elements of $\mathfrak{m}$. Define

$$\mathcal{E}(t,h) = \exp\left(\sum_{k=1}^r \left(\int_0^t h_k(s) dW_k(s) - \frac{1}{2}\int_0^t |h_k(s)|^2 ds\right)\right). \tag{3.9}$$

We also introduce the notations

$$\mathcal{E}(h) = \mathcal{E}(T,h); \quad h_{ik} = \int_0^T h_k(t) m_i(t) dt,\ m_i \in \mathfrak{m}; \quad h^\alpha = \prod_{i,k} h_{ik}^{\alpha_i^k},\ \alpha \in \mathcal{J}.$$

The following properties of the process $\mathcal{E}(t,h)$ are verified by direct calculation:

$$\mathcal{E}(h) = \sum_{\alpha\in\mathcal{J}} \frac{h^\alpha}{\sqrt{\alpha!}}, \tag{3.10}$$

$$\mathcal{E}(t,h) = 1 + \int_0^t \mathcal{E}(s,h) \sum_{i,k} h_{ik} m_i(s) dW_k(s). \tag{3.11}$$

In particular,

$$\mathbb{E}\big(\mathcal{E}(h)|\mathcal{F}_t^W\big) = \mathcal{E}(t,h). \tag{3.12}$$

The main consequence of (3.10) is an alternative representation of ξ_α:

$$\xi_\alpha = \frac{1}{\sqrt{\alpha!}} \frac{\partial^{|\alpha|}}{\partial h^\alpha} \mathcal{E}(h)\bigg|_{h=0}. \tag{3.13}$$

Taking the conditional expectation, using the martingale property (3.12) of $\mathcal{E}(t,h)$, and exchanging the expectation and differentiation operations (for example, by Theorem 3.3.3(iii) in [38]), we conclude that

$$\xi_\alpha(t) = \frac{1}{\sqrt{\alpha!}} \frac{\partial^{|\alpha|}}{\partial h^\alpha} \mathcal{E}(t,h)\bigg|_{h=0}. \tag{3.14}$$

Equality (3.6) now follows from (3.11). □

We are now ready to derive the Cameron-Martin expansion for the solution of a linear stochastic evolution equation.

Theorem 3.3. *Let $U = U(t)$, $0 \le t \le T$, be a square integrable, $\mathcal{F}_t^W$ – adapted random process with values in a Hilbert space X. Denote by $(\cdot,\cdot)_X$ and $\|\cdot\|_X$ the inner product and the norm in X. Assume that*

$$U(t) = U(0) + \int_0^t \mathcal{A}(s)U(s)ds + \sum_{k=1}^r \int_0^t \mathcal{B}_k(s)U(s)dW_k(s), \tag{3.15}$$

where the initial condition $U(0)$ is independent of $\mathcal{F}_T^W$ and the linear operators $\mathcal{A}(s)$, $\mathcal{B}_k(s)$ are non-random. Then $U(t) = \sum_{\alpha\in\mathcal{J}} U_\alpha(t)\xi_\alpha$, and

$$U_\alpha(t) = U_0 I(|\alpha| = 0) + \int_0^t \mathcal{A}(s)U_\alpha(s)ds + \sum_{i,k} \sqrt{\alpha_i^k}\, \mathcal{B}_k(s)U_{\alpha^-(i,k)}(s)m_i(s)ds, \tag{3.16}$$

with multi-index $\alpha^-(i,k)$ defined in (3.8). If, in addition, the operator $\mathcal{A}$ generates a semi-group $\Phi_{t,s}$, $t \ge s \ge 0$, that is, a family of continuous operator on X so that, for every $u, v \in H$, $(\Phi_{t,s}v, u)_X = (v,u)_X + \int_s^t \big(\mathcal{A}(\tau)\Phi_{\tau,s}v, u\big)_X d\tau$, then

$$\sum_{|\alpha|=n} \|U_\alpha(t)\|_X^2 = \sum_{k_1,\dots,k_n=1}^r \int_0^t \int_0^{s_n} \cdots \int_0^{s_2} \tag{3.17}$$
$$\|\Phi_{t,s_n}\mathcal{B}_{k_n}(s_n)\cdots\Phi_{s_2,s_1}\mathcal{B}_{k_1}(s_1)\Phi_{s_1,0}U_0(s_1)\|_X^2 ds_1\cdots ds_n.$$

Proof. By Theorem 3.1 we have $U_\alpha(t) = \mathbb{E}\big(U(t)\xi_\alpha\big)$ and then $\mathcal{F}_t^W$-measurability of $U(t)$ implies

$$\mathbb{E}\big(U(t)\xi_\alpha\big) = \mathbb{E}\Big(U(t)\mathbb{E}(\xi_\alpha|\mathcal{F}_t^W)\Big) = \mathbb{E}\big(U(t)\xi_\alpha(t)\big).$$

To derive (3.16), it remains to apply the Itô formula to the product $U(t)\xi_\alpha(t)$ using equation (3.7).

To prove (3.17), iterate (3.15) and write the random variable $U(t)$ as

$$U(t) = U(0) + \sum_{n\ge1}\sum_{k_1,\dots,k_n=1}^r \int_0^t \int_0^{s_n} \cdots \int_0^{s_2} \tag{3.18}$$
$$\Phi_{t,s_n}\mathcal{B}_{k_n}(s_n)\cdots\Phi_{s_2,s_1}\mathcal{B}_{k_1}(s_1)\Phi_{s_1,0}U_0(s_1)dW_{k_1}(s_1)\cdots dW_{k_n}(s_n).$$

Then the relation between the multiple Itô integrals and Hermite polynomials [24, Theorem 3.1] implies the equality

$$\sum_{|\alpha|=n} U_\alpha(t)\xi_\alpha = \sum_{k_1,\dots,k_n=1}^r \int_0^t \int_0^{s_n} \cdots \int_0^{s_2} \tag{3.19}$$
$$\Phi_{t,s_n}\mathcal{B}_{k_n}(s_n)\cdots\Phi_{s_2,s_1}\mathcal{B}_{k_1}(s_1)\Phi_{s_1,0}U_0(s_1)dW_{k_1}(s_1)\cdots dW_{k_n}(s_n),$$

and (3.17) follows. □

Existence of $\Phi_{t,s}$ usually follows from the existence and uniqueness of a solution of (3.15); see, for example, Section 3.1 in [62] for the list of the corresponding sufficient conditions on the operators $\mathcal{A}$ and $\mathcal{B}_k$.

Given the semi-group $\Phi_{t,s}$, the system of equations (3.16) is solvable by induction: definition of $\alpha^-(i,k)$ implies that, for $n \ge 1$, the coefficients U_α with $|\alpha| = n-1$ determine the coefficients U_α with $|\alpha| = n$. There is only one $U_\alpha = U_{(0)}$ corresponding to $|\alpha| = 0$, and $U_{(0)}(t) = \Phi_{t,0}U(0)$. For $|\alpha| = 1$, with $\alpha_i^k = 1$, we have

$$U_\alpha(t) = \int_0^t \mathcal{A}(s)U_\alpha(s)ds + \int_0^t \mathcal{B}_k(s)U_{(0)}(s)m_i(s)ds,$$

so that $U_\alpha = \int_0^t \Phi_{t,s}\mathcal{B}_k(s)\Phi_{s,0}U(0)\, m_i(s)ds$. Similarly, an explicit formula for all other U_α can be derived by induction; see [47, Proposition A1].

Representation (3.18) is an alternative form of the Wiener chaos expansion and uses multiple Itô integrals rather than the Cameron-Martin basis (3.2). Multiple integral expansion is more explicit than the Cameron-Martin expansion, and has been widely used in the study of stochastic equations. For example, Kunita [37] used it to sharpen existence and uniqueness results for the Zakai equation and Krylov and Veretennikov [35], to establish a new criterion for existence and uniqueness of strong solutions for stochastic *ordinary* differential equations. A clear advantage of the Cameron-Martin version (3.4) is computational flexibility: numerically, it is much easier to work with the system of equations (3.16) than with the integrands in (3.18).

4. Spectral Separating Schemes

In this section, we describe two algorithms for solving the Zakai equation (2.8) using the Cameron-Martin expansion of the solution. Consider the filtering model (2.1). By Theorem 2.2, given a suitable function $f = f(x)$, the optimal filter $\widehat{f_t} = \mathbb{E}\big(f(X(t))|\mathcal{F}_t^Y\big)$ has the representation $\widehat{f_t} = \phi_t[f]/\phi_t[1]$, where $\phi[f] = \int_{\mathbb{R}^d} f(x)u(t,x)dx$ and u is the solution of the Zakai equation (2.8). Using Theorem 3.3, we will implement the idea described in the Introduction and construct recursive approximations of $u(t,x)$ and $\varphi_t[f]$, as well as study the quality of the approximations.

Let $0 = t_0 < t_1 < \ldots < t_M = T$ be a uniform partition of the interval $[0,T]$ with step Δ so that $t_i = i\Delta,\ i = 0,\ldots,M$. Fix an orthonormal basis $\mathfrak{m} = \{m_k(s),\, k \ge 1\}$ in $L_2([0,\Delta])$ and define random variables

$$\xi_\alpha^i = \prod_{k,l}\left(\frac{H_{\alpha_k^l}(\xi_{k,l}^i)}{\sqrt{\alpha_k^l!}}\right), \quad \alpha \in \mathcal{J}, \tag{4.1}$$

where $\mathcal{J}$ is the set of multi-indices (3.1), $\xi_{k,l}^i = \int_{t_{i-1}}^{t_i} m_k(s - t_{i-1})dY_l(s)$, and H_n is the n-th Hermite polynomial (3.3). We also fix an orthonormal basis $\mathfrak{e} = \{e_k(x),\, k \ge 1\}$ in $L_2(\mathbb{R}^d)$.

4.1. Spectral Separating Scheme of the First Kind.

Consider the following system of equations:

$$\begin{aligned}
&\frac{\partial\varphi_\alpha(s,x,g)}{\partial s} = \mathcal{L}^*\varphi_\alpha(s,x,g) + \sum_{k,l}\sqrt{\alpha_k^l}m_k(s)\mathcal{M}_l^*\varphi_{\alpha^-(k,l)}(s,x,g), \quad 0 < s \le \Delta \\
&\varphi_\alpha(0,x,g) = g(x)1_{\{|\alpha|=0\}}.
\end{aligned} \tag{4.2}$$

Define the numbers

$$q^l_{\alpha k} = (\varphi_\alpha(\Delta, \cdot, e_k), e_l)_0, \tag{4.3}$$

and then by induction

$$\psi_j(0) = (p_0, e_j)_0; \ \psi_j(i) = \sum_{\alpha\in\mathcal{J}} \sum_{k=1}^{\infty} \psi_k(i-1) q^j_{\alpha k} \xi^i_\alpha. \tag{4.4}$$

Theorem 4.1. *The unnormalized filtering density has the representation*

$$u(t_i, x) = \sum_{j=1}^{\infty} \psi_j(i) e_j(x), \ 0 \le i \le M, \tag{4.5}$$

and, for $f \in L_2(\mathbb{R}^2)$, the unnormalized optimal filter $\phi_t[f]$ has the representation

$$\phi_{t_i}[f] = \sum_{j=1}^{\infty} \psi_j(i) f_j, \ \text{where } f_k = \int_{\mathbb{R}^d} f(x) e_k(x) dx. \tag{4.6}$$

Proof. Recall that u is a square integrable $\mathcal{F}^Y_t$-adapted solution of the Zakai equation (2.8), and Y is an r-dimensional Wiener process under measure $\widetilde{\mathbb{P}}$. By Theorem 3.3 we have $u(t,x) = \varphi_\alpha(t, x, u(t_{i-1}, x))\xi^i_\alpha$, $t \in [t_{i-1}, t_i]$. To establish (4.5), it remains to write $u(t_{i-1}, x) = \sum_{k\ge1}(u(t_{i-1}, \cdot), e_k)_0 e_k(x)$ and use linearity of equations (2.8) and (4.2). Equality (4.6) is a direct consequence of (4.5), because $\phi_t[f] = \int_{\mathbb{R}^d} f(x)u(t,x)dx$. $\square$

Even when $f \notin L_2(\mathbb{R}^d)$, the integrals $\int_{\mathbb{R}^d} f(x)e_k(x)dx$ may still be defined for all $k \ge 1$, and, under some additional assumptions, representation (4.6) of the unnormalized optimal filter will still hold; see Theorems 4.4, 4.5, and 4.6 below.

Next, we use Theorem 4.1 to construct recursive approximations of $u(t_i, x)$ and $\phi_{t_i}[f]$. for all $j \ge 1$. Let K, n, N be positive integers, and $\mathcal{J}^n_N$, the collection of those multi-indices in $\mathcal{J}$ for which $|\alpha| \le N$ and $\alpha^l_k = 0$ if $k > n$; note that $\mathcal{J}^n_N$ is a *finite* set. With the numbers $q^l_{\alpha k}$ from (4.3), we define $\psi^K_l(i, N, n)$ by truncating the sums in (4.4):

$$\begin{aligned} &\psi^K_l(0, N, n) = (p_0, e_l)_0, \ l = 1, \dots, K; \\ &\psi^K_l(i, N, n) = \sum_{\alpha\in\mathcal{J}^n_N} \sum_{k=1}^{K} \psi^K_k(i-1, N, n) q^l_{\alpha k} \xi^i_\alpha, \end{aligned} \tag{4.7}$$

and then define the approximations of $u(t_i, x)$ and $\phi_{t_i}[f]$ by

$$u^n_{N,K}(t_i, x) = \sum_{j=1}^{K} \psi^K_j(i, N, n) e_j(x), \quad \widetilde{\phi}_i[f] = \sum_{j=1}^{K} \psi^K_j(i, N, n) f_j, \quad 0 \le i \le M, \tag{4.8}$$

where we assume that $f_j = \int_{\mathbb{R}^d} f(x)e_j(x)dx$ exists for each $j \ge 1$. The following is an algorithm for computing the approximations of the unnormlized filtering density and filter using (4.8).

1. *Preliminary computations (before the observations are available)*

   (1) *Choose suitable basis functions $\{e_k, k = 1, \dots, K\}$ in $L_2(\mathbb{R}^d)$ and $\{m_i, i = 1, \dots, n\}$ in $L_2([0, \Delta])$.*

(2) *For $\alpha \in J_N^n$ and $k,l = 1,\dots,K$ compute $q^l_{\alpha k} = (\varphi_\alpha(\Delta,\cdot,e_k),e_l)_0$* (using (4.2)),

$$f_k = \int_{\mathbb{R}^d} f(x)e_k(x)dx, \quad \psi^K_l(0,N,n) = \int_{\mathbb{R}^d} p_0(x)e_k(x)dx.$$

2. Real − time computations, i − th step (as the observations become available): compute ξ^i_α according to (4.1) and update the coefficients ψ:

$$\psi^K_l(i,N,n) = \sum_{\alpha\in J_N^n}\sum_{k=1}^K \psi^K_k(i-1,N,n)q^l_{\alpha k}\xi^i_\alpha \quad l = 1,\dots,K;$$

then, if necessary, compute

$$u^{n,K}_N(t_i,x) = \sum_{l=1}^K \psi^K_l(i,N,n)e_l(x) \tag{4.9}$$

and/or

$$\widetilde{\phi}_i[f] = \sum_{j=1}^K \psi^K_j(i,N,n)f_j, \quad \widetilde{f}_i = \frac{\widetilde{\phi}_i[f]}{\widetilde{\phi}_i[1]}. \tag{4.10}$$

We call this algorithm the Spectral Separating Scheme of the First Kind. It was first suggested in [51] and analyzed in [47] for the filtering model (2.1) with $\rho \equiv 0$. Analysis for arbitrary ρ became possible after the recent work [49]. We discuss the properties of this algorithm in Section 4.3 and analyze its convergence in Section 4.4.

4.2. **Spectral Separating Scheme of the Second Kind.** We now present an alternative algorithm for solving the Zakai equation (2.8) in the filtering model (2.1). In the spectral separating scheme of the first kind, the truncation of the expansion in $L_2(\mathbb{R}^d)$ is done after the truncation of the Cameron-Martin expansion. Now, we will do the truncation in $L_2(\mathbb{R}^d)$ first.

Let $\mathfrak{e}$ be an orthonormal basis in $L_2(\mathbb{R}^d)$ so that every function $e_k = e_k(x)$ belongs to $\mathbf{H}^1$.

Fix a positive integer number K. Define the matrices $A^K = (A^K_{ij},\ i,j = 1,\dots,K)$ and $B^K_l = (B^K_{l,ij},\ i,j = 1,\dots,K;\ l = 1,\dots,r)$, by

$$A^K_{ij} = (\mathcal{L}^*e_j,e_i)_0, \quad B^K_{l,ij} = (\mathcal{M}^*_l e_j,e_i)_0,$$

and consider the Galerkin approximation $u^K(t,x)$ of $u(t,x)$:

$$u^K(t,x) = \sum_{i=1}^K u^K_i(t)e_i(x), \tag{4.11}$$

where the vector $u^K(t) = \{u^K_i(t),\ i = 1,\dots,K\}$ is the solution of the system of stochastic ordinary differential equations

$$du^K(t) = A^K u^K(t)dt + \sum_{l=1}^r B^K_l u^K(t)dY_l(t) \tag{4.12}$$

with the initial condition $u_i^K(0) = (p_0, e_i)_0$. Note that the matrices $B_l^K,\ l = 1, \ldots, r$, do not, in general, commute with each other even if $\rho(x) \equiv 0$, and so system (4.12) must be solved numerically because there is no closed-form solution.

Define random variables ξ_α^i according to (4.1). Theorem 3.3 implies the following result.

Theorem 4.2. *For every $i = 1, \ldots, M$, the solution of (4.12) can be written in $L_2(\Omega, \widetilde{\mathbb{P}}; \mathbb{R}^K)$ as*

$$u^K(t_i) = \sum_{\alpha \in \mathcal{J}} \varphi_\alpha^K(\Delta; u^K(t_{i-1}))\xi_\alpha^i, \quad i = 1, \ldots, M, \tag{4.13}$$

where, for $s \in (0, \Delta]$ and $\zeta \in \mathbb{R}^K$, the functions $\varphi_\alpha^K(s; \zeta)$ are the solutions of

$$\begin{aligned} \frac{\partial \varphi_\alpha^K(s;\zeta)}{\partial s} &= A^K \varphi_\alpha^K(s;\zeta) + \sum_{k,l} \sqrt{\alpha_k^l}\, m_k(s) B_l^K \varphi_{\alpha^-(k,l)}^K(s;\zeta),\ 0 < s \le \Delta, \\ \varphi_\alpha^K(0;\zeta) &= \zeta 1_{\{|\alpha|=0\}}, \end{aligned} \tag{4.14}$$

and $\alpha^-(i,j)$ stands for the multi-index (3.8).

To construct a recursive approximation of u^K, fix positive integers N and n and define the set $\mathcal{J}_N^n$ as the collection of multi-indices α from $\mathcal{J}$ such that $|\alpha| \le N$ and $\alpha_k^l = 0$ if $k > n$. The approximation $u_N^{K,n}(t_i)$ of $u^K(t_i)$ is defined by

$$u_N^{K,n}(t_0) = u^K(0), \quad u_N^{K,n}(t_i) = \sum_{\alpha \in \mathcal{J}_N^n} \varphi_\alpha^K(\Delta; u_N^{K,n}(t_{i-1}))\xi_\alpha^i, \quad i = 1, \ldots, M. \tag{4.15}$$

To establish a representation of $u_N^{K,n}(t_i)$ similar to (4.9), note that $u_N^{K,n}(t_i)$ is a vector in $\mathbb{R}^K$. Let $\mathfrak{U} = \{\mathbf{u}^j,\ j = 1, \ldots, K\}$ be a basis in $\mathbb{R}^K$. The vector $u_N^{K,n}(t_i)$ can then be written as

$$u_N^{K,n}(t_i) = \sum_{j=1}^K u_{N,j}^{K,n}(t_i; \mathfrak{U})\mathbf{u}^j,$$

and by the recursive definition of $u_N^{K,n}(t_i)$,

$$\begin{aligned} u_N^{K,n}(t_i) &= \sum_{\alpha \in \mathcal{J}_N^n} \varphi_\alpha^K(\Delta; u_N^{K,n}(t_{i-1}))\xi_\alpha^i \\ &= \sum_{\alpha \in \mathcal{J}_N^n} \sum_{j=1}^K \varphi_\alpha^K(\Delta; \mathbf{u}^j) u_{N,j}^{K,n}(t_{i-1}; \mathfrak{U})\xi_\alpha^i. \end{aligned}$$

Once again, $\varphi_\alpha^K(\Delta, \mathbf{u}^i)$ is a vector in $\mathbb{R}^K$, so we write

$$\varphi_\alpha^K(\Delta, \mathbf{u}^j) = \sum_{k=1}^K q_{jk}^{K,\alpha}(\mathfrak{U})\mathbf{u}^k, \tag{4.16}$$

and conclude that

$$u_{N,j}^{K,n}(t_i; \mathfrak{U}) = \sum_{\alpha \in \mathcal{J}_N^n} \sum_{k=1}^K q_{jk}^{K,\alpha}(\mathfrak{U}) u_{N,k}^{K,n}(t_{i-1}; \mathfrak{U})\xi_\alpha^i. \tag{4.17}$$

If $f_k = \int_{\mathbb{R}^d} f(x)e_k(x)dx$ is defined for all $k \geq 1$, then

$$u_N^{K,n}(t_i, x) = \sum_{j,k=1}^{K} u_{N,j}^{K,n}(t_i; \mathfrak{U})\mathbf{u}_k^j e_k(x), \quad \widetilde{\phi}_i[f] = \sum_{j,k=1}^{K} u_{N,j}^{K,n}(t_i; \mathfrak{U})\mathbf{u}_k^j f_k \tag{4.18}$$

are the approximations of the unnormalized filtering density and filter.

The following is an algorithm for computing the approximations of the unnormlized filtering density and filter using (4.18).

1. Preliminary computations (before the observations are available):

(1) *Choose suitable basis functions* $\{e_k,\, k = 1, \ldots, K\}$ *in* $L_2(\mathbb{R}^d)$, $\{m_k,\, k = 1, \ldots, n\}$ *in* $L_2([0, \Delta])$, *and a* **standard unit basis** $\{\mathbf{u}^j, j = 1, \ldots, K\}$ *in* $\mathbb{R}^K$, *that is,* $\mathbf{u}_j^j = 1$, $\mathbf{u}_l^j = 0$ *otherwise.*

(2) *For* $\alpha \in \mathcal{J}_N^n$ *and* $j, k = 1, \ldots, K$, *compute* $q_{jk}^{K,\alpha} = \varphi_{\alpha,j}^K(\Delta; \mathbf{u}^k)$ (using (4.14)), $f_k = \displaystyle\int_{\mathbb{R}^d} f(x)e_k(x)dx$, $u_{N,k}^{K,n}(t_0) = \displaystyle\int_{\mathbb{R}^d} p_0(x)e_k(x)dx$.

2. Real − time computations, i − th step (as the observations become available): compute ξ_α^i, $\alpha \in \mathcal{J}_N^n$ *according to (4.1) and update the coefficients* $u_{N,k}^{K,n}$ *as follows:*

$$Q_{jk}^K(\xi^i) = \sum_{\alpha \in \mathcal{J}_N^n} q_{jk}^{K,\alpha}\xi_\alpha^i, \quad u_{N,j}^{K,n}(t_i) = \sum_{k=1}^{K} Q_{jk}^K(\xi^i)u_{N,k}^{K,n}(t_{i-1}), \quad j = 1, \ldots, K; \tag{4.19}$$

then, if necessary, compute

$$u_N^{K,n}(t_i, x) = \sum_{j=1}^{K} u_{N,j}^{K,n}(t_i)e_j(x) \tag{4.20}$$

and/or

$$\widetilde{\phi}_{t_i}[f] = \sum_{j=1}^{K} f_j u_{N,j}^{K,n}(t_i), \quad \widetilde{f}_{t_i} = \frac{\widetilde{\phi}_{t_i}[f]}{\widetilde{\phi}_{t_i}[1]}. \tag{4.21}$$

We call this algorithm the Spectral Separating Scheme of the Second Kind. It was suggested and analyzed in [22] when $\rho = 0$ and in [46] for the general model (2.1).

4.3. **Discussion.** The main advantage of the spectral separating schemes, as compared to most other nonlinear filtering algorithms, is that the time consuming computations, including solving partial differential equations and evaluation of integrals, are performed in advance, while the real-time part is relatively simple even when the dimension d of the state process is large. Here are some other features of the spectral separating schemes:

(1) The overall amount of preliminary computations does not depend on the number of the on-line time steps;

(2) Formulas (4.10) and (4.21) can be used to compute an approximation to $\widehat{f}_{t_i}$, for example, conditional moments, without the time consuming computations of the unnormalized filtering density and the related integrals;
(3) Only the coefficients $\psi_j^K(i,n,N)$ or $u_{N,j}^{K,n}(t_i)$ must be updated at every time step; the filtering density and/or filter can be computed independently of each other as needed, for example, at the final time moment.
(4) The real-time part of the algorithms can be easily parallelized.
(5) If $n = 1$, then each ξ_α^i depends only on the increments $Y_l(t_i) - Y_l(t_{i-1})$ of the observation process, and the corresponding algorithms can be used for filtering with discrete time observations [50]. For $n > 1$ and $k > 1$, the integral $\int_{t_{i-1}}^{t_i} m_k(s - t_{i-1})dY_l(s)$ can be reduced to a usual Riemann integral and then approximated by the trapezoidal rule.
(6) The implementation of both algorithms does not depend on whether the model is noise-correlated ($\rho \not\equiv 0$) or not.

Successful implementation of the algorithms requires effective numerical methods for solving deterministic parabolic equations [61] and evaluating integrals [13], but no special tools from numerical stochastics. On the other hand, successful testing and tuning of the algorithms will require effective numerical methods for stochastic ODEs to simulate the processes X, Y. The books [32, 54] describe many such methods.

Theoretical analysis of the algorithms is possible with little or no change if the model is not time homogeneous, that is, the functions b, ρ, σ, h in (2.1) depend on time. This time dependence certainly decrease the computational advantages, as the number of preliminary computations will grow substantially and will depend on the number of the on-line time steps.

The Wiener chaos approach is far less effective if the functions b, ρ, σ, h in (2.1) depend on the observation process Y, because the corresponding systems (4.2) and (4.14) have a much more complicated structure and are no longer solvable by induction. The corresponding analysis is still an open problem.

4.4. **Approximation error.** The quality of the approximation for the spectral separating schemes is controlled by four numbers: K, n, N, and Δ. On the other hand, the amount of the preliminary computations and the storage space are controlled by the size of the array q; the size of this array is $K^2|\mathcal{J}_N^n|$, where K is the number of basis functions in $L_2(\mathbb{R}^d)$, and $|\mathcal{J}_N^n|$, the size of the set $\mathcal{J}_N^n$, is the number of the Cameron-Martin basis functions. By construction, it is impossible to improve the quality of approximation without increasing K. While increasing n and N should also lead to better approximation, it is essentially impossible to use large values of n and N because of the prohibitively large size of the set $\mathcal{J}_N^n$. For example, if $r = 1$, the number of the elements in the set $\mathcal{J}_5^{10}$ is 740, and this number *more than doubles* for $r = 2$. A rough asymptotic of $|\mathcal{J}_N^n|$ is $(nN)^r$. Accordingly, the convergence of the approximations must be studied with fixed values of n and N: to improve the quality of approximation, we should decrease the time step Δ and increase the number K of the spatial basis functions.

The study of convergence of the spectral separating schemes requires a special choice of the bases $\mathfrak{e}$ and $\mathfrak{m}$, as well as extra regularity of the filtering model (2.1).

We begin by specifying the basis $\mathfrak{e}$ in $L_2(\mathbb{R}^d)$. Denote by Γ the set of ordered d-tuples $\gamma = (\gamma_1, \dots, \gamma_d)$ with $\gamma_j = 0, 1, 2, \dots$. For $\gamma \in \Gamma$ define

$$\mathcal{H}_\gamma(x) = \prod_{j=1}^{d} \mathcal{H}_{\gamma_j}(x_j),$$

where

$$\mathcal{H}_n(t) = \frac{(-1)^n}{\sqrt{2^n \pi^{1/2} n!}}\, e^{t^2/2}\, \frac{d^n}{dt^n}\, e^{-t^2},\ n = 0, 1, 2, \dots.$$

If Λ is the operator

$$\Lambda = -\nabla^2 + (1 + |x|^2), \tag{4.22}$$

where ∇^2 is the Laplace operator, then direct computations show that

$$\Lambda \mathcal{H}_\gamma = \lambda_\gamma e_\gamma, \tag{4.23}$$

with $\lambda_\gamma = (2\sum_{j=1}^{d} \gamma_j + d + 1)$.

Next, we introduce an ordering of the set Γ as follows: define $|\gamma| = \sum_{j=1}^{d} \gamma_j$ and then say that $\gamma < \tau$ if $|\gamma| < |\tau|$ or if $|\gamma| = |\tau|$ and $\gamma < \tau$ under the lexicographic ordering, that is, $\gamma_{i_0} < \tau_{i_0}$, where i_0 is the first index for which $\gamma_i \neq \tau_i$. Finally, we define the basis $\mathfrak{e}$, known as the *Hermite basis*, as the collection $\{\mathcal{H}_\gamma(x),\ \gamma \in \Gamma\}$ together with the above ordering of Γ. By construction, the elements e_k of $\mathfrak{e}$ satisfy

$$\Lambda e_k = \lambda_k e_k, \tag{4.24}$$

where $c_1 k^{1/d} \leq \lambda_k \leq c_2 k^{1/d}$ and $0 < c_1 < c_2$ do not depend on k. The construction of the Hermite basis implies that each e_k decays at infinity faster than every power of $|x|$, and therefore $f_k = \int_{\mathbb{R}^d} f(x) e_k(x) dx$ is defined for every $k \geq 1$ and every measurable function of polynomial growth.

As far as the basis $\mathfrak{m}$ in $L_2([0, \Delta])$, we use the Fourier cosine basis

$$m_1(s) = \frac{1}{\sqrt{\Delta}}; \quad m_k(s) = \sqrt{\frac{2}{\Delta}} \cos\left(\frac{\pi(k-1)s}{\Delta}\right),\ k > 1;\ 0 \leq s \leq \Delta. \tag{4.25}$$

Definition 4.3. *The filtering model (2.1) is called ν-regular for some positive integer ν if the functions σ and ρ belong to $\mathbf{C}_b^{2\nu+3}$, the functions b and h belong to $\mathbf{C}_b^{2\nu+2}$, and $\Lambda^\nu p_0$ belongs to $\mathbf{H}^1$, with Λ as in (4.22).*

We are now ready to study the convergence of the spectral separating schemes. Recall that the Spectral Separating Scheme of the First Kind defines the approximations $u^n_{N,K}(t_i, x)$, $\widetilde{\phi}_{t_i}$ of the unnormalized filtering density and filter according to (4.8). The following theorem presents the quality of these approximations and establishes the convergence in the limit $\lim_{\Delta\to 0} \lim_{K\to\infty}$ for the noise uncorrelated model.

Theorem 4.4. *Assume that $N \geq 2$, $\rho(x) \equiv 0$, and the matrix $\sigma\sigma^*$ is uniformly positive definite, that is, condition (2.6) holds. If the filtering model (2.1) is ν-regular*

for some $\nu > d+1$, then

$$\max_{0\le i\le M} \mathbb{E}\|u(t_i,\cdot) - u^n_{N,K}(t_i,\cdot)\|_0 \le C_0 \left(\frac{(C_{11}\Delta)^{N/2}}{\sqrt{(N+1)!}} + \frac{C_{12}\Delta}{\sqrt{n}} \right) + \frac{C_2}{K^{(\nu-d-1)/d}\Delta}. \quad (4.26)$$

The number C_0 depends on T and the parameters of the model, that is, the coefficients and the initial condition in the Zakai equation (2.8); the numbers C_{11}, C_{12} depend only on the parameters of the model; the number C_2 depends on ν, T, and the parameters of the model.

If, in addition, $(1+|x|^2)^{-w} f \in L_2(\mathbb{R}^d)$ for some $w \ge 0$ so that $\nu > d+1+w$ and $\Lambda^\nu((1+|x|^2)^w p_0) \in \mathbf{H}^1$, then

$$\max_{0\le i\le M} \mathbb{E}|\phi_{t_i}[f] - \widetilde{\phi}_{t_i}[f]| \le C_3 \left(\frac{(C_{11}\Delta)^{N/2}}{\sqrt{(N+1)!}} + \frac{C_{12}\Delta}{\sqrt{n}} \right) + \frac{C_4}{K^{(\nu-d-1)/d}\Delta}. \quad (4.27)$$

The numbers C_3, C_4 depend on ν, T, the function f, and the parameters of the model; the numbers C_{11} and C_{12} are the same as in (4.26).

Proof. We only indicate the main steps; the details are in [47]. Consider first the local error $\widetilde{\mathbb{E}}\|u(\Delta,\cdot) - u^n_{N,K}(\Delta,\cdot)\|_0^2$. Define $u_N(\Delta,x) = \sum_{\alpha\in\mathcal{J}^n_N} \varphi_\alpha(\Delta,x,p_0)$. By Theorem 2.2 in [47],

$$\widetilde{\mathbb{E}}\|u(\Delta,\cdot) - u^n_N(\Delta,\cdot)\|_0^2 \le c_1 e^{c_2\Delta} \left(\frac{(c_3\Delta)^{N+1}}{(N+1)!}\|p_0\|_0^2 + \frac{\Delta^3}{n}\|p_0\|_2^2 \right), \quad (4.28)$$

where the numbers c_1, c_2, c_3 depend only on the coefficients of (2.1); recall that $\|\cdot\|_2$ is the norm in the Sobolev space $\mathbf{H}^2$. Next, by Theorem 2.6 in [47],

$$\widetilde{\mathbb{E}}\|u^n_N(\Delta,\cdot) - u^n_{N,K}(\Delta,\cdot)\|_0^2 \le c_4 e^{c_5\Delta} K^{-2(\nu-d-1)/d} \|\Lambda^\nu p_0\|_0^2, \quad (4.29)$$

where the numbers c_3, c_4 depend on ν and the parameters of the model. We combine (4.28), in which $N \ge 2$, and (4.29) to get the overall local error

$$\widetilde{\mathbb{E}}\|u(\Delta,\cdot) - u^n_{N,K}(\Delta,\cdot)\|_0^2 \le \left(c_6\Delta^3 + c_7 K^{-2(\nu-d-1)/d}\right) e^{c_8\Delta};$$

the global error is then

$$\widetilde{\mathbb{E}}\|u(t_i,\cdot) - u^n_{N,K}(t_i,\cdot)\|_0^2 \le c_9\Delta^2 + c_{10} K^{-2(\nu-d-1)/d}\Delta^{-2};$$

see [47] for details. Error bound (4.26) now follows from (2.4).

Error bound (4.27) follows from (4.26) by the Cauchy-Schwartz inequality. □

The following properties of the functions m_k were essential in the proof of (4.28): if $M_k(t) = \int_0^t m_k(s)ds$, then $M_k(\Delta) = 0$, $|M_k(t)| \le \sqrt{\Delta}/n$. Any other basis with these properties can also be used, but for now the Fourier cosine basis (4.25) appears to be the only one for which these properties are easily verified. The Haar basis, while simplifying calculations of ξ^i_α, results in a *local* error bound (4.28) with a slower rate of decay in Δ [6, Corollary 3.8], [7, Corollary 4.5]; the corresponding *global* error bound for the Haar basis has not been derived yet.

The assumption $\rho \equiv 0$ was also essential for the proof of (4.28); without this assumption, a slightly weaker error bound holds.

Theorem 4.5. *Assume that the matrix $\sigma\sigma^*$ is uniformly positive definite. If the filtering model (2.1) is ν-regular for some $\nu > \max(4, d+1)$, then*

$$\max_{0\le i\le M}\mathbb{E}\|u(t_i,\cdot)-u^n_{N,K}(t_i,\cdot)\|_0 \le C_1\left(\frac{1}{(1+\delta)^{N/2}}+\frac{1}{\sqrt{n}}\right)\Delta^{1/2}+\frac{C_2}{K^{(\nu-d-1)/d}\Delta}. \quad (4.30)$$

The number C_1 depends on T and the parameters of the model, that is, the coefficients and the initial condition in the equation (2.1); the number $\delta > 0$ depends only on the parameters of the model; C_2 depends on ν, T, and the parameters of the model.

If, in addition, $(1+|x|^2)^{-w}f \in L_2(\mathbb{R}^d)$ for some $w \ge 0$ so that $\nu > d+1+w$ and $\Lambda^\nu((1+|x|^2)^w p_0) \in \mathbf{H}^1$, then

$$\max_{0\le i\le M}\mathbb{E}|\phi_{t_i}[f]-\widetilde{\phi}_{t_i}[f]| \le C_3\left(\frac{1}{(1+\delta)^{N/2}}+\frac{1}{\sqrt{n}}\right)\Delta^{1/2}+\frac{C_4}{K^{(\nu-d-1)/d}\Delta}. \quad (4.31)$$

The numbers C_3, C_4 depend on ν, T, the function f, and the parameters of the model.

Proof. Once we establish the local error bound of the type (4.28), which in this case turns out to be

$$\widetilde{\mathbb{E}}\|u(\Delta,\cdot)-u^n_N(\Delta,\cdot)\|_0^2 \le c_1 e^{c_2\Delta}\left(\frac{\Delta^2}{(1+\delta)^N}\|p_0\|_2^2+\frac{\Delta^2}{n}\|p_0\|_4^2\right), \quad (4.32)$$

for a suitable $\delta > 0$, the proof is completed by the same arguments as in Theorem 4.4.

To establish (4.32), we use equality (3.17), in which we replace equation (3.15) with the Zakai equation (2.8), and also put $X = \mathbf{H}^\gamma$ for a suitable γ. Then

$$\sum_{|\alpha|=n}\|\varphi_\alpha(t,\cdot,p_0)\|_\gamma^2 = \sum_{k_1,\dots,k_n=1}^{r}\int_0^t\int_0^{s_n}\cdots\int_0^{s_2} \|\Phi_{t-s_n}\mathcal{M}^*_{k_n}\cdots\Phi_{s_2-s_1}\mathcal{M}^*_{k_1}\Phi_{s_1}p_0\|_\gamma^2 ds_1\dots ds_n, \quad (4.33)$$

where $\Phi = \Phi_t$ is the semi-group of the operator $\mathcal{L}^*$. The assumptions of the current theorem imply that the semi-group Φ is bounded above by the heat kernel:

$$\|\Phi_t f\|_\gamma \le C_1\int_{\mathbb{R}^d} e^{-C_2|y|^2 t}|\breve{f}(y)|^2(1+|y|^2)^\gamma dy \quad (4.34)$$

for some positive numbers C_1, C_2, where $\breve{f}$ is the Fourier transform of f; see [18] for details. Notice also that

$$\|\mathcal{M}^*_k f\|_\gamma \le C_3(\|f\|_\gamma + \|\nabla f\|_\gamma),$$

where ∇f is the gradient of f. Then direct computations show that

$$\int_0^t\int_0^s \|\mathcal{M}^*_k\Phi_{s-s_1}f(s_1)\|_\gamma^2 ds_1 ds \le C_4\int_0^t\|f(s)\|_\gamma^2 ds.$$

For $n \geq 2$, we combine the last inequality with Theorem 9.5 in [49] to conclude that

$$\begin{aligned}
&\sum_{k_1,\dots,k_n=1}^{r} \int_0^\Delta \int_0^{s_n} \dots \int_0^{s_2} \|\Phi_{\Delta-s_n}\mathcal{M}_{k_n}^* \cdots \Phi_{s_2-s_1}\mathcal{M}_{k_1}^*\Phi_{s_1}p_0\|_0^2 ds_1 \dots ds_n \\
\leq & C_5 \sum_{k_1,\dots,k_{n-3}}^{r} \int_0^\Delta \int_0^{s_n} \dots \int_0^{s_2} \qquad (4.35) \\
& \|\mathcal{M}_{k_{n-3}}^*\Phi_{s_{n-2}-s_{n-3}}\mathcal{M}_{k_{n-3}}^* \cdots \Phi_{s_2-s_1}\mathcal{M}_{k_1}^*\Phi_{s_1}p_0\|_2^2 ds_1 \dots ds_n \\
& \leq C_6(1+\delta)^{-n}\Delta^2\|p_0\|_2^2
\end{aligned}$$

for some $\delta > 0$. Then local error bound (4.32) follows by same arguments as in the proof of Theorem 4.4. The main reason for the factor Δ^2 rather than Δ^3 in (4.35) is that the operators $\mathcal{M}_k^*$ do not commute with one another when $\rho \not\equiv 0$. □

If the matrix $\sigma\sigma^*$ is not uniformly positive definite, then the rate of convergence is an open question.

We now establish the rate of convergence for the Spectral Separating Scheme of the Second Kind. Recall that this algorithm defines the approximations $u_N^{K,n}(t_i,x)$, $\widetilde{\phi}_{t_i}$ of the unnormalized filtering density and filter according to (4.18). The following theorem presents the quality of these approximations and establishes the convergence in the limit $\lim_{K\to\infty}\lim_{\Delta\to 0}$.

Theorem 4.6. *If the filtering model (2.1) is ν-regular for some $\nu > d+1$, then*

$$\max_{0\leq i\leq M} \mathbb{E}\|u(t_i,\cdot) - u_N^{K,n}(t_i,\cdot)\|_0 \leq \frac{C_1}{K^{(\nu-d-1)/d}} + C_2\left(\frac{(C_{21}\Delta)^{N/2}}{\sqrt{(N+1)!}} + \frac{C_{22}\Delta^{1/2}}{\sqrt{n}}\right). \quad (4.36)$$

The number C_1 depends on ν, T, and the parameters of the model, that is, the coefficients and the initial condition in the Zakai equation (2.8); the number C_2 depends on T, K and the parameters of the model; the numbers C_{21}, C_{22} depend on K and the parameters of the model.

If, in addition, $(1+|x|^2)^{-w}f \in L_2(\mathbb{R}^d)$ for some $w \geq 0$ so that $\nu > d+1+w$ and $\Lambda^\nu((1+|x|^2)^w p_0) \in \mathbf{H}^1$, then

$$\max_{0\leq i\leq M} \mathbb{E}|\phi_{t_i}[f] - \widetilde{\phi}_{t_i}[f]| \leq \frac{C_3}{K^{(\nu-w-d-1)/d}} + C_4\left(\frac{(C_{21}\Delta)^{N/2}}{\sqrt{(N+1)!}} + \frac{C_{22}\Delta^{1/2}}{\sqrt{n}}\right). \quad (4.37)$$

The number C_3 depends on ν, T, the function f, and the parameters of the model; the number C_4 depends on K, T, the function f, and the parameters of the model; the numbers C_{21}, C_{22} are the same as in (4.36).

Proof. This theorem is proved in [46]. The reference also contains a more detailed information about the constants C. □

Note that, in the Spectral Separating Scheme of the Second Kind, the approximation in space is carried out first, and the Winer chaos expansion is applied to a system of ordinary differential equations (4.12). As a result, unlike Theorems 4.4 and 4.5,

the error bound can be established with no additional non-degeneracy assumptions about the matrix $\sigma\sigma^*$. According to [4, 10], the rate of convergence in Δ for an approximation of the optimal filter for (2.1) is, in general, not better than $\Delta^{1/2}$, and both spectral separating schemes achieve this rate. Indeed, for $N \geq 2$, formulas (4.31) and (4.37) can be written as

$$\max_{0\leq i\leq M} \mathbb{E}|\phi_{t_i}[f] - \widetilde{\phi}_{t_i}[f]| \leq C_3\Delta^{1/2} + \frac{C_4}{K^{(\nu-d-1)/d}\Delta}. \tag{4.38}$$

and

$$\max_{0\leq i\leq M} \mathbb{E}|\phi_{t_i}[f] - \widetilde{\phi}_{t_i}[f]| \leq \frac{C_3}{K^{(\nu-w-d-1)/d}} + C_4\Delta^{1/2}, \tag{4.39}$$

respectively. Note that the error due to truncation in space is $C_3K^{-(\nu-w-d-1)/d}$ in both cases, but, since computation of $\widetilde{\phi}_{t_i}[f]$ in (4.38) involves truncation in space on every time step, this error is multiplied by the number of time steps, which is proportional to $1/\Delta$. The rate of convergence in time is still $\Delta^{1/2}$, since we first take the limit $K \to \infty$.

5. Other Directions

The most general model of the processes X, Y for continuous-time filtering problem in Gaussian white noise is

$$\begin{aligned} dX(t) &= A(t, X_{0,t}, Y_{0,t})dt + B(t, X_{0,t}, Y_{0,t})d\mathcal{V}(t), \\ dY(t) &= C(t, X_{0,t}, Y_{0,t})dt + D(t, X_{0,t}, Y_{0,t})d\mathcal{V}(t), \end{aligned} \tag{5.1}$$

where A, B, C, D are measurable functionals of suitable dimensions, and $\mathcal{V}$ is a Wiener process, also of a suitable dimension. Since the process $\mathcal{V}$ is multi-dimensional and the functions B, D are matrix-valued, the state and observation equations are, in general, driven by *different* noise processes, even though the first look at equation (5.1) might suggest otherwise. Under natural regularity conditions on A, B, C, D, the system of equations (5.1) has a unique strong solution [44, Theorem 4.6]. In the diffusion filtering model, the functionals A, B, C, D at time moment t depend only on $X(t)$ and $Y(t)$.

Solution of the filtering problem for (5.1) requires additional conditions on the coefficients in the equation, and under these conditions the system (5.1) is reduced to an upper triangular form

$$\begin{aligned} dX(t) &= b(t, X_{0,t}, Y_{0,t})dt + \sigma(t, X_{0,t}, Y_{0,t})dV + \rho(t, X_{0,t}, Y_{0,t})dW(t), \\ dY(t) &= h(t, X_{0,t}, Y_{0,t})dt + H(t, Y_{0,t})dW(t), \end{aligned} \tag{5.2}$$

with modified coefficients b, σ, ρ, h, H and new Wiener processes V, W; Section 6.0.2 in [62], together with Lemma 10.4 in [44], provide the details of this reduction. The square matrix H does not depend on X and is uniformly positive definite; with no dependence on t and Y, it will be just a constant invertible matrix. As a result, (2.1) describes the most general time homogeneous diffusion filtering with no observation process Y in the coefficients.

There are specific reasons for the additional assumptions about the function H in (5.2): without these assumptions, the procedure of estimating X from the observations of Y is somewhat different. In particular, if H depends on the state process X, then some information about X can be obtained from the quadratic variation of Y. If H is not uniformly positive definite, then we have the filtering problem with degenerate observation noise [5].

5.1. **Representations of the Optimal Filter.** In this section, we review the main representations of the optimal filter using the Zakai, Kushner-Stratonovich, and Fujisaki-Kallianpur-Kunita equations, as well as the Kallianpur-Striebel formula. The Wiener chaos expansion can be used to study any of these representations.

Consider the filtering model (5.2). If $f = f(t, X(t), Y_{0,t})$ is a square integrable functional, and the functionals b, σ, ρ, h at time t depend only on $X(t)$ in a sufficiently smooth way, then the optimal filter $\widehat{f} = \mathbb{E}\big(f|\mathcal{F}_t^Y\big)$ has the representation

$$\widehat{f} = \frac{\int_{\mathbb{R}^d} f(t,x,Y_{0,t})u(t,x)dx}{\int_{\mathbb{R}^d} u(t,x)dx}, \tag{5.3}$$

where u satisfies the Zakai equation similar to (2.8). The original reference, the paper by Zakai [68], provides the derivation for the diffusion model with the coefficients independent of Y and with $\rho \equiv 0$. A more general derivation, together with the detailed investigation of the analytical properties of the solution, is in [36, 62]. Some of the results of [68] were obtained independently by Duncan [17] and Mortensen [56]; accordingly, equation (2.8) is also known as the DMZ (Duncan-Mortensen-Zakai) equation. Being a linear equation, (2.8) is well-suited for analysis using various forms of the Wiener chaos expansion, especially if the coefficients do not depend on Y.

If the unnormalized filtering density $u = u(t,x)$ exists, then the time evolution of the NORMALIZED FILTERING DENSITY $p(t,x) = u(t,x)/\int_{\mathbb{R}^d} u(t,x)$ is described by a non-linear integro-differential equation, which for the time-homogeneous diffusion model (2.1) becomes

$$\begin{aligned} p(t,x) = p_0(x) + \int_0^t \mathcal{L}^* p(s,x)ds + \sum_{l=1}^r \Big(\mathcal{M}_l^* p(s,x) \\ - p(s,x)\int_{\mathbb{R}^d} p(s,x)h_l(x)dx\Big)\left(dY_l - \left(\int_{\mathbb{R}^d} p(s,x)h_l(x)dx\right)ds\right). \end{aligned} \tag{5.4}$$

The nonlinear structure of this equation complicates the analysis by means of the Wiener chaos expansion, leaving it an open problem.

The time evolution of the normalized filtering density was originally derived, in various forms and with various degrees of mathematical rigor, in [40, 42, 64, 65], and is known as the Kushner-Stratonovich equation. Some of the computations in [65] do not agree with the accepted standards of the Itô calculus and were initially dismissed as a mistake. A more careful analysis of the computations later lead to the creation of the now famous Stratonovich stochastic calculus [66]. Many later works, such as Chapter 6 in [62], or Section 8 in [34], present a modern approach, both to the derivation and to the study of the analytical properties of the filtering density p.

It can happen that representation (2.3) of the optimal filter holds, while representation (5.3) does not; one example is the general model (5.2). In fact, for (2.3) to hold, the process X does not need any particular structure. Papers [45, 52, 57, 67] successfully study the optimal filter using (2.3) and various versions of the Wiener chaos expansion. The results illustrate the power of the Wiener chaos method by providing an insight into the structure of the optimal filter when the Zakai equation is not available. On the other hand, this level of generality prevents a detailed analysis of the potential numerical methods based on the expansions.

Another application of representation (2.3) is the PARTICLE SYSTEM APPROXIMATIONS of the optimal filter [14, 15, 16], which are deep extensions of the Monte Carlo method.

Kallianpur and Striebel [29] were among the first to realize the importance of the representation (2.3) for the solving the filtering and other estimation problems. Accordingly, (2.3) is known as the Kallianpur-Striebel formula.

One disadvantage of (2.3) is that this representation does not provide the time evolution of the optimal filter. It turns out that this time evolution can be written even when the filtering density does not exist. The Fujisaki-Kallianpur-Kunita equation [21] describes the time evolution of the conditional expectation $\widehat{F}(t) = \mathbb{E}\big(F(t)|\mathcal{F}_t^Y\big)$, $t \geq 0$, for two semi-martingales F, Y and can serve as the starting point in the derivation of all other filtering equations. Assume that $F(t) = F(0) + \int_0^t B(s)ds + M(t)$ and $Y_l(t) = Y_l(0) + \int_0^t h_l(s)ds + \sum_{k=1}^t H_{lk}(s)dW_k(s)$, $l = 1, \ldots, r$, so that the matrix $H(s)$ is $\mathcal{F}_t^Y$-measurable and invertible. Define $\overline{W}_k(s) = \int_0^t H_{kl}^{-1}(s)(dY_l(s) - \widehat{h}_l(s)ds)$ and let D_l be the process so that $\langle M, W_l\rangle(t) = \int_0^t D_l(s)ds$, where $\langle M, W_l\rangle$ is the quadratic covariation of the martingale M and the Wiener process W_l. Then

$$\begin{aligned}\widehat{F}(t) = \widehat{F}(0) + \int_0^t \widehat{B}(s)ds \\ + \sum_{k=1}^r \int_0^t \left(\widehat{D}(s) + \sum_{l=1}^r \left(\widehat{Fh_l}(s) - \widehat{F}(s)\widehat{h}_l(s)\right) H_{lk}^{-1}(s)\right) d\overline{W}_k(s)\end{aligned} \tag{5.5}$$

The complicated nonlinear structure of this time evolution prevents a direct numerical analysis, including the use of the Wiener chaos expansion.

Despite the technical difficulties, the Wiener chaos can be used to study nonlinear equations [53], and so the analysis of the Fujisaki-Kallianpur-Kunita and Kushner-Stratonovich equations using Wiener chaos is an open problem.

5.2. **Solving the Filtering Problem.** If the pair of the process (X, Y) is jointly Gaussian, then, by the Normal Correlation Theorem, the conditional distribution of $X(t)$ given $\mathcal{F}_t^Y$ is Gaussian and uniquely defined by the conditional mean $m(t) = \mathbb{E}(X(t)|\mathcal{F}_t^Y)$ and variance $P(t) = \mathbb{E}((X(t) - m(t))(X(t) - m(t))^*|\mathcal{F}_t^Y)$. The system of stochastic ordinary differential equations describing time evolution of m and P is the foundation of the KALMAN-BUCY FILTER [30]. Various extensions of this filter to conditionally Gaussian processes have been derived [44].

In the Gaussian model, the filtering density is characterized by a finite number of parameters. Even though *exact finite dimensional filters* can exist for non-Guassian processes X, Y [2, 55], a result of Ocone and Pardoux [58, 60] shows that for most models the optimal filter is infinite-dimensional. Some special infinite-dimensional optimal filters can be computed exactly [12, 33, 63], but a typical filtering problem requires an approximation of the optimal filter.

The EXTENDED KALMAN FILTER [27] is one of the most straightforward approximations and is derived by applying the Kalamn filter to the linearization of the model around the current estimate of X. This approximation preserves the relative analytical and computational simplicity of the Kalman filter and works well in many applications. The main drawback is that there is no rigorous justification for such approximations, that is, we have no reliable bounds on the deviation of the computed filter estimate from the optimal value. More theoretically sound are the geometrically intrinsic filter [11], and the moment approximation of the optimal filter [41, 39]. Still, the error analysis of these approximations is an open problem. A large class of numerical methods is based on approximating the original filtering model with a simpler one, for a example, a finite state Markov chain, and using the actual observations in the optimal filter for the approximating model [43].

Still, the most straightforward way to get an approximation together with an error bound is to solve numerically the equation for the filtering density, either normalized or unnormalized. The unnormalized filtering density is a more popular object to study because the corresponding equation is linear. For some numerical methods, the step-by-step solution of the Zakai equation with normalization at every step provides an approximate solution of the Kushner-Stratonovich equation, with approximation error under control [26].

A large class of numerical methods for the Zakai equation employ the corresponding algorithms used for deterministic partial differential equations: Galerkin approximation [3, 25], finite element [23], or operator splitting [4, 19, 20]. All these methods require the solution of certain partial differential equation at every time step, which is usually impossible to achieve in real time if the dimension of the process X is bigger than three. The spectral separating schemes deal with this "curse of dimensionality" by doing all the time consuming calculations in advance. With a careful choice of the branching mechanism, the particle system approximations [14, 15, 16] also have a potential for the real-time implementation.

For $t < T$, the problems of computing $\mathbb{E}\big(f(X(t))|\mathcal{F}_T^Y\big)$ (interpolation, or smoothing) and $\mathbb{E}\big(f(X(T))|\mathcal{F}_t^Y\big)$ (extrapolation, or prediction) are studied using the same technical tools as in the filtering problem; see [44, Sections 8.4, 8.5] or [62, Section 6.3]. Application of the Wiener chaos expansions to the problems of interpolation and extrapolation is mostly an open question.

The traditional formulation of the filtering problem assumes that the probability distributions of X and Y are completely known. Multiple model filtering [1] provides a more realistic setting and can be studied using Wiener chaos expansions [48]. An even more realistic setting is simultaneous filtering and estimation [9], where application of the Wiener chaos is an open problem.

6. Acknowledgment

I am grateful to the referee for a number of helpful comments and suggestions.

References

[1] Y. Bar-Shalom and X.-R. Li, *Estimation and tracking. Principles, techniques, and software*, Artech House, Inc., Boston, MA, 1993.
[2] V. E. Benesh, *Exact finite-dimensional filters for certain diffusions with nonlinear drift*, Stochastics **5** (1981), 65–92.
[3] J. F. Bennaton, *Discrete time Galerkin approximation to the nonlinear filtering solution*, J. Math. Anal. Appl. **110** (1985), no. 2, 364–383.
[4] A. Bensoussan, R. Glowinski, and R. Rascanu, *Approximations of the Zakai equation by splitting-up method*, SIAM Journal on Control and Optimization **28** (1990), no. 6, 1420–1431.
[5] A. V. Borisov, *Optimal filtering in systems with degenerate noise in the observations*, Automat. Remote Control **59** (1999), no. 11, 1526–1537.
[6] A. Budhiraja and G. Kallianpur, *Approximations to the solution of the Zakai equation using multiple Wiener and Stratonovich integral expansions*, Stochastics Stochastics Rep. **56** (1996), no. 3–4, 271–315.
[7] ———, *The Feynman-Stratonovich semigroup and Stratonovich integral expansions in nonlinear filtering*, Appl. Math. Optim. **35** (1997), no. 1, 91–116.
[8] R. H. Cameron and W. T. Martin, *The orthogonal development of nonlinear functionals in a series of Fourier-Hermite functions*, Ann. Math. **48** (1947), no. 2, 385–392.
[9] F. Campillo and F. Le Gland, *MLE for partially observed diffusions: direct maximization vs. the EM algorithm*, Stochastic Process. Appl. **33** (1989), no. 2, 245–274.
[10] J. M. C. Clark and R. J. Cameron, *The maximum rate of convergence of discrete approximation for stochastic differential equations*, Springer Lecture Notes in Control and Inform. Sc. **25** (1980), 162–171.
[11] R. W. R. Darling, *Geometrically intrinsic nonlinear recursive filters I: Algorithms*, Tech. Report 494, UC Berkeley, Depertment of Statistics, March 1998, http://www.stat.berkeley.edu/tech-reports/494.pdf.
[12] F. E. Daum, *New exact nonlinear filters*, Bayesian Analysis of Time Series and Dynamic Models (J. C. Spall, ed.), Marsel Dekker, 1988, pp. 199–226.
[13] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*, Academic Press, 1984.
[14] P. Del Moral, *Non-linear filtering: interacting particle resolution*, Markov Proc. Rel. Fields **2** (1996), 555–579.
[15] P. Del Moral, J. Jacod, and P. Protter, *The Monte-Carlo method for filtering with discrete-time observations*, Probab. Theory Related Fields **120** (2001), no. 3, 346–368.
[16] P. Del Moral and L. Miclo, *Branching and intercting particle systems approximations of feynman-kac formula with applications to non-linear filtering*, Séminaire de Probabilités, XXXIV, Lecture Notes in Mathematics, vol. 1729, Springer, Berlin, 2000, pp. 1–145.
[17] T. E. Duncan, Tech. Report 7001-4, Stanford University, Center for Systems Research, May 1967.
[18] S. D. Eidelman, *Parabolic systems.*
[19] R. J. Elliott and R. Glowinski, *Approximations to solutions of the Zakai filtering equation*, Stoch. Anal. Appl. **7** (1989), no. 2, 145–168.
[20] P. Florchinger and F. LeGland, *Time discretization of the Zakai equation for diffusion processes observed in correlated noise*, Stoch. and Stoch. Rep. **35** (1991), no. 4, 233–256.
[21] M. Fujisaki, G. Kallianpur, and H. Kunita, *Stochastic differential equations for the nonl inear filltering problem*, Osaka J. Math. **9** (1972), no. 1, 19–40.
[22] C. P. Fung and S. V. Lototsky, *Nonlinear filtering: separation of parameters and observations using Galerkin approximation and Wiener chaos decomposition*, Tech. Report 1458, IMA Preprint Series, February 1997.

[23] A. Germani and M. Piccioni, *Semi-discretisation of stochastic partial differential equations on* $\mathbf{R}^d$ *by a finite element technique*, Stochastics **23** (1988), no. 2, 131–148.

[24] K. Itô, *Multiple Wiener integral*, J. Math. Soc. Japan **3** (1951), 157–169.

[25] K. Ito, *Approximation of the Zakai equation for nonlinear filtering*, SIAM J. Control Optim. **34** (1996), no. 2, 620–634.

[26] K. Ito and B. L. Rozovskii, *Approximation of the Kushner equation for nonlinear filtering*, SIAM J. Control Optim. **38** (2000), no. 3, 893–915.

[27] A. H. Jazwinski, *Stochastic processes and filtering theory*, Academic Press, New York, 1970.

[28] G. Kallianpur, *Stochastic filtering theory*, Applications of Mathematics, vol. 13, Springer, Berlin, 1980.

[29] G. Kallianpur and C. Striebel, *Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors*, Ann. Math. Stat. **39** (1968), 785–801.

[30] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction problems*, J. Basic Engineering, Trans. ASME **83D** (1961), 95–108.

[31] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus, 2nd ed.*, Springer, New York, 1991.

[32] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Springer, Berlin, 1992.

[33] M. A. Kouritzin, *Exact infinite dimensional filters and explicit solutions*, Stochastic Models, CMC Conference Proceedings, vol. 26, AMS, 2000, pp. 265–282.

[34] N. V. Krylov, *An analytic approach to SPDEs*, Stochastic Partial Differential Equations. Six Perspectives (B. L. Rozovskii and R. Carmona, eds.), Mathematical Surveys and Monographs, AMS, 1999, pp. 185–242.

[35] N. V. Krylov and A. J. Veretennikov, *On explicit formula for solutions of stochastic equations*, USSR Math. Sbornik **29** (1976), no. 2, 239–256.

[36] N. V. Krylov and A. Zatezalo, *A direct approach to deriving filtering equations for diffusion processes*, Appl. Math. Optim. **42** (2000), no. 3, 315–332.

[37] H. Kunita, *Cauchy problem for stochastic partial differential equations arising in nonlinear filtering theory*, Syst. & Cont. Letters **1** (1981), no. 1, 37–41.

[38] H. Kunita, *Stochastic flows and stochastic differential equations*, Cambridge University Press, Cambridge, 1982.

[39] H. Kushner and A. Budhiraja, *A nonlinear filtering algorithm based on an approximation of the conditional distribution*, IEEE Trans. Automat. Control **45** (2000), no. 3, 580–585.

[40] H. J. Kushner, *On the dynamical equations of conditional probability density functions, with applications to optimal stochastic control theory*, J. Math. Anal. Appl. **8** (1964), 332–344.

[41] ———, *Approximations to optimal nonlinear filters*, IEEE Trans. AC **3** (1967), 179–190.

[42] ———, *Dynamical equations for optimal nonlinear filtering*, J. Differential Equations **3** (1967), 179–190.

[43] H. J. Kushner, *Probability methods for approximations in stochastic control and for elliptic equations*, Academic Press, New York, 1977.

[44] R. Sh. Liptser and A. N. Shiryayev, *Statistics of random processes, I, II*, Springer, New York, 2002.

[45] J. T.-H. Lo and S.-K. Ng, *Optimal orthogonal expansion for estimation I: signal in white Gaussian noise*, Nonlinear Stochastic Problems (R. S. Bucy and J. M. F Moura, eds.), D. Reidel Publ. Company, 1983, pp. 291–309.

[46] S. Lototsky, *Nonlinear filtering of diffusion processes in correlated noise: analysis by separation of variables*, Appl. Math. Optim. **47** (2003), no. 2, 167–194.

[47] S. V. Lototsky, R. Mikulevicius, and B. L. Rozovskii, *Nonlinear filtering revisited: a spectral approach*, SIAM J. Contr. Optim. **35** (1997), no. 2, 435–461.

[48] S. V. Lototsky, C. Rao, and B. L. Rozovskii, *Fast nonlinear filter for continuous-discrete time multiple models*, Proceedings of the 35th IEEE Conference on Decision and Control, Kobe, Japan, 1996, vol. 4, pp. 4071–4076.

[49] S. V. Lototsky and B. L. Rozovskii, *Stochastic differential equations: a Wiener chaos approach*, A volume in honor of A. N. Shiryaev, to appear.

[50] S. V. Lototsky and B. L. Rozovskii, *Recursive nonliner filter for a continuous - discrete time model: separation of parameters and observations*, IEEE Trans. AC **43** (1998), no. 8, 1154–1158.
[51] R. Mikulevicius and B. L. Rozovskii, *Separation of observations and parameters in nonlinear filtering*, Proceedings of the 32nd IEEE Conference on Decision and Control, San Antonio, TX, 1993, vol. 2, IEEE Control Systems Society, 1993, pp. 1564–1559.
[52] R. Mikulevicius and B. L. Rozovskii, *Fourier-Hermite expansion for nonlinear filtering*, Theory Probab. Appl. **44** (2000), no. 3, 606–612.
[53] R. Mikulevicius and B. L. Rozovskii, *Global L_2-solutions of stochasic Navier-Stokes equations*, Ann. Probab. **33** (2005), no. 1, 137–176.
[54] G. Milstein, *Numerical integration of stochastic differential equations*, Kluwer Academic Publishers, Dordrecht, 1995.
[55] S. K. Mitter, *On the analogy between mathematical problems of nonlinear filtering and quantum physics*, Ricerche Automat. **10** (1979), no. 2, 163–216.
[56] R. E. Mortensen, Tech. Report ERL–66–1, UC Berkeley, Electronics Research Laboratory, August 1966.
[57] D. Ocone, *Multiple integral expansions for nonlinear filtering*, Stochastics **10** (1983), no. 1, 1–30.
[58] D. Ocone and E. Pardoux, *A Lie algebraic criterion for nonexistence of finite-dimensionally computable filters*, Stochastic partial differential equations and applications, II (Trento, 1988), Lecture Notes in Math., vol. 1390, Springer, Berlin, 1989, pp. 197–204.
[59] B. K. Øksendal, *Stochastic differential equations: an introduction with applications, 5th ed.*, Springer, 1998.
[60] E. Pardoux, *Filtrage Non Lineaire et Equations aux Derivees Partielles Stochastiques Associees*, Ecole d'été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, pp. 67–163.
[61] A. Quarteroni and A. Vali, *Numerical approximation of partial differential equations*, Springer, 1994.
[62] B. L. Rozovskii, *Stochastic evolution systems*, Kluwer Academic Publishers, Dordrecht, 1990.
[63] G. C. Schmidt, *Designing nonlinear filters based on Daum's theory*, Journal of Guidance, Control, and Dynamics **16** (1993), no. 2, 371–376.
[64] A. N. Shiryaev, *On stochastic equations in theory of conditional Markov processes*, Theor. Probab. Appl. **11** (1966), 179–184.
[65] R. I. Stratonovich, *Conditional Markov processes*, Theor. Probab. Appl. **5** (1960), 156–178.
[66] ———, *A new representation for stochastic integrals and equations*, Vestn. Moskov. Univer., Ser. Mat. Mekhan. **1** (1964), 3–12.
[67] E. Wong, *Explicit solutions to a class of nonlinear filtering problems*, Stochastics **5** (1981), no. 4, 311–321.
[68] M. Zakai, *On the optimal filtering of diffusion processes*, Z. Wahrsch. Verw. Gebiete **11** (1969), no. 3, 230–243.

Current address, S. V. Lototsky: Department of Mathematics, USC, Los Angeles, CA 90089

E-mail address, S. V. Lototsky: `lototsky@math.usc.edu`

URL: `http://math.usc.edu/~lototsky`