SIAM J. CONTROL OPTIM.
Vol. 35, No. 2, pp. 435–461, March 1997

004

NONLINEAR FILTERING REVISITED: A SPECTRAL APPROACH*

SERGEY LOTOTSKY†, REMIGIJUS MIKULEVICIUS‡, AND BORIS L. ROZOVSKII†

Abstract. The objective of this paper is to develop an approach to nonlinear filtering based on the Cameron–Martin version of Wiener chaos expansion. This approach gives rise to a new numerical scheme for nonlinear filtering. The main feature of this algorithm is that it allows one to separate the computations involving the observations from those dealing only with the system parameters and to shift the latter off-line.

Key words. Cameron–Martin development, Wick polynomials, Wiener chaos, Zakai equation

AMS subject classifications. 60G35, 60H15, 62M20

PII. S0363012993248918

1. Introduction.

Nonlinear filtering is a classic problem of applied stochastic analysis (see, e.g., Kallianpur [19], Kunita [23], Kushner [24], Liptser and Shiryayev [27], etc.). It is of notable theoretical and practical importance by itself and also as a part of control theory for partially observable stochastic systems (see, e.g., Fleming and Pardoux [11]).

In this paper we consider the filtering scheme where the signal process $x(t)$ is a Markov diffusion process and the observation process is of the form

$$y(t) = y_0 + \int_0^t h(x(s))ds + w(t),$$

where $w(t)$ is a Brownian motion independent of the process $x(t)$.

Let f be a given bounded function on $\mathbf{R}^d$ and $\hat{f}(x(t))$ be the optimal filter (the best in the mean-square estimate for $f(x(t))$ based on observations $y(s)$, $s \le t$). A fundamental result of filtering theory says that the optimal filter is given by the formula

$$\hat{f}(x(t)) = \frac{\int_{\mathbf{R}^d} f(x)u(t,x)dx}{\int_{\mathbf{R}^d} u(t,x)dx}, \tag{1.1}$$

where $u(t,x)$ is the so-called unnormalized filtering density (UFD); of course, some regularity assumptions are needed to ensure the existence of the density.

A standard way to study the UFD (analytically or numerically) is to treat it as a solution of the Zakai equation

$$du(t,x) = \mathcal{L}^* u(t,x)dt + h(x)u(t,x)dy(t), \tag{1.2}$$

where $\mathcal{L}^*$ is the formally adjoint operator to the generator of the Markov process $x(t)$ (see, e.g., Baras [2]; Benesh [3]; Bensoussan, Glowinski, and Rascanu [4]; Clark [8];

*Received by the editors May 14, 1993; accepted for publication (in revised form) December 21, 1995. This research was supported in part by Office of Naval Research grant N00014-95-1-0229 and Army Research Office grant DAAH 04-95-1-0164. A version of this paper was presented at the 35th IEEE Conference on Decision and Control, Kobe, Japan, December 11–13, 1996.

http://www.siam.org/journals/sicon/35-2/24891.html

†Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, CA 90089-1113 (rozovski@cams.usc.edu).

‡Institute of Mathematics and Informatics, Akademijos 4, Vilnius 2600, Lithuania.

DiMasi and Runggaldier [9]; Elliott and Glowinski [10]; Florchinger and LeGland [12]; Krylov and Rozovskii [20]; Kunita [22]; Pardoux [33]; Rozovskii [34]; Zakai [37]; etc.).

Another comparatively recent approach is based on the Wiener chaos expansion (WCE) (see references below). In this paper we further develop a version of this approach based on the Cameron–Martin orthogonal decomposition of L_2-functionals of a Gaussian process (see Cameron and Martin [7]). We prove that the UFD can be written in the form

$$u(t,x) = \sum_{\alpha} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t,x)\xi_\alpha(y), \tag{1.3}$$

where $\xi_\alpha(y)$ are Wick polynomials (certain products of Hermite polynomials; see, e.g., [14]) of Wiener integrals $\int_0^t m_i(s)dy(s)$, where $\{m_k\}$ is a complete orthonormal system in $L_2([0,t])$, and $\varphi_\alpha(t,x)$ are deterministic Hermite–Fourier coefficients in the Cameron–Martin orthogonal decomposition of $u(t,x)$ (see Mikulevicius and Rozovskii [30, 31]). The Wick series expansion (1.3) converges in L_2-sense on the reference probability space.

We prove that the set of functions $\{\varphi_\alpha(t,x)\}$ is a solution to a simple recursive system of Kolmogorov-like equations (see (2.6)). Below it will be referred to as the S-system.

Our interest in the WCE was motivated mainly by computational purposes. One important feature of the expansion (1.3) is that it separates observations and parameters in that the Wick polynomials are completely defined by the observation process $y(t)$ but the Hermite–Fourier coefficients $\varphi_\alpha(t,x)$ are determined only by the coefficients of the signal process $x(t)$, its initial distribution, and the observation function h.

Unfortunately, direct application of the above expansion for numerical computations is impractical, limited, at best, to short time intervals. The main reason is possible exponential growth of the errors inflicted by truncation of the infinite series (1.3) as the time interval $[0,t]$ increases (Theorem 2.2).

One important objective of the paper is to develop a numerical approximation scheme for the UFD which retains the separation of observations and parameters but is not subject to the aforementioned limitations (Theorem 2.5 and the accompanying algorithm).

This recursive scheme splits into two parts: deterministic and stochastic. The deterministic part (solving the S-system) might be time consuming but can be performed off-line since in many applications the coefficients of the processes $x(t), y(t)$ and also of the S-system are known a priori. The stochastic part (determining the Wick polynomials $\xi_\alpha(y)$) is computationally simple and can be performed in real time. In this paper this scheme is referred to as the spectral separating scheme (S^3).

We prove the strong convergence of S^3 both in L_2 and $\mathbf{C}$ spaces and demonstrate that the overall rate of convergence (on- and off-line) is of order $O(\Delta)$, where Δ is the time step (Theorems 2.2 and 2.4).

S^3 can be also viewed as a time-discretization scheme for a solution of the Zakai equation. In section 4 we demostrate that some well-known discretization algorithms for this equation (e.g., explicit Euler scheme, splitting-up method (see [4, 26])) can be derived from a multistep version of (1.3). In this section we also discuss the computational complexity of S^3, compare it with the complexity of the splitting-up method, and present some results of numerical simulations.

We conclude the introduction with some historical remarks. The idea of obtaining an "explicit" WCE solution of a stochastic (ordinary) differential equation can be

traced back to the paper [21] by Krylov and Veretennikov (see also Zvonkin and Krylov [38]). Kunita [22] applied this idea to prove uniqueness of the Zakai equation. Wong [35] obtained the solution of a special class of nonlinear filtering problems in the form of the WCE. Ocone [32] pioneered finite-order WCEs of normalized nonlinear filters (see also references therein).

In these works the multiple Wiener integral version of the WCE was used. The Cameron–Martin development is analytically equivalent to this version of the WCE (see, e.g., Ito [17]). However, it has some computational advantage since only ordinary Wiener integrals are required in this approach. Lo and Ng [28] were the first ones to utilize the above fact. They modified Ocone's approximation using the Cameron–Martin expansion. Unfortunately, the equations for the deterministic coefficients of the finite-order approximations in [28] are quite complex. To solve them one needs to know the Hermite–Fourier coefficients for the corresponding unnormalized filters. Computation of the latter was not discussed in [28].

The S-system (2.6) was introduced by Mikulevicius and Rozovskii [30, 31]. The upper bound $ce^{ct}t^{N+1}/(N+1)!$ on the error of the Nth-order approximation to (1.3) was obtained in [30]. Recently, Budhiraja and Kallianpur [5] developed a different WCE-type approximation of the unnormalized filtering density using the Haar-type basis. They also established an upper bound on the error of truncation with respect to the stochastic and deterministic bases.

2. Main results. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $w(t)$ be an r-dimensional Brownian motion on the space. Let $x(t)$ be a d-dimensional (unobservable) signal process and $y(t)$ be the r-dimensional observation process given by

$$y(t) = \int_0^t h(x(s))ds + w(t), \quad 0 \le t \le T, \tag{2.1}$$

where $h = (h^l)_{1\le l\le r}$ is an r-dimensional vector function on $\mathbf{R}^d$. We assume in addition that the signal $x(t)$ is a diffusion Markov process of the form[1]

$$\begin{aligned} dx^i(t) &= b^i(x(t))dt + \sigma^{ij}(x(t))d\tilde{w}^j(t), \quad 0 < t \le T, \\ x(0) &= x_0, \end{aligned} \tag{2.2}$$

where $b = (b^i)_{1\le i\le d}$ is a d-dimensional vector function on $\mathbf{R}^d$, $\sigma = (\sigma^{ij})_{1\le i\le d,\ 1\le j\le d_1}$ is a $d\times d_1$ dimensional matrix function on $\mathbf{R}^d$, and $\tilde{w} = (\tilde{w}^i)_{1\le i\le d_1}$ is a d_1-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathrm{P})$.

The following is assumed about the model (2.1), (2.2):

(A1) the functions b, σ, and h are infinitely differentiable and bounded with all derivatives;

(A2) the processes w and $\tilde{w}$ are independent;

(A3) the random vector x_0 is independent of both w and $\tilde{w}$ and has density[2] $p(x) \in \mathbf{H}^n$ for $n = 0, 1, 2, \ldots$.

(Then by the Sobolev embedding theorem, $p(x)$ is also in $\mathbf{C}_b^n$ for any n.) Some of these assumptions can be weakened, and we will discuss them at the end of this section.

Let $\mathcal{F}_t^y$ be the σ-algebra generated by $y(s)$, $s \le t$. Denote

$$\rho(t) = \exp\Big\{ -\int_0^t h^l(x(s))dw^l(s) - \frac{1}{2}\sum_{l=1}^r \int_0^t |h^l(x(s))|^2 ds\Big\}.$$

[1] When the sum is finite, we assume summation over repeated indices and omit the $\sum$ sign.

[2] Here and below $\mathbf{H}^n$ is the Sobolev space $W_2^n(\mathbf{R}^d)$ (see, e.g., [25]), and $\mathbf{C}_b^n$ is the space of n times continuously differentiable on $\mathbf{R}^d$ functions bounded with all the derivatives.

It is well known (see, e.g., [27] or [19]) that the measure $\tilde{\mathrm{P}}$ defined by $d\tilde{\mathrm{P}} = \rho(T)d\mathrm{P}$ is a probability measure on $(\Omega, \mathcal{F})$ with the following properties:

(i) on the reference probability space $(\Omega, \mathcal{F}, \tilde{\mathrm{P}})$, $y(\cdot)$ is a Brownian motion independent of $x(\cdot)$;

(ii) the optimal filter $\hat{f}(x(t)) = \mathbf{E}[f(x(t))|\mathcal{F}_t^y]$ is given by

$$\hat{f}(x(t)) = \frac{\tilde{\mathbf{E}}[f(x(t))\rho(t)^{-1}|\mathcal{F}_t^y]}{\tilde{\mathbf{E}}[\rho(t)^{-1}|\mathcal{F}_t^y]}, \tag{2.3}$$

where $\tilde{\mathbf{E}}$ is the expectation with respect to measure $\tilde{\mathrm{P}}$. If assumptions (A1)–(A3) hold, the unnormalized filtering measure $\Phi_t(dx) = \tilde{\mathbf{E}}[1_{\{x(t)\in dx\}}\rho(t)^{-1}|\mathcal{F}_t^y]$ admits the density $u(t,x) = \Phi_t(dx)/dx$, called the UFD, which is a solution of the Zakai equation

$$du(t,x) = \mathcal{L}^* u(t,x)dt + h^l(x)u(t,x)dy^l(t), \tag{2.4}$$

where $\mathcal{L}^* u := \frac{1}{2}\frac{\partial^2}{\partial x_i \partial x_j}((\sigma\sigma^*)^{ij}u) - \frac{\partial}{\partial x_i}(b^i u)$ and such that for every $n \in N$

$$\tilde{\mathbf{E}}\sup_{t\le T}||u(t,\cdot)||^2_{\mathbf{C}_b^n} < \infty,$$
$$\tilde{\mathbf{E}}\sup_{t\le T}||u(t,\cdot)||^2_{\mathbf{H}^n} < \infty.$$

Using the UFD $u(t,x)$, one can rewrite (2.3) in the form (1.1).

DEFINITION. *A collection $\alpha = (\alpha_k^l)_{1\le l\le r,\ k\ge 1}$ of nonnegative integers is called an r-dimensional multiindex if only finitely many of α_k^l are different from zero.*

The set of all r-dimensional multiindices will be denoted by J. For $\alpha \in J$ we use the following definitions:

$|\alpha| := \sum_{l,k}\alpha_l^k$, length of α;

$d(\alpha) := \max\{k \ge 1 :\ \alpha_k^l > 0 \text{ for some } 1 \le l \le r\}$, order of α.

We also write $\alpha! = \prod_{k,l}(\alpha_k^l!)$.

Let us fix an arbitrary orthonormal system $\{m_k\} = \{m_k(s)\}_{k\ge 1}$ in the space $L_2([0,t])$ of square integrable functions on $[0,t]$ and set

$$\xi_{k,l} = \int_0^t m_k(s)dy^l(s).$$

Note that due to property (i) of the measure $\tilde{\mathrm{P}}$, $\xi_{k,l}$ are independent Gaussian random variables with zero mean and unit variance.

Let H_n be the nth Hermite polynomial defined by

$$H_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}. \tag{2.5}$$

It is well known (see, e.g., [7] or Theorem A.1) that the collection

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

is a complete orthonormal system (CONS) in $L_2(\Omega, \mathcal{F}_t^y, \tilde{\mathrm{P}})$.

To illustrate how the system is constructed, consider the case $r = 1$. Then α is a multiindex of the form $(\alpha_1, \alpha_2, \dots)$. If $|\alpha| = 0$ (i.e., $\alpha = (0, 0, \dots)$), then obviously $\xi_\alpha \equiv 1$.

If $|\alpha| = 1$, then the multiindex α is of the form $(0, \dots, 0, 1, 0, \dots)$ (i.e., $\alpha_i = 1$, $\alpha_k = 0$, $k \neq i$). In this case, $\xi_\alpha = \int_0^t m_i(s)dy(s)$.

Similarly, if $|\alpha| = 2$, then α is of either the form

$$(0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots)$$

(if $i < j$ and $\alpha_i = \alpha_j = 1$, $\alpha_k = 0, k \neq i, j$) or the form $(0, \dots, 0, 2, 0, \dots)$ (if $i = j$). For such α we have

$$\xi_\alpha = \Big(\int_0^t m_i(s)dy(s) \Big) \Big(\int_0^t m_j(s)dy(s) \Big)$$

in the first case,

$$\xi_\alpha = \frac{1}{\sqrt{2}} \Big[\Big(\int_0^t m_i(s)dy(s) \Big)^2 - 1 \Big]$$

in the second case, and so on. See also Remark A.2.

First, we will focus on the expansion of the UFD in the Wick polynomials ξ_α. To determine the coefficients of the expansion we consider the following system of deterministic PDEs:

$$\begin{aligned} \frac{\partial \varphi_\alpha(s,x)}{\partial s} &= \mathcal{L}^* \varphi_\alpha(s,x) + \textstyle\sum_{k,l} \alpha_k^l m_k(s) h^l(x) \varphi_{\alpha(k,l)}(s,x), \quad 0 < s \leq t, \\ \varphi_\alpha(0,x) &= p(x) 1_{\{|\alpha|=0\}}, \end{aligned} \tag{2.6}$$

where $\alpha = (\alpha_k^l)_{1 \leq l \leq r,\ k \geq 1} \in J$ and $\alpha(i,j)$ stands for the multiindex $\tilde{\alpha} = (\tilde{\alpha}_k^l)_{1 \leq l \leq r,\ k \geq 1}$ with

$$\tilde{\alpha}_k^l = \begin{cases} \alpha_k^l & \text{if } k \neq i \text{ or } l \neq j \text{ or both}, \\ \max(0, \alpha_i^j - 1) & \text{if } k = i \text{ and } l = j. \end{cases} \tag{2.7}$$

This system is recursive in $|\alpha|$: once we know the functions φ_α for all α of length $|\alpha| = k$, we can compute all φ_α for $|\alpha| = k + 1$. To illustrate the idea, again consider the case $r = 1$. Let us write φ_0 for the φ_α with $\alpha = (0, 0, \dots, 0, \dots)$ ($|\alpha| = 0$). Then $\varphi_0(s,x)$ satisfies the forward Kolmogorov equation corresponding to the state process:

$$\begin{aligned} \frac{\partial \varphi_0(s,x)}{\partial s} &= \mathcal{L}^* \varphi_0(s,x), \\ \varphi_0(0,x) &= p(x). \end{aligned}$$

If $|\alpha| = 1$ with $\alpha_i = 1$ and we write φ_i for φ_α with this α, then the corresponding equation in (2.6) becomes

$$\begin{aligned} \frac{\partial \varphi_i(s,x)}{\partial s} &= \mathcal{L}^* \varphi_i(s,x) + m_i(s) h(x) \varphi_0(s,x), \\ \varphi_i(0,x) &= 0. \end{aligned}$$

For $|\alpha| = 2$, the corresponding function $\varphi_{ij},\ i \leq j$, satisfies the equation

$$\begin{aligned} \frac{\partial \varphi_{ij}(s,x)}{\partial s} &= \mathcal{L}^* \varphi_{ij}(s,x) + m_i(s) h(x) \varphi_j(s,x) + m_j(s) h(x) \varphi_i(s,x), \\ \varphi_{ij}(0,x) &= 0, \end{aligned}$$

and so on.

Under assumptions (A1) and (A3), system (2.6) has a unique smooth solution (see Proposition A.1 for details).

Our approach is based on the following expansion of the UFD.

THEOREM 2.1 (Mikulevicius and Rozovskii [30, 31]). *Assume* (A1)–(A3). *Then for each* $x \in \mathbf{R}^d$ *the UFD is given by*

$$u(t,x) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t,x)\xi_\alpha \quad (\mathrm{P}-\text{a.s.}). \tag{2.8}$$

This series converges in $L_2(\Omega, \tilde{\mathrm{P}})$, *and* $L_1(\Omega, \mathrm{P})$, *and the following Parseval's equality holds:*

$$\tilde{\mathbf{E}}|u(t,x)|^2 = \sum_{\alpha \in J} \frac{1}{\alpha!} |\varphi_\alpha(t,x)|^2. \tag{2.9}$$

Proof of this theorem is given in the appendix.

For the computational purposes one needs to truncate the sum in the expansion of u. This sum is "double infinite." Writing

$$u(t,x) = \sum_{k=0}^{\infty} \sum_{|\alpha|=k} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t,x)\xi_\alpha, \tag{2.10}$$

one can see that for $k \geq 1$ there are infinitely many multiindices α with $|\alpha| = k$. To make it finite, we have to bound the length $|\alpha|$ of α and also the order $d(\alpha)$ of α: if $d(\alpha) \leq n$, then there are at most $(nr)^k$ multiindices α with $|\alpha| = k$.

Recall that if $\alpha = (\alpha_k^l)_{1 \leq l \leq r,\ k \geq 1}$, then α_k^l defines the degree of the Hermite polynomial of $\int_0^t m_k(s)dy^l(s)$ used in the construction of ξ_α. If $d(\alpha) \leq n$, then $\alpha_k^l = 0$ for all $k > n$, so the truncation of the order of α is equivalent to keeping only the first n elements of the (deterministic) basis $\{m_k(s)\}_{k \geq 1}$.

On the other hand, by restricting the length of α, we eliminate a number of elements of the stochastic basis $\{\xi_\alpha\}$, which are otherwise available with the retained collection of $\{m_k\}$.

Thus, restriction of the order of α makes the inner sum in (2.10) finite and is equivalent to the truncation of the deterministic basis $\{m_k\}$, while restriction of the length of α makes the outer sum in (2.10) finite and is equivalent to the truncation of the stochastic basis ξ_α.

The following theorem gives the upper bound on the error that one makes by doing both truncations for a particular choice of the basis $\{m_k\}$.

THEOREM 2.2. *Suppose that assumptions* (A1)–(A3) *hold and the deterministic basis* $\{m_k\}$ *is chosen as follows:*

$$m_1(s) = \frac{1}{\sqrt{t}}, \quad m_k(s) = \sqrt{\frac{2}{t}} \cos\Big(\frac{\pi(k-1)s}{t}\Big),\ k > 1,\ 0 \leq s \leq t.$$

Write $J_N^n = \{\alpha \in J : |\alpha| \leq N,\ d(\alpha) \leq n\}$ *and define*

$$u_N^n(t,x) := \sum_{\alpha \in J_N^n} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t,x)\xi_\alpha. \tag{2.11}$$

Then

$$\tilde{\mathbf{E}}||u_N^n(t,\cdot) - u(t,\cdot)||_{L_2}^2 \leq Be^{Bt}\Big(\frac{(h_0 t)^{N+1}}{(N+1)!}||p||_{L_2}^2 + \frac{t^3}{n}||p||_{\mathbf{H}^2}^2\Big), \tag{2.12}$$

$$\sup_{x\in\mathbf{R}^d}\tilde{\mathbf{E}}|u_N^n(t,x)-u(t,x)|^2\le Ce^{Ct}\Big(\frac{(h_0t)^{N+1}}{(N+1)!}||p||^2_{\mathbf{C}_b^0}+\frac{t^3}{n}||p||^2_{\mathbf{C}_b^2}\Big). \tag{2.13}$$

Constants B and C depend only on the coefficients b, σ, and h of the model and $h_0:=\sum_{l=1}^r\sup_{x\in\mathbf{R}^d}|h^l(x)|^2$.

This and the following theorems will be proven in section 3.

REMARK 2.1. *For different k and l, random variables $\int_0^t m_k(s)dy^l(s)$, which make the stochastic basis ξ_α, are idependent and identically distributed $\mathcal{N}(0,1)$ under measure $\tilde{\mathrm{P}}$ for any CONS $\{m_k\}$. This suggests that the part of the error due to the truncation in the length of α should be independent of the choice of $\{m_k\}$, and the analysis of the proof shows that this is indeed the case. On the other hand, the error due to the truncation of the order of α crucially depends on the choice of $\{m_k\}$ (see also Remark* 3.1).

Truncations in the order and in the length can be done independently of each other. If $n=\infty$, we have truncation in length only; this case was studied by Mikulevicius and Rozovskii [30].

The Hermite–Fourier coefficients φ_α in (2.10) and (2.11) can be computed off-line, since system (2.6) does not involve the observation process y. In spite of this important property, approximation (2.11) does not yet provide an effective numerical algorithm for computing the UFD. The major reason for this is that the error of truncation may grow exponentially with t, so we can expect (2.11) to give a good approximation only for sufficiently small t. The above is a typical problem for approximations of solutions of parabolic equations (both deterministic and stochastic). One can try to offset this effect by choosing a higher-order approximation (in our case by taking larger N and n). However, higher-order numerical schemes are slower and often numerically unstable. A standard way to overcome the exponential growth of the truncation errors is to develop a recursive procedure by iterating the one-step approximation.

REMARK 2.2. *Of course, for the recursive approximation to converge, it is necessary that the error of the one-step approximation converges to zero fast enough as $t\downarrow 0$. By Theorem* 2.2*, the short time asymptotics of the error of approximation* (2.11) *are of order t if $N=1$ and of order $t^{3/2}$ if $N>1$, so it is possible to use* (2.11) *to costruct a multistep approximation* (*Theorem* 2.4).

In what follows, we present a recursive version of the expansion (2.8). It will allow us to modify the corresponding numerical scheme and eliminate the possible error growth.

Let $0=t_0<t_1<\cdots<t_M=T$ be a uniform partition of the interval $[0,T]$ with step Δ (so that $t_i=i\Delta$, $i=0,\dots,M$). Let $\{m_k^i\}=\{m_k^i(s)\}_{k\ge1}$ be a CONS in $L_2([t_{i-1},t_i])$. We also define random variables

$$\left\{\xi_\alpha^i:=\prod_{k,l}\left(\frac{H_{\alpha_k^l}(\xi_{k,l}^i)}{\sqrt{\alpha_k^l!}}\right),\quad \alpha\in J\right\}, \tag{2.14}$$

where $\xi_{k,l}^i=\int_{t_{i-1}}^{t_i}m_k^i(s)dy^l(s)$ and H_n is the nth Hermite polynomial (2.5).

Consider the following system of equations:

$$\begin{aligned}\frac{\partial\varphi_\alpha^i(s,x,g)}{\partial s}&=\mathcal{L}^*\varphi_\alpha^i(s,x,g)+\sum_{k,l}\alpha_k^lm_k^i(s)h^l(x)\varphi_{\alpha(k,l)}^i(s,x,g),\qquad t_{i-1}<s\le t_i,\\ \varphi_\alpha^i(t_{i-1},x,g)&=g(x)1_{\{|\alpha|=0\}},\end{aligned} \tag{2.15}$$

where $g(x)$ is a function to be determined. For each $i = 1, \dots, M$ this system is similar to (2.6). The main new feature is that the initial time moment is no longer zero and we now allow *that an arbitrary initial condition g may be different for different i*; this dependence on g is indicated explicitly in the arguments of φ.

The following is the recursive version of Theorem 2.1.

THEOREM 2.3. *Define $u(t_0, x) := p(x)$. Then for each $x \in \mathbf{R}^d$ and each t_i, $i = 1, \dots, M$, the UFD is given by*

$$u(t_i, x) = \sum_{\alpha \in J} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha^i(t_i, x, u(t_{i-1}, \cdot)) \xi_\alpha^i, \quad i = 1, \dots, M \ (P - \ a.s.). \tag{2.16}$$

This series converges in $L_2(\Omega, \tilde{\mathrm{P}})$ and $L_1(\Omega, \mathrm{P})$, and the following Parseval's equality holds:

$$\tilde{\mathbf{E}} |u(t_i, x)|^2 = \sum_{\alpha \in J} \frac{1}{\alpha!} \tilde{\mathbf{E}} |\varphi_\alpha(t, x, u(t_{i-1}, \cdot))|^2, \qquad i = 1, \dots, M.$$

This result follows easily from Theorem 2.1, since (2.4) is linear with a unique solution, and random variables $u(t_{i-1}, x)$ and ξ_α^i are independent under measure $\tilde{\mathrm{P}}$.

Again, for computational purposes, we need to perform truncations in (2.16). For that purpose, as in Theorem 2.2, we will use the special basis $\{m_k^i\}$:

$$\begin{aligned} &m_k^i(s) = m_k(s - t_{i-1}), \qquad t_{i-1} \le s \le t_i, \\ &m_1(s) = \frac{1}{\sqrt{\Delta}}, \quad m_k(s) = \sqrt{\frac{2}{\Delta}} \cos\Big(\frac{\pi(k-1)s}{\Delta}\Big), \ k > 1, \quad 0 \le s \le \Delta, \\ &m_k(s) = 0, \qquad k \ge 1, \ s \notin [0, \Delta]. \end{aligned} \tag{2.17}$$

THEOREM 2.4. *Suppose that basis $\{m_k^i\}$ is given by* (2.17) *and assumptions* (A1)–(A3) *hold. Define $u_N^n(t_0, x) := p(x)$ and by induction*

$$u_N^n(t_i, x) := \sum_{\alpha \in J_N^n} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha^i(\Delta, x) \xi_\alpha^i, \tag{2.18}$$

where $J_N^n = \{\alpha \in J : |\alpha| \le N, \ d(\alpha) \le n\}$ and φ_α^i are solutions of the system

$$\begin{aligned} &\frac{\partial \varphi_\alpha^i(s, x)}{\partial s} = \mathcal{L}^* \varphi_\alpha^i(s, x) + \textstyle\sum_{k,l} \alpha_k^l m_k(s) h^l(x) \varphi_{\alpha(k,l)}^i(s, x), \quad 0 < s \le \Delta, \\ &\varphi_\alpha^i(0, x) = u_N^n(t_{i-1}, x) 1_{\{|\alpha| = 0\}}. \end{aligned} \tag{2.19}$$

Then

$$\max_{1 \le i \le M} \tilde{\mathbf{E}} ||u_N^n(t_i, \cdot) - u(t_i, \cdot)||_{L_2}^2 \le B e^{BT} \Big(\frac{(h_0 \Delta)^N}{(N+1)!} ||p||_{L_2}^2 + \frac{\Delta^2}{n} ||p||_{\mathbf{H}^2}^2 \Big), \tag{2.20}$$

$$\max_{1 \le i \le M} \sup_{x \in \mathbf{R}^d} \tilde{\mathbf{E}} |u_N^n(t_i, x) - u(t_i, x)|^2 \le C e^{CT} \Big(\frac{(h_0 \Delta)^N}{(N+1)!} ||p||_{\mathbf{C}_b^0}^2 + \frac{\Delta^2}{n} ||p||_{\mathbf{C}_b^2}^2 \Big). \tag{2.21}$$

Constants B and C depend only on the coefficients b, σ, and h of the model and $h_0 := \sum_{l=1}^r \sup_{x \in \mathbf{R}^d} |h^l(x)|^2$.[3]

[3] Of course, B and C here are, in general, different from constants B and C in Theorem 2.2.

The sequence $\{u_N^n(t_i,x)\}_{1\le i\le M}$ gives an approximation to the UFD at all points of the time grid. This is a flexible and comparatively universal approximation. Many well-known numerical schemes for the Zakai equation can be obtained as particular cases of (2.18). In section 4 we will demonstrate this for two well-known algorithms: the explicit Euler scheme and the splitting-up method.

REMARK 2.3. *Analysis of the proofs of Theorems* 2.2–2.4 *shows that the wavelet type structure of our "global" basis* $\mathcal{M}=\cup_{i=1}^M\cup_{k=1}^\infty\{m_k^i\}$ *is of central importance.*

Specifically the following two properties of the basis are crucial:

(1) *The global basis* $\mathcal{M}$ *is a direct sum of "local" bases* $\mathcal{M}^i=\cup_{k=1}^\infty\{m_k^i\}$ *with nonoverlapping supports* (*Theorems* 2.3–2.5).

(2) *The functions* $m_k^i(s)$ *are smooth and* $\int_{t_{i-1}}^{t_i} m_k^i(s)ds=0$ *for* $k\ge 2$, $i=1,\dots,M$ (*see Theorem* 2.4).

The recursive version (2.18) of the spectral approximation of the unnormalized filtering density has one important disadvantage as compared to the one-step approximation (2.11). Indeed, to compute $u_N^n(t_i,x)$ we have to solve a certain number of equations from system (2.19). Although these equations are the same on every time interval and their coefficients do not involve the observation process y, the initial condition for the first equation of the system, $u_N^n(t_{i-1},x)$, does. This fact of course rules out off-line computation of the Fourier–Hermite coefficients $\varphi_\alpha(t,x)$, which is one of the important objectives of our study. For this reason, we present below a modification of the expansion (2.18) which admits off-line computations. Loosely speaking, the idea is to expand the initial condition for the first equation of (2.19) in a Fourier series as a function of spatial variable x, $u_N^n(t_{i-1},x)=\sum_l c_l e_l(x)$, and to exploit the obvious relation

$$\varphi_\alpha(t_i,x,u(t_{i-1},x))=\sum_l c_l\varphi_\alpha(t_i,x,e_l).$$

Note that the functions $\varphi_\alpha(t_i,x,e_l)$ can be computed off-line.

THEOREM 2.5. *Let* $\{e_l\}=\{e_l(x)\}_{l\ge1}$, $e_l\in\cap_n\mathbf{H}^n$, *be a CONS in* $L_2(\mathbf{R}^d)$ *and* $(\cdot,\cdot)$ *be the inner product in that space. Suppose that assumptions* (A1)–(A3) *hold and* $\{m_k^i\}$ *are given by* (2.17). *Consider the following system of equations:*

(2.22)
$$\frac{\partial\varphi_\alpha(s,x,g)}{\partial s}=\mathcal{L}^*\varphi_\alpha(s,x,g)+\sum_{k,l}\alpha_k^l m_k(s)h^l(x)\varphi_{\alpha(k,l)}(s,x,g),\qquad 0<s\le\Delta,$$
$$\varphi_\alpha(0,x,g)=g(x)1_{\{|\alpha|=0\}}.$$

Define $q_{\alpha k}^l:=(\varphi_\alpha(\Delta,\cdot,e_k),e_l)$ *and then by induction*

(2.23)
$$\psi_l(0,N,n):=(p,e_l),$$
$$\psi_l(i,N,n):=\sum_{\alpha\in J_N^n}\sum_k\frac{1}{\sqrt{\alpha!}}\psi_k(i-1,N,n)q_{\alpha k}^l\xi_\alpha^i.$$

Then

(2.24)
$$u_N^n(t_i,x)=\sum_l\psi_l(i,N,n)e_l(x),\ 0\le i\le M\quad(P-\ a.s.).$$

Now we can describe an approximation algorithm which stems from Theorem 2.5.

(1) Before the observations become available, (a) choose a finite collection $\{e_l\}_{1\le l\le\kappa}$;

(b) compute $\psi_l(0,N,n,\kappa) := (p,e_l),\ 1 \le l \le \kappa$, where p is the initial density;
(c) for all $\alpha \in J_N^n$ and $l = 1,\dots,\kappa$ compute $\varphi_\alpha(\Delta,x,e_l)$;
(d) compute $q_{\alpha k}^l = (\varphi_\alpha(\Delta,\cdot,e_k),e_l)$.
(2) On the ith step, when the observations become available,
(a) compute ξ_α^i;
(b) compute $\psi_l(i,N,n,\kappa) := \sum_{\alpha\in J_N^n}\sum_{k=1}^{\kappa}(1/\sqrt{\alpha!})\psi_k(i-1,N,n,\kappa)q_{\alpha k}^l\xi_\alpha^i$ for $l = 1,\dots,\kappa$;
(c) compute

$$u_N^{n,\kappa}(t_i,x) := \sum_{l=1}^{\kappa}\psi_l(i,N,n,\kappa)e_l(x). \tag{2.25}$$

.

We refer to this algorithm as the spectral separating scheme (S^3).

REMARK 2.4. *The amount of on-line operations and the amount of information that has to be stored in each step of* S^3 *do not depend on the number of steps to be performed. Also in contrast to the standard time-discretization schemes for the Zakai equation,* S^3 *does not require computing of the UFD at all the grid points* t_i, $i = 1,\dots,M$. *Specifically, step* 2(c) *of the algorithm can be omitted on any subset of time grid points (e.g., everywhere except the final point* t_M*). Note that computing of* (2.25) *is time consuming since it has to be done at all points of the space mesh.*

The truncation of the basis $\{e_l\}$ assumed in the above algorithm is necessary for computational reasons. Obviously it adds an extra error to (2.20). It is also clear that the error depends on the choice of the basis $\{e_l\}$ and is very much related to the particular numerical scheme used to solve (2.22).

It is beyond the scope of this work to study the above questions in detail, so we restrict ourselves to one particular case.

THEOREM 2.6. *Suppose that* $\{e_l\}$ *is the Hermite basis in* $L_2(\mathbf{R}^d)$ [16].

Let $0 = t_0 < \cdots < t_M = T$ *be a uniform partition of* $[0,T]$ *and* $u_N^n(t_i,x)$ *and* $u_N^{n,\kappa}(t_i,x)$ *be defined by* (2.18) *and* (2.25)*, respectively. Assume that* (A1)–(A3) *hold and in addition the initial density* p *and all its derivatives decay faster than any negative power of* $|x|$ *as* $|x| \to \infty$.

Then for any positive integer γ *there is a real number* $C_\gamma > 0$ *depending only on* γ *and the parameters* σ, b, p, *and* d *of the model such that*

$$\max_{1\le i\le M}\sqrt{\tilde{\mathbf{E}}||u_N^n(t_i,\cdot) - u_N^{n,\kappa}(t_i,\cdot)||_{L_2}^2} \le \frac{MC_\gamma(e^{C_\gamma T}-1)}{T\kappa^{\gamma-1/2}}. \tag{2.26}$$

This theorem shows that for sufficiently smooth initial condition p and with appropriate choice of the basis $\{e_l\}$, the error due to the truncation of the basis decays faster than any power of κ; i.e., our approximation is of a "spectral quality" (see, e.g., [15]).

REMARK 2.5. *The overall error of approximation for the spectral separating scheme follows from* (2.20) *and* (2.26) *and is given by*

$$\max_{1\le i\le M}\tilde{\mathbf{E}}||u(t_i,\cdot) - u_N^{n,\kappa}(t_i,\cdot)||_{L_2}^2 \le C\Big(\frac{(h_0\Delta)^N}{(N+1)!} + \frac{\Delta^2}{n} + \frac{C(\gamma)}{\Delta^2\kappa^{2\gamma-1}}\Big), \tag{2.27}$$

where C *is a constant depending on the parameters of the model (including the initial density* p *and the length of the time interval* T*) and it is assumed that the Wiener integrals* $\int_{t_{i-1}}^{t_i} m_k(t)dy^l(t)$ *are computed exactly. If* $n = 1$, *then only increments of the*

observation process are needed at each step and the computation of the integrals does not introduce any additional error. For $n > 1$, the integrals $\int_{t_{i-1}}^{t_i} m_k(t)dy^l(t),\ k > 1$, can be reduced to Riemann integrals and then approximated by subdividing the interval $[t_{i-1}, t_i]$ with some step $\delta << \Delta$. The error of the corresponding approximation will depend on the new asymptotic parameter δ. Still, formula (2.27) *implies that, in the limit $\lim_{\Delta\to 0}\lim_{\kappa\to\infty}$, the schemes with $n = 1$ and $n > 1$ have the same rate of convergence.*

REMARK 2.6. *Another approximation based on the Haar basis was proposed by Budhiraja and Kallianpur* [5]. *The approximation in* [5] *converges when $N \uparrow \infty$ and $\Delta \downarrow 0$. For computational purposes, though, it can be difficult to take arbitrarily large values of N because of the growing complexity and possible numerical instability. On the other hand, it follows from* (2.27) *that the spectral separating scheme converges in the limit $\lim_{\Delta\to 0}\lim_{\kappa\to\infty}$ for every $N \geq 1$ and the rate of convergence is the same for all $N \geq 2$.*

We remark that Theorems 2.1–2.6 can be extended to the case of time-dependent coefficients. Theorems 2.1–2.6 hold if the coefficients belong to the Hölder space $\mathbf{C}^{2+\alpha}(\mathbf{R}^d)$ for each t. The generalization is straightforward yet a bit cumbersome. Theorems 2.1–2.2 can be carried over to the case of correlated noises without many changes in the proofs [29, 31]. On the other hand the error estimates in the latter case are more delicate.

By no means is our approach a universal one. For example, it requires advanced knowledge of the parameters of the system, which are not always readily available. Also, it is not clear if it could be extended to the case of a non-Markov state process.

3. Proofs. In this section we will prove Theorems 2.2, 2.4, and 2.5. Everywhere C stands for a positive constant depending only on the parameters of the system; its actual value may be different in different places.

We introduce the following notation:

$\{T_s\}_{s\geq 0}$, the semigroup generated by the operator $\mathcal{L}^*$;

s^k, the the ordered set $(s_1, \dots, s_k)$; $ds^k := ds_1 \dots ds_k$;

$F(t; s^k; x) := T_{t-s_k} h T_{s_k - s_{k-1}} \dots h T_{s_1} p(x),\ k \geq 1$;

$\int^{(k)}(\cdots)ds^k := \int_0^t \int_0^{s_k} \dots \int_0^{s_2} (\cdots) ds_1 \dots ds_k$.

When $r = 1$, each multiindex $\alpha = (\alpha_1, \alpha_2, \dots)$ of length $|\alpha| = k$ can be identified with a vector $K_\alpha = (i_1^\alpha, \dots, i_k^\alpha)$, where $i_1^\alpha \leq i_2^\alpha \leq \dots \leq i_k^\alpha$. The first entry i_1^α of K_α is the number of the first nonzero element of α. The second entry i_2^α is equal to i_1^α if that first nonzero element $\alpha_{i_1^\alpha}$ is greater than 1; otherwise i_2^α is the number of the second nonzero element and so on. As a result, if $\alpha_j > 0$, then exactly α_j entries of the vector K_α are equal to j. We will call this vector the *characteristic set* of multiindex α. For example, if $\alpha = (0, 1, 0, 2, 3, 0, \dots)$, then nonzero elements are $\alpha_2 = 1,\ \alpha_4 = 2,\ \alpha_5 = 3$, and the characteristic set is $(2, 4, 4, 5, 5, 5)$. A similar construction is possible for general $r > 1$. In the future, when there is no danger of confusion, we will omit the upper index in i (i.e., write i_j rather than i_j^α).

Let $\mathcal{P}^k$ be the permutation group of the set $\{1, \dots, k\}$. For a given $\alpha \in J$ with $|\alpha| = k$ and the characteristic set $(i_1, \dots, i_k)$ $(r = 1)$ define

$$E_\alpha(s^k) := \sum_{\sigma\in\mathcal{P}^k} m_{i_1}(s_{\sigma(1)}) \cdots m_{i_k}(s_{\sigma(k)}).$$

Proof of Theorem 2.2. We will prove inequality (2.12); the other can be proven in a similar way.

Set

$$u_N(t,x) := \sum_{|\alpha|\le N} \frac{\varphi_\alpha(t,x)\xi_\alpha}{\sqrt{\alpha!}}.$$

Suppose that we know that

$$\tilde{\mathbf{E}}||u(t,\cdot) - u_N(t,\cdot)||^2_{L_2} \le \frac{(h_0 t)^{N+1}}{(N+1)!} e^{Ct} ||p||^2_{L_2} \tag{3.1}$$

and

$$\tilde{\mathbf{E}}||u_N(t,\cdot) - u^n_N(t,\cdot)||^2_{L_2} \le C\frac{t^3}{n} e^{Ct} ||p||^2_{\mathbf{H}^2}. \tag{3.2}$$

Then (2.12) will follow immediately from the inequality $(a+b)^2 \le 2(a^2+b^2)$.

The problem is thus to prove (3.1) and (3.2). To simplify the presentation, we assume that $r = 1$.

Proof of (3.1). We will use the following results:

$$\sum_{|\alpha|=k} \frac{\varphi^2_\alpha(t,x)}{\alpha!} = \int^{(k)} |F(t;s^k;x)|^2 ds^k, \tag{3.3}$$

where φ_α is the solution of (2.6) with any CONS $\{m_k\}$, and

$$||T_s f||^2_{L_2} \le e^{Cs} ||f||^2_{L_2}. \tag{3.4}$$

The first equality is established in the appendix, Proposition A.1 (see also [30]); inequality (3.4) is a standard fact (see [25]).

Since ξ_α are uncorrelated under $\tilde{\mathrm{P}}$, we have

$$\tilde{\mathbf{E}}|u(t,x) - u_N(t,x)|^2 = \sum_{k>N} \sum_{|\alpha|=k} \frac{\varphi^2_\alpha(t,x)}{\alpha!} = \sum_{k>N} \int^{(k)} |F(t;s^k;x)|^2 ds^k.$$

Then by the Fubini theorem

$$\begin{aligned}\tilde{\mathbf{E}}||u(t,\cdot) - u_N(t,\cdot)||^2_{L_2} &= \sum_{k>N} \int^{(k)} \Big(\int_{\mathbf{R}^d} |F(t;s^k;x)|^2 dx \Big) ds^k \\ &= \sum_{k>N} \int^{(k)} ||F(t,s^k,\cdot)||^2_{L_2} ds^k.\end{aligned}$$

Since h is bounded, it follows from the definition of F and (3.4) that

$$\begin{aligned}||F(t;s^k;\cdot)||^2_{L_2} &\le h_0 e^{C(t-s_k)} ||T_{s_k - s_{k-1}} h \dots h T_{s_1} p||^2_{L_2} \\ &\le \dots \le h_0^k e^{t - s_k + s_k - \dots + s_2 - s_1 + s_1} ||p||^2_{L_2} = h_0^k e^{Ct} ||p||^2_{L_2}.\end{aligned}$$

Finally, from $\int^{(k)} ds^k = t^k/k!$, we conclude that

$$\begin{aligned}\tilde{\mathbf{E}}||u(t,\cdot) - u_N(t,\cdot)||^2_{L_2} &\le e^{Ct} \sum_{k>N} \frac{(th_0)^k}{k!} \\ &\le \frac{(th_0)^{N+1}}{(N+1)!} e^{(C+h_0)t},\end{aligned}$$

and (3.1) follows. Note that it holds for any CONS $\{m_k\}$.

Proof of (3.2). If α is a multiindex with $|\alpha| = k$ and the characteristic set $(i_1^\alpha, \dots, i_k^\alpha)$, then $i_k^\alpha = d(\alpha)$, the order of α, and so the set J_N^n can be described as $\{\alpha \in J : |\alpha| \le N;\ i_{|\alpha|}^\alpha \le n\}$. Thus

$$\tilde{\mathbf{E}}|u_N^n(t,x) - u_N(t,x)|^2 = \sum_{l=n+1}^{\infty} \sum_{k=1}^{N} \sum_{|\alpha|=k; i_k^\alpha = l} \frac{\varphi_\alpha^2(t,x)}{\alpha!}.$$

The problem is thus to estimate $\sum_{l=n+1}^{\infty} \sum_{k=1}^{N} \sum_{|\alpha|=k; i_k^\alpha=l} \frac{\varphi_\alpha^2(t,x)}{\alpha!}$.

By Proposition A.1 (see also [30]) the corresponding solution φ_α of (2.6) can be written as

$$\varphi_\alpha(t,x) = \int^{(k)} F(t; s^k; x) E_\alpha(s^k) ds^k. \tag{3.5}$$

Note that we can also write

$$E_\alpha(s^k) = \sum_{j=1}^{k} m_{i_k}(s_j) E_{\alpha(i_k)}(s_j^k),$$

where s_j^k denotes the same set $(s_1, \dots, s_k)$ with omitted s_j (e.g., $s_1^k = (s_2, \dots, s_k)$) and $\alpha(i_k)$ is the multiindex with this characteristic set $(i_1, \dots, i_{k-1})$ (cf. (2.7); recall that $r = 1$).

This allows us to write (3.5) as

$$\varphi_\alpha(t,x) = \sum_{j=1}^{k} \int^{(k-1)} \Big(\int_{s_{j-1}}^{s_{j+1}} F(t; s^k; x) m_{i_k}(s_j) ds_j \Big) E_{\alpha(i_k)}(s_j^k) ds_j^k, \tag{3.6}$$

where $s_0 := 0;\ s_{k+1} := t$. (We just change the order of integration in the multiple integral.)

Denote

$$M_k(s) := \frac{\sqrt{2t}}{\pi(k-1)} \sin\Big(\frac{\pi(k-1)}{t} s\Big), \qquad k > 1,\ 0 \le s \le t,$$

and also $F_j := \frac{\partial F(t; s^k; x)}{\partial s_j}$. Then, as long as $i_k = l > 1$, we can integrate by parts the inner integral on the right of (3.6) to get

$$\begin{aligned} &\int_{s_{j-1}}^{s_{j+1}} F(t; s^k; x) m_l(s_j) ds_j \\ &\quad = F(t; s^k; x) M_l(s_j) \Big|_{s_j = s_{j-1}}^{s_j = s_{j+1}} - \int_{s_{j-1}}^{s_{j+1}} F_j(t, s^k, x) M_l(s_j) ds_j. \end{aligned}$$

For each j, let us rename the remaining variables s_j^k in (3.6) as follows: $t_i := s_i, i \le j-1;\ t_i := s_{i+1}, i > j-1$, or, symbolically, $t^{k-1} := s_j^k$. We will set $t_0 := 0$, $t_k := t$ and denote by $t^{k-1,j}, j = 1, \dots, k-1$, the set t^{k-1} in which t_j is repeated twice (e.g., $t^{k-1,1} = (t_1, t_1, \dots, t_{k-1})$, etc.); also $t^{k-1,0} := (t_0, t_1, t_2, \dots, t_{k-1})$, $t^{k-1,k} := (t_1, \dots, t_{k-1}, t_k)$.

Then

$$F(t;s^k;x)M_l(s_j)\Big|_{s_j=s_{j-1}}^{s_j=s_{j+1}}$$
$$= F(t;t^{k-1,j};x)M_l(t_j) - F(t;t^{k-1,j-1};x)M_l(t_{j-1}), \qquad j=1,\dots,k.$$

As a result, since $M_l(t_0)=M_l(t_k)=0$ (and this is the only place where the choice of $\{m_k\}$ really makes the difference), all these terms will cancel out as we sum over j. What remains can be written as

$$\int^{(k-1)} f_l(t;t^{k-1};x)E_{\alpha(l)}(t^{k-1})dt^{k-1},$$

where

$$f_l(t;t^{k-1};x) = -\int_0^{t_1} F_1(t;\tau,t^{k-1};x)M_l(\tau)d\tau$$
$$-\sum_{j=2}^{k-1}\int_{t_{j-1}}^{t_j} F_j(t;\dots,t_{j-1},\tau,t_j,\dots;x)M_l(\tau)d\tau$$
$$-\int_{t_{k-1}}^{t_k} F_k(t;t^{k-1},\tau;x)M_l(\tau)d\tau.$$

Then, since $|\alpha(i_{|\alpha|})| = |\alpha|-1$, $\alpha! \geq \alpha(i_{|\alpha|})!$, we get

$$\sum_{|\alpha|=k;i_k^\alpha=l} \frac{\varphi_\alpha^2(t,x)}{\alpha!}$$
$$= \sum_{|\alpha|=k;i_k^\alpha=l} \Big(\frac{1}{\sqrt{\alpha!}}\int^{(k-1)} f_l(t;t^{k-1};x)E_{\alpha(l)}(t^{k-1})dt^{k-1}\Big)^2$$
$$\leq \sum_{|\beta|=k-1} \Big(\frac{1}{\sqrt{\beta!}}\int^{(k-1)} f_l(t;t^{k-1};x)E_{\beta}(t^{k-1})dt^{k-1}\Big)^2,$$

and arguments similar to those used in the proof of Proposition A.1 show that the last expression is equal to

$$\int^{(k-1)} |f_l(t;t^{k-1};x)|^2 dt^{k-1}. \tag{3.7}$$

Direct computations yield

$$F_j(t,s^k,x) = T_{t-s_k}h\dots T_{s_{j+1}-s_j}h\mathcal{L}^*T_{s_j-s_{j-1}}\dots T_{s_1}p(x)$$
$$-T_{t-s_k}h\dots\mathcal{L}^*T_{s_{j+1}-s_j}hT_{s_j-s_{j-1}}\dots T_{s_1}p(x).$$

Since $\mathcal{L}^*$ is a continuous operator from $\mathbf{H}^2$ to L_2, it follows from (3.4) and a similar inequality for $\mathbf{H}^2$-norms that

$$||F_j(t;s^k;\cdot)||_{L_2}^2 \leq e^{Ct}C^k||p||_{\mathbf{H}^2}^2.$$

Then the definition of f_l and obvious inequalities

$$(a_1+\dots+a_k)^2 \leq k(a_1^2+\dots+a_k^2)$$

and

$$\left(\int_0^x f(y)dy\right)^2 \le x\int_0^x (f(y))^2 dy$$

imply

$$\begin{aligned}||f_l(t;t^{k-1};\cdot)||_{L_2}^2 &\le kC^k e^{Ct}||p||_{\mathbf{H}^2}^2 t\int_0^t (M_l(s))^2 ds\\ &\le \frac{kC^k t^3 e^{Ct}}{(l-1)^2}||p||_{\mathbf{H}^2}^2;\end{aligned}$$

so, since $\int^{(k-1)} dt^{k-1} = t^{k-1}/(k-1)!$, (3.7) and the last inequality yield

$$\sum_{|\alpha|=k, i_k^\alpha = l} \frac{||\varphi_\alpha^2(t,\cdot)||_{L_2}^2}{\alpha!} \le e^{Ct}||p||_{\mathbf{H}^2}^2 t^3 \frac{kC^k}{(l-1)^2(k-1)!}.$$

Now we collect everything to get

$$\begin{aligned}\tilde{\mathbf{E}}||u_N(t,\cdot) - u_N^n(t,\cdot)||_{L_2}^2 &= \sum_{l\ge n+1}\sum_{k=1}^N \sum_{|\alpha|=k; i_k^\alpha = l} \frac{\varphi_\alpha^2(t,x)}{\alpha!}\\ &\le Ct^3 e^{Ct}\Big(\sum_{k\ge 1}\frac{k(Ct)^{k-1}}{(k-1)!}\Big)\sum_{l\ge n}\frac{1}{l^2} \le C\frac{t^3}{n}e^{Ct}||p||_{\mathbf{H}^2}^2.\end{aligned}$$

This completes the proof of (3.2) and the theorem as a whole. □

Proof of Theorem 2.4. We again prove only the first inequality.

First of all notice that time homogeneity of (2.15) and the special choice of $\{m_k^i\}$ as $m_k^i(s) = m_k(s - t_{i-1})$ imply

$$\varphi_\alpha^i(\Delta, x) = \varphi_\alpha^i(t_i, x, u(t_{i-1},\cdot))$$

(see (2.15) and (2.19)). Then by Fubini's theorem and Theorem 2.3 and due to linearity of system (2.15),

$$\begin{aligned}\tilde{\mathbf{E}}||u_N^n(t_i,\cdot) - u(t_i,\cdot)||_{L_2}^2 &= \sum_{\alpha\in J_N^n}\frac{1}{\alpha!}\tilde{\mathbf{E}}||\varphi_\alpha^i(t_i,\cdot,u_N^n(t_{i-1},\cdot) - u(t_{i-1},\cdot))||_{L_2}^2\\ &\quad + \sum_{\alpha\notin J_N^n}\frac{1}{\alpha!}\tilde{\mathbf{E}}||\varphi_\alpha^i(t_i,\cdot,u(t_{i-1},\cdot))||_{L_2}^2\\ &\le \sum_{\alpha\in J}\frac{1}{\alpha!}\tilde{\mathbf{E}}||\varphi_\alpha^i(t_i,\cdot,u_N^n(t_{i-1},\cdot) - u(t_{i-1},\cdot))||_{L_2}^2\\ &\quad + \sum_{\alpha\notin J_N^n}\frac{1}{\alpha!}\tilde{\mathbf{E}}||\varphi_\alpha^i(t_i,\cdot,u(t_{i-1},\cdot))||_{L_2}^2.\end{aligned} \tag{3.8}$$

By Theorem 2.3 and linearity of equation (2.4), we have

$$\begin{aligned}&\sum_{\alpha\in J}\frac{1}{\alpha!}\tilde{\mathbf{E}}||\varphi_\alpha^i(t_i,\cdot,u_N^n(t_{i-1},\cdot) - u(t_{i-1},\cdot))||_{L_2}^2\\ &\qquad = \tilde{\mathbf{E}}||U(t_i,x;u_N^n(t_{i-1},\cdot) - u(t_{i-1},\cdot))||_{L_2}^2,\end{aligned} \tag{3.9}$$

where $U(t,x;u_N^n(t_{i-1},\cdot)-u(t_{i-1},\cdot))$ is the solution of

$$\begin{aligned} dv(t,x) &= \mathcal{L}^* v(t,x)dt + h^l(x)v(t,x)dy^l(t), \quad t\in(t_{i-1},t_i],\\ v(t_{i-1},x) &= u_N^n(t_{i-1},x)-u(t_{i-1},x). \end{aligned}$$

It is a standard fact that under assumptions (A1) and (A3),

$$\tilde{\mathbf{E}}||U(t_i,\cdot;u_N^n(t_{i-1},\cdot)-u(t_{i-1},\cdot))||_{L_2}^2 \le e^{C\Delta}\tilde{\mathbf{E}}||u_N^n(t_{i-1},\cdot)-u(t_{i-1},\cdot)||_{L_2}^2 \tag{3.10}$$

(see, e.g., [34]).

Repeating the same arguments as in the proof of Theorem 2.2, one can check that

$$\begin{aligned} &\sum_{\alpha\notin J_N^n}\frac{1}{\alpha!}\tilde{\mathbf{E}}||\varphi_\alpha^i(t_i,\cdot,u(t_{i-1},\cdot))||_{L_2}^2\\ &\qquad \le Ce^{C\Delta}\Big(\frac{(h_0\Delta)^{N+1}}{(N+1)!}\tilde{\mathbf{E}}||u(t_{i-1},\cdot)||_{L_2}^2+\frac{\Delta^3}{n}\tilde{\mathbf{E}}||u(t_{i-1},\cdot)||_{\mathbf{H}^2}^2\Big). \end{aligned} \tag{3.11}$$

Finally, we use the inequalities

$$\tilde{\mathbf{E}}||u(t_{i-1},\cdot)||_{L_2}^2 \le e^{CT}||p||_{L_2}^2 \tag{3.12}$$

and

$$\tilde{\mathbf{E}}||u(t_{i-1},\cdot)||_{\mathbf{H}^2}^2 \le e^{CT}||p||_{\mathbf{H}^2}^2. \tag{3.13}$$

These inequalities are similar to (3.10) and can also be found in [34].

If we now denote $\tilde{\mathbf{E}}||u_N^n(t_i,\cdot)-u(t_i,\cdot)||_{L_2}^2$ by ε_i, then, combining (3.8)–(3.13), we arrive at

$$\varepsilon_i \le \left(\varepsilon_{i-1}+Ce^{CT}\Big(\frac{(h_0\Delta)^{N+1}}{(N+1)!}||p||_{L_2}^2+\frac{\Delta^3}{n}\tilde{\mathbf{E}}||p||_{\mathbf{H}^2}^2\Big)\right)e^{C\Delta},$$

and since $\varepsilon_0=0$, the statement of the theorem follows from the discrete Gronwall lemma. □

Proof of Theorem 2.5. By construction, $u_N^n(t_i,\cdot)\in L_2(\mathbf{R}^d)$ (P − a.s.), so

$$u_N^n(t_i,x)=\sum_{l\ge1}\psi_l(i,N,n)e_l(x) \quad (\mathrm{P}-\text{a.s.})$$

with some $\psi_l(i,N,n)$. Then all we have to do is to establish (2.23), which means

$$\sum_{\alpha\in J_N^n}\sum_k\frac{1}{\sqrt{\alpha!}}\psi_k(i-1,N,n)q_{\alpha k}^l\xi_\alpha^i=(u_N^n(t_i,\cdot),e_l). \tag{3.14}$$

We will prove this by induction. For $i=0$, $\psi_l(0,N,n)=(u_N^n(t_0,\cdot),e_l)$ by definition. Assume that $u_N^n(t_{i-1},x)=\sum_l\psi_l(i-1,N,n)e_l(x)$ for some $i\ge1$.

The proof of Theorem 2.2 shows that operator $g\mapsto\varphi_\alpha(t_i,\cdot,g)$ is continuous and linear from $L_2(\mathbf{R}^d)$ to $L_2(\mathbf{R}^d)$ for all $\alpha\in J$, where $\varphi_\alpha(t_i,\cdot,g)$ is the solution of (2.22). Then

$$\begin{aligned} \sum_k\psi_k(i-1,N,n)q_{\alpha k}^l &= \sum_k\psi_k(i-1,N,n)(\varphi_\alpha(\Delta,\cdot,e_k),e_l)\\ &= \Big(\varphi_\alpha\Big(\Delta,\cdot,\sum_k\psi_k(i-1,N,n)e_k\Big),e_l\Big), \end{aligned}$$

and by an induction assumption the right-hand side of the above formula is equal to

$$(\varphi_\alpha(\Delta, \cdot, u_N^n(t_{i-1}, \cdot)), e_l).$$

On the other hand, comparing (2.19) and (2.22) we conclude that

$$\varphi_\alpha(\Delta, x, u_N^n(t_{i-1}, \cdot)) = \varphi_\alpha^i(\Delta, x).$$

As a result,

$$\sum_{\alpha \in J_N^n} \sum_k \frac{1}{\sqrt{\alpha!}} \psi_k(i-1, N, n) q_{\alpha k}^l \xi_\alpha^i = \Big(\sum_{\alpha \in J_N^n} \frac{1}{\sqrt{\alpha!}} \varphi_\alpha^i(t_i, \cdot, u_N^n(t_{i-1}, \cdot)) \xi_\alpha^i, e_l \Big),$$

and by (2.18) this is equal to $(u_N^n(t_i, \cdot), e_l)$. This completes the proof of (3.14) and the theorem as a whole. □

REMARK 3.1. *Analysis of the proof shows that the result (with obvious modifications) is also true for the exact solution $u(t_i, x)$.*

Proof of Theorem 2.6. In what follows, C_γ denotes a constant depending on γ and (maybe) the parameters of the model. As before, C is a constant depending only on the parametes of the model. Values of C and C_γ may be different in different places.

If $d = 1$, then

$$e_l(x) = \frac{1}{\sqrt{(2\pi)^{1/2} l!}} e^{-x^2/4} H_l(x), \tag{3.15}$$

where H_l is the lth Hermite polynomial (2.5) [15, 16].

For $d > 1$, the elements of the basis are

$$e_l(x_1, \dots, x_d) = e_{l_1}(x_1) \dots e_{l_d}(x_d),$$

where $l_i \geq 0$ and e_{l_i} are given by (3.15), $i = 1, \dots, d$ [16]. The system $\{e_l\}$ is thus indexed by the set of d-dimensional multiindices $l = (l_1, \dots, l_d)$ ordered in some natural way. We will say that $l \leq \kappa$ if $\max_{1 \leq i \leq d} l_i \leq \kappa$.

To simplify the presentation, we assume from now on that $d = 2$. Then $l = (l_1, l_2)$.

Direct computations show that e_{l_i} satisfies

$$\mathcal{A}_i e_{l_i} = (l_i + 1) e_{l_i}, \quad i = 1, 2,$$

where operator $\mathcal{A}_i$ is defined by

$$\mathcal{A}_i f(x) = -\frac{\partial^2 f(x)}{\partial x_i^2} + \frac{2 + x_i^2}{4} f(x).$$

As a result, $\mathcal{A}_2 \mathcal{A}_1 e_l(x_1, x_2) = (l_1 + 1) \mathcal{A}_2 (e_{l_1}(x_1) e_{l_2}(x_2)) = (l_1 + 1)(l_2 + 1) e_l(x_1, x_2)$. This means that if f and all its derivative decay fast enough, then

$$|(f, e_l)| \leq \frac{||\mathcal{A}_1 \mathcal{A}_2 f||_{L_2}}{(l_1 + 1)(l_2 + 1)} \leq \dots \leq \frac{||(\mathcal{A}_1 \mathcal{A}_2)^\gamma f||_{L_2}}{(l_1 + 1)^\gamma (l_2 + 1)^\gamma}. \tag{3.16}$$

If $\mathbf{H}^s(r)$, $s, r \in \mathbf{R}$, is the weighted Sobolev space $W_2^s(r, \mathbf{R}^2)$ ([34]; see also [16]), then definition of $\mathcal{A}_i$ implies that

$$||(\mathcal{A}_1 \mathcal{A}_2)^\gamma f||_{L_2} \leq C_\gamma ||f||_{\mathbf{H}^{4\gamma}(4\gamma)},$$

and (3.16) becomes

$$|(f,e_k)| \le \frac{C_\gamma ||f||_{\mathbf{H}^{4\gamma}(4\gamma)}}{(l_1+1)^\gamma (l_2+1)^\gamma}. \tag{3.17}$$

Introduce the following notation:
$|||f||| := \sqrt{\tilde{\mathbf{E}}||f||^2_{L_2}}$,
$\varepsilon_i := |||u^n_N(t_i,\cdot) - u^{n,\kappa}_N(t_i,\cdot)|||$,
Π^κ, the L_2-orthogonal projection on the subspace generated by $e_l,\ l \le \kappa$,
$V^{n,i}_N(f)$, the operator $f \mapsto \sum_{\alpha\in J^n_N} \varphi_\alpha(\Delta,\cdot,f)\xi^i_\alpha;\ f \in L_2(\Omega,\mathcal{F}^y_{t_{i-1}},\tilde{\mathrm{P}})$ and $U^i(f) := V^{\infty,i}_\infty$. Since $V^{n,i}_N(f)$ is the $L_2(\Omega,\tilde{\mathrm{P}})$-orthogonal projection of $U^i(f)$ on the subspace generated by $\{\xi_\alpha,\ \alpha \in J^n_N\}$,

$$|||V^{n,i}_N(f)||| \le |||U^i(f)|||. \tag{3.18}$$

Below we will be dealing with a fixed set (n,i,N) and to simplify notation will write V instead of $V^{n,i}_N$. We also omit the dot in $u^n_N(t_i,\cdot)$, etc.

Since the coefficients of the model are time independent,

$$u^n_N(t_i) = V(u^n_N(t_{i-1})), \qquad u^{n,\kappa}_N(t_i) = \Pi^\kappa V(u^{n,\kappa}_N(t_{i-1})).$$

The second equality follows from (2.25), the definition of $\psi_l(i,N,n,\kappa)$, and the linearity of the map $f \mapsto \varphi_\alpha(\Delta,f)$. Then by the triangle inequality

$$\begin{aligned}\varepsilon_i \le {} & |||\Pi^\kappa V(u^n_N(t_{i-1})) - \Pi^\kappa V(u^{n,\kappa}_N(t_{i-1}))||| \\ & + |||u^n_N(t_i) - \Pi^\kappa u^n_N(t_i)|||.\end{aligned} \tag{3.19}$$

By the definition of Π^κ,

$$\begin{aligned}&|||\Pi^\kappa V(u^n_N(t_{i-1})) - \Pi^\kappa V(u^{n,\kappa}_N(t_{i-1}))||| \\ &\quad \le |||V(u^n_N(t_{i-1})) - V(u^{n,\kappa}_N(t_{i-1}))||| \le e^{C\Delta}\varepsilon_{i-1},\end{aligned} \tag{3.20}$$

$\Delta = t_i - t_{i-1}$, where the last inequality follows from (3.18) and (3.10).

Under the assumptions of the theorem it is easy to show, using the standard estimates from [25] or [34], that for any $i = 1,\dots,M$, $u^n_N(t_i) \in \cap_s \mathbf{H}^s(r)$ $(\mathrm{P}-\text{a.s.})$ for any $r \in \mathbf{R}$. In addition,

$$\sum_{\alpha\in J^n_N} \frac{\tilde{\mathbf{E}}||\varphi^i_\alpha(\Delta)||^2_{\mathbf{H}^\gamma(\gamma)}}{\alpha!} \le e^{C_\gamma T}||p||^2_{\mathbf{H}^\gamma(\gamma)} \tag{3.21}$$

for any positive integer γ, where $\varphi^i_\alpha(\Delta) = \varphi_\alpha(\Delta,u(t_{i-1}))$. As a result, from (2.18), (3.17), (3.21), and the obvious estimates $\sum_{j>\kappa} 1/(j+1)^\gamma \le C_\gamma/(\kappa+1)^{\gamma-1} \le C_\gamma/\kappa^{\gamma-1}$ (valid for $\gamma > 1$), we conclude that

$$\begin{aligned}|||u^n_N(t_{i-1}) - \Pi^\kappa u^n_N(t_{i-1})|||^2 &= \sum_{\alpha\in J^n_N}\sum_{l>\kappa} \frac{\tilde{\mathbf{E}}(\varphi^i_\alpha(\Delta),e_l)^2}{\alpha!} \\ &\le 2\sum_{\alpha\in J^n_N}\Big(\sum_{l_1>\kappa,l_2\ge 0}\frac{1}{(l_1+1)^{2\gamma}(l_2+1)^{2\gamma}}\Big)\frac{\tilde{\mathbf{E}}||\varphi^i_\alpha(\Delta)||^2_{\mathbf{H}^{4\gamma}(4\gamma)}}{\alpha!} \\ &\le \frac{C_\gamma e^{C_\gamma T}}{\kappa^{2\gamma-1}}||p||^2_{\mathbf{H}^{4\gamma}(4\gamma)}.\end{aligned} \tag{3.22}$$

Combining (3.19), (3.20), and (3.22), we arrive at

$$\varepsilon_i \le e^{C\Delta}\varepsilon_{i-1} + \frac{C_\gamma e^{C_\gamma T}}{\kappa^{\gamma-1/2}}||p||_{\mathbf{H}^{4\gamma}(4\gamma)},$$

which by the discrete Gronwall lemma implies

$$\varepsilon_i \le \frac{C_\gamma(e^{C_\gamma T}-1)}{\Delta\ \kappa^{\gamma-1/2}}||p||_{\mathbf{H}^{4\gamma}(4\gamma)}.$$

Since $\Delta = T/M$ and by assumption $||p||_{\mathbf{H}^{4\gamma}(4\gamma)} \le C_\gamma$, (2.26) follows. □

4. Comparison with other algorithms and numerical simulations. The Wiener chaos approximations (2.8), (2.16) can be viewed as higher-order time-discretization schemes for the Zakai equation.

For $N \ge 2$, the rate of convergence of the Wiener chaos approximation u_N^n is $O(\Delta)$, where Δ is the time step (Theorem 2.4). This is similar to the rates of convergence of the splitting-up algorithm (see [26]) and the implicit Euler–Milstein scheme (see [18]) for the Zakai equation.

In fact, many well-known time-discretization schemes can be obtained as particular cases of the Wiener chaos approximation.

One of the simplest is the explicit Euler scheme. Take a uniform partition of the interval $[0,T]$ with step Δ. Then the explicit Euler approximation $u_i(x)$ to the Zakai equation is obtained from

(4.1)
$$u_0(x) = p(x), \quad u_i(x) = (1+\Delta\cdot\mathcal{L}^*)u_{i-1}(x) + \sum_{l=1}^{r} h^l(x)u_{i-1}(x)(y^l(t_i)-y^l(t_{i-1})).$$

Now we will derive the same result from Theorem 2.4. Take $n = N = 1$. Then set J_1^1 contains $r+1$ elements, and on each step we need to solve $r+1$ equations from (2.19):

$$\frac{\partial\varphi_0^i(s,x)}{\partial s} = \mathcal{L}^*\varphi_0^i(s,x), \quad 0 < s \le \Delta,$$
$$\varphi_0^i(0,x) = u_1^1(t_{i-1},x)$$

(for $|\alpha| = 0$);

$$\frac{\partial\varphi_l^i(s,x)}{\partial s} = \mathcal{L}^*\varphi_l^i(s,x) + \textstyle\sum_l \frac{h^l(x)}{\sqrt{\Delta}}\varphi_0^i(s,x), \quad 0 < s \le \Delta,$$
$$\varphi_l^i(0,x) = 0, \qquad l = 1,\dots,r$$

(for $|\alpha| = 1$ with $\alpha_1^l = 1$) and $u_1^1(t_0,x) = p(x)$.

We solve these equations using the explicit Euler scheme; the (approximate) solutions are then given by

$$\varphi_0^i(\Delta,x) = (1+\Delta\cdot\mathcal{L}^*)u_1^1(t_{i-1},x),$$

$$\varphi_l^i(\Delta,x) = h^l(x)\sqrt{\Delta}u_1^1(t_{i-1},x).$$

By definition,

$$\xi_l^i = \int_{t_{i-1}}^{t_i} m_1(s)dy^l(s) = \frac{y^l(t_i)-y^l(t_{i-1})}{\sqrt{\Delta}},$$

and by Theorem 2.4,

$$u_1^1(t_i,x)=\varphi_0^i(\Delta,x)+\sum_{l=1}^{r}\varphi_l^i(\Delta,x)\xi_l^i,$$

and this, due to the above relations, coincides with (4.1).

Another well-known algorithm for solving the Zakai equation (2.4) is the splitting-up approximation (see Bensoussan, Glowinski, and Rascanu [4]; Florchinger and LeGland [12]; etc.). For simplicity, we consider the case $r=1$. Take a uniform partition of $[0,T]$ with step Δ and let $\{T_t\}$ be the semigroup generated by operator $\mathcal{L}^*$ (or some approximation of that semigroup). Then the splitting-up approximation $u_i(x)$ to $u(t_i,x)$ is computed from the recursion

$$u_0(x)=p(x),\qquad u_i(x)=T_\Delta\exp([y(t_i)-y(t_{i-1})]h-0.5h^2\Delta)u_{i-1}(x). \tag{4.2}$$

Let us see how the same result can be obtained from Theorem 2.4. Set $n=1$, $N=\infty$. Then the set J_N^n consists of multiindices $\alpha=(k,0,0,\dots)$; the corresponding φ_α will be denoted by φ_k. We need to solve the following system:

$$\frac{\partial\varphi_0^i(s,x)}{\partial s}=\mathcal{L}^*\varphi_0^i(s,x),\quad 0<s\le\Delta,$$
$$\varphi_0^i(0,x)=u_\infty^1(t_{i-1},x)$$

(for $|\alpha|=0$);

$$\frac{\partial\varphi_k^i(s,x)}{\partial s}=\mathcal{L}^*\varphi_k^i(s,x)+k\tfrac{h(x)}{\sqrt{\Delta}}\varphi_{k-1}^i(s,x),\quad 0<s\le\Delta,$$
$$\varphi_k^i(0,x)=0,\quad k\ge1$$

(for $|\alpha|=k$) and $u_\infty^1(t_0,x)=p(x)$. An approximate solution to this system is given by

$$\varphi_k^i(t,x)=T_t\Big(\frac{th}{\sqrt{\Delta}}\Big)^k u_\infty^1(t_{i-1},\cdot)\ (x),\qquad k\ge0. \tag{4.3}$$

Indeed, for $k=0$, this is the exact solution (if T_t is exact); assuming (4.3) for some $k=n-1\ge0$, we get for $k=n$

$$\begin{aligned}\varphi_n^i(t,x)&=n\int_0^t T_{t-s}\frac{h}{\sqrt{\Delta}}\varphi_{n-1}^i(s,\cdot)(x)ds\\&=\frac{n}{\Delta^{n/2}}\int_0^t T_{t-s}h^nT_su_\infty^1(t_{i-1},\cdot)(x)s^{n-1}ds\\&\approx\frac{n}{\Delta^{n/2}}T_th^nu_\infty^1(t_{i-1},\cdot)(x)\int_0^t s^{n-1}ds\\&=T_t\Big(\frac{th}{\sqrt{\Delta}}\Big)^n u_\infty^1(t_{i-1},\cdot)\ (x),\end{aligned}$$

so (4.3) follows by induction. Note that, if $T_t(hf)(x)=hT_t(f)$ for all $f(x)$, it would be an exact solution.

Clearly, (4.3) implies that

$$\varphi_k^i(\Delta,x)=T_\Delta(h\sqrt{\Delta})^ku_\infty^1(t_{i-1},\cdot)\ (x),\qquad k\ge0.$$

It is also clear that

$$\xi_k^i = \frac{1}{\sqrt{k!}} H_k\Big(\frac{y(t_i) - y(t_{i-1})}{\sqrt{\Delta}}\Big), \tag{4.4}$$

and then by Theorem 2.4

$$\begin{aligned} u_\infty^1(t_i, x) &= T_\Delta u_\infty^1(t_{i-1}, \cdot)(x) + T_\Delta \sum_{k\geq 1} \frac{1}{k!} (h\sqrt{\Delta})^k H_k\Big(\frac{y(t_i) - y(t_{i-1})}{\sqrt{\Delta}}\Big) \\ &= T_\Delta \exp([y(t_i) - y(t_{i-1})]h - 0.5h^2\Delta) u_\infty^1(t_i, \cdot)(x). \end{aligned}$$

(The last equality follows from the well-known expansion

$$\exp(ax - 0.5x^2) = \sum_{k\geq 0} \frac{1}{k!} H_k(a) x^k$$

if we set $a = (y(t_i) - y(t_{i-1}))/\sqrt{\Delta},\ x = h\sqrt{\Delta}$.)

An alternative form of the splitting-up approximation, namely,

$$u_o(x) = p(x), \qquad u_i(x) = \exp((y(t_i) - y(t_{i-1}))h(x) - 0.5|h(x)|^2\Delta) T_\Delta u_{i-1}(\cdot)(x)\ , \tag{4.5}$$

can be obtained by Theorem 2.4 in the same way.

Next, we present an estimate on the number of on-line operations required by S^3 and compare it with a corresponding estimate for the splitting-up method.

We introduce the following parameters: N_s, the number of grid points in the spatial domain; N_J, the number of elements in J_N^n; κ, the number of basis functions e_l.

Assume that one needs to compute an approximation to the solution of (2.4) at moment $t = N_\tau \Delta$.

To do this using S^3, one has to find $\psi_l(i, N, n, \kappa)$, $i = 1, \dots, N_\tau$, for every $l = 1, \dots, \kappa$, which requires about $2\kappa^2 N_J N_\tau$ flops, and then compute the sum in (2.25)—κN_s more flops. The Wiener integral $\xi_{k,l} = \int_0^\Delta m_k(s) dy^l(s)$ reduces to a one-dimensional Riemann integral by integrating by parts. In addition, computations of the integrals $\xi_{k,l}$ for different k and l can be performed in parallel. As a result, computational complexity of the Wick polynomials ξ_α is negligible as compared to other procedures of S^3.

The total number of flops N_{S^3} is then $N_{S^3} = 2\kappa^2 N_J N_\tau + N_s \kappa$. Given the precision of the approximation, the number κ^2 will grow with d as C^d, where C is some constant depending on the type of the basis (but not on d), so $N_{S^3} \leq C^d(2N_J N_\tau + N_s)$.

If the splitting-up algorithm is used, one has to perform N_τ steps of the type (4.2). Each step requires solving a parabolic equation. To estimate the corresponding number of operations, assume that a finite element method is used and the resulting linear system is solved using an iterative procedure without preconditioning. The matrix of the system is of dimension $N_s \times N_s$, sparse and nonsymmetric (since operator $\mathcal{L}^*$ is not self-adjoint). Then one iteration requires about $C_d N_s$ flops, where C_d is a constant depending on d and on the particular numerical algorithm (see [1]), and the total number of iterations is proportional to the condition number of the matrix [36]. For nonsymmetric matrices, the condition number is proportional to at least $(\ln N_s)^{d-1}$ [1, 6]. Thus the total number of operations required to solve the equation on one step is $C_d N_s (\ln N_s)^{d-1}$. One also has to compute a certain number of exponential

TABLE 4.1
Comparison of the splitting-up approximation and the S^3.

	$t = 1$ (Step 100)		$t = 2$ (Step 200)	
	Splitting-up	S^3	Splitting-up	S^3
Flops	8161001	397431	16321601	788431
N_{50}	32	27	20	16
N_{75}	61	55	47	36
N_{95}	93	90	85	81

functions, but this can be done much faster and we disregard it. The total number of on-line operations is then $N_{sp-up} = N_\tau C_d N_s (\ln N_s)^{d-1}$.

As a result,

$$\frac{N_{\mathrm{S}^3}}{N_{sp-up}} \le \frac{C}{C_d}\Big(\frac{C}{\ln N_s}\Big)^{d-1}\Big[\frac{2N_J}{N_s} + \frac{1}{N_\tau}\Big].$$

Theorem 2.4 shows that for any d the splitting-up algorithm and S^3 have errors of the same order in Δ already for $N = 2,\ n = 1$, so we can take $1 + 2r + r(r-1)/2$ as the lower bound on N_J, where r is the dimension of the observation process. Since N_s usually grows with d, we can expect S^3 to have an advantage over the splitting-up algorithm in the following situations:

(1) when the estimation of u is required at one time moment after a long observation period ($N_\tau \gg 1$). This is characteristic for some tracking problems.

(2) when the dimension d of the state process is large.

To conclude this section we compare (numerically) the on-line performance of S^3 and the splitting-up method for one simple example.

For the test model, both signal and observation processes were chosen one-dimensional with the signal

$$dx(t) = 0.1\cos(2x(t))dt + 0.14 d\tilde{w}(t), \quad x(0) \sim \mathcal{N}(0, 0.1),$$

and the observations

$$y(t) = \int_0^t \arctan(x(s))ds + 0.04 w(t);$$

obvious modifications were made to reduce the last equation to the standard form (2.1). We took $T = 2$ and $\Delta = 0.01$.

The interval $[-1, 1]$ was taken as the spatial domain; it was discretized uniformly with step 0.01. Functions $\sin(\pi l(x-1)/2),\ 1 \le l \le 15$, sampled at the points of the spatial grid served as the basis $\{e_l\}$.

For the S^3, multiindices α with $|\alpha| \le 8,\ d(\alpha) \le 1$ were used. (This corresponds to the set J_8^1 in Theorem 2.4.)

Given the trajectory of the signal process, 100 independent observation trajectories were simulated; for each trajectory, the filtering density was computed at moments $25\Delta,\ 50\Delta, \dots, 200\Delta$, using both the S^3 and the splitting-up method.

The results are presented in Table 4.1. They are borrowed from [13]. In the table, "flops" stands for the total number of the *on-line* floating point operations (additions and multiplications) that it took to compute the filtering density at the given time moment; N_{50} (resp., N_{75}, N_{95}) is the number of times the value of the signal process was in the 50% (resp., 75%, 95%) confidence interval defined by the *computed* density.

We see that S^3 results in substantial reduction (up to 20 times) in the number of on-line computations without significant loss of accuracy. The decrease in the number of on-line computations should be even more conspicuous as the dimension of the observation process grows.

Appendix. To make the exposition as self-contained as possible, we will prove Theorem 2.1 and give some other results used in the proof of Theorem 2.2. Most of the results come from [30].

The summation over repeated indices convention is still in force. We also use the notations introduced at the beginning of section 3.

To begin with we recall the celebrated Cameron–Martin development (see, e.g., [7] and also [16, 19]).

THEOREM A.1 (Cameron–Martin development). *Let $B_s = (B_s^1, \dots, B_s^r)$, $0 \le s \le T$, be an r-dimensional Brownian motion and η be a measurable functional of the path $\{B_s, s \le T\}$ such that $\mathbf{E}\eta^2 < \infty$. Let $\{c_i(t)\}_{i\ge 1}$ be an arbitrary complete orthonormal system in $L_2([0,T])$. For $\alpha = \{\alpha_k^l\} \in J$ set*

$$\xi_\alpha(B) = \prod_{k,l} \frac{H_{\alpha_k^l}\Big(\int_0^T c_k(s)dB^l(s)\Big)}{\sqrt{\alpha_k^l!}}.$$

Then $(\xi_\alpha)_{\alpha\in J}$ is a CONS in $L_2(\Omega, \mathcal{F}_T^B, \mathrm{P})$, where $\mathcal{F}_T^B = \sigma(B_s,\ s \le T)$, and

$$\eta = \sum_{\alpha\in J} \mathbf{E}[\eta\xi_\alpha(B)]\xi_\alpha(B), \tag{A.1}$$

$$\mathbf{E}\eta^2 = \sum_{\alpha\in J} (\mathbf{E}[\eta\xi_\alpha(B)])^2. \tag{A.2}$$

The series (A.1) *converges in $L_2(\Omega, \mathrm{P})$.*

Let $\{z_k^l\},\ l = 1, \dots, r,\ k = 1, 2, \dots$, be a sequence of real numbers such that $\sum_{k,l} |z_k^l|^2 < \infty$. Set $m_z^l = m_k(s)z_k^l$, where $\{m_k\}$ is a CONS in $L_2([0,t])$. We also define

$$P_s(z) = \exp\Big\{ \int_o^s m_z^l(\tau)dy^l(\tau) - 0.5 \int_0^s \sum_{l=1}^r |m_z^l(\tau)|^2 d\tau \Big\}$$

and denote

$$\frac{\partial^\alpha}{\partial z^\alpha} := \prod_{k,l} \frac{\partial^{\alpha_k^l}}{(\partial z_k^l)^{\alpha_k^l}}.$$

Proof of Theorem 2.1. It is known (see, e.g., [34]) that for every t, x the UFD $u(t,x)$ is a measurable functional of the observation process $y(s),\ s \le t$. By Girsanov's theorem $y(s)$ is a Brownian motion on the new probability space $(\Omega, \mathcal{F}, \tilde{\mathrm{P}})$ (recall that $d\tilde{\mathrm{P}} = \rho(T)d\mathrm{P}$). Then by Theorem A.1 we have

$$u(t,x) = \sum_{\alpha\in J} \tilde{\mathbf{E}}[u(t,x)\xi_\alpha(y)]\xi_\alpha(y), \tag{A.3}$$

$$\tilde{\mathbf{E}}|u(t,x)|^2 = \sum_{\alpha\in J} (\tilde{\mathbf{E}}[u(t,x)\xi_\alpha(y)])^2, \tag{A.4}$$

where $\tilde{\mathbf{E}}$ stands for the expectation symbol with respect to measure $\tilde{\mathrm{P}}$, and the right-hand side of (A.3) converges in $L_2(\Omega, \tilde{\mathrm{P}})$.

Let us denote

$$\varphi_\alpha(s,x) := \sqrt{\alpha!}\tilde{\mathbf{E}}[u(s,x)\xi_\alpha(y)].$$

It is a standard fact (see, e.g., [16]) that

$$\xi_\alpha(y) = \frac{1}{\sqrt{\alpha!}}\frac{\partial^\alpha}{\partial z^\alpha}P_t(z)|_{z=0},$$

and so for every $s \le t$

$$\varphi_\alpha(s,x) = \frac{\partial^\alpha}{\partial z^\alpha}\tilde{\mathbf{E}}[u(s,x)P_t(z)]|_{z=0} = \frac{\partial^\alpha}{\partial z^\alpha}\tilde{\mathbf{E}}[u(s,x)P_s(z)]|_{z=0},$$

where the second equality follows from the martingale property of $P_s(z)$ on $(\Omega, \tilde{\mathrm{P}})$. Now to prove (2.8) and (2.9) it remains to show that the system of functions $\{\varphi_\alpha\}$, $\alpha \in J$, is a solution to the S-system (2.6). For this purpose it will be convenient to treat the UFD $u(t,x)$ as the solution of the Zakai equation (2.4). Since $P_s(z)$ satisfies the Ito stochastic differential equation

$$dP_s(z) = m_z^l(s)P_s(z)dy^l(s), \quad s \le t;\ P_0(z) = 1, \tag{A.5}$$

by the Ito chain rule

$$\begin{aligned} u(t,x)P_t(z) = p(x) &+ \int_0^t (\mathcal{L}^* u(s,x)P_s(z) + h^l(x)m_z^l(z)u(s,x)P_s(z))ds \\ &+ \int_0^t (h^l(x)u(s,x)P_s(z) + u(s,x)m_z^l(s)P_s(z))dy^l(s). \end{aligned}$$

Taking expectation $\tilde{\mathbf{E}}$ on both sides of the last equality and setting $\varphi(s,x,z) := \tilde{\mathbf{E}}u(s.x)P_s(z)$ we obtain

$$\begin{aligned} &\frac{\partial\varphi(s,x,z)}{\partial s} = \mathcal{L}^*\varphi(s,x,z) + m_z^l(s)h^l(x)\varphi(s,x,z), \quad 0 < s \le t, \\ &\varphi(0,x,z) = p(x)1_{\{|\alpha|=0\}}. \end{aligned} \tag{A.6}$$

Applying the operator $\frac{1}{\sqrt{\alpha!}}\frac{\partial^\alpha}{\partial z^\alpha}$ on both sides of (A.6) and setting $z = 0$ we get (2.6).

To complete the proof of Theorem 2.1 one needs to prove that the right-hand side of (2.8) converges also in $L_1(\Omega, \mathrm{P})$. This follows in a simple way from the convergence in $L_2(\Omega, \tilde{\mathrm{P}})$ and Cauchy–Schwartz inequality [30]. □

In what follows we give some additional properties of the solution of (2.6) in the case $r = 1$; these properties are used in the proof of Theorem 2.2. Generalizations to the general case $r > 1$ are straightforward.

PROPOSITION A.1 (see [30]). *Let $\{\varphi_\alpha(t,x)\}_{\alpha\in J}$ be a solution of* (2.6). *Then for each α with $|\alpha| = k$*

$$\begin{aligned} &\varphi_\alpha(t,x) = \sum_{\sigma\in\mathcal{P}^k}\int^{(k)} F(t;s^k;x)m_{i_{\sigma(k)}}(s_k)\dots m_{i_{\sigma(1)}}(s_1)ds^k, \qquad k > 1, \\ &\varphi_\alpha(t,x) = \int_0^t T_{t-s_1}hT_{s_1}p(x)m_i(s_1)ds_1, \qquad k = 1, \\ &\varphi_\alpha(t,x) = T_t p(x), \qquad k = 0, \end{aligned} \tag{A.7}$$

where $(i_1, \dots, i_k)$ *is the characteristic set of* α *(see the beginning of section* 3 *for notation).*

In addition,

$$\sum_{|\alpha|=k} \frac{\varphi_\alpha^2(t,x)}{\alpha!} = \int^{(k)} |F(t;s^k;x)|^2 ds^k. \tag{A.8}$$

Proof. Representation (A.7) is obviously true for $|\alpha| = 0$. Then the general case $|\alpha| \geq 1$ follows by induction from the variation of parameters formula.

To prove (A.8), first of all note that

$$\sum_{\sigma\in\mathcal{P}^k} m_{i_{\sigma(k)}}(s_k) \dots m_{i_{\sigma(1)}}(s_1) = \sum_{\sigma\in\mathcal{P}^k} m_{i_k}(s_{\sigma(k)}) \dots m_{i_1}(s_{\sigma(1)}).$$

Indeed, any term on the left corresponding to a given $\sigma_0 \in \mathcal{P}^k$ is equal to the term on the right corresponding to $\sigma_0^{-1} \in \mathcal{P}^k$.

Then we can write (A.7) as

$$\varphi_\alpha(t,x) = \int^{(k)} F(t;s^k;x) E_\alpha(s^k) ds^k.$$

Introducing

$$G(s^k;x) := \sum_{\sigma\in\mathcal{P}^k} T_{t-s_{\sigma(k)}} h \dots T_{s_{\sigma(2)}-s_{\sigma(1)}} h T_{s_{\sigma(1)}} p(x) 1_{s_{\sigma(1)}<\dots<s_{\sigma(k)}},$$

we can rewrite it further as

$$\varphi_\alpha(t,x) = \frac{1}{k!} \int\limits_{[0,t]^k} G(s^k) E_\alpha(s^k) ds^k. \tag{A.9}$$

Since for each x G is a symmetric function from $L_2([0,t]^k)$ and $\{E_\alpha/\sqrt{\alpha!k!}, |\alpha| = k\}$ form a CONS for the symmetric part of the space, we have

$$G = \sum_{|\alpha|=k} \frac{c_\alpha E_\alpha}{\sqrt{\alpha!k!}}$$

with some $c_\alpha \in \mathbf{R}$. Then from (A.9) $\varphi_\alpha^2/\alpha! = c_\alpha^2/k!$ and so

$$\sum_{|\alpha|=k} \frac{\varphi_\alpha^2(t,x)}{\alpha!} = \frac{1}{k!} \sum_{|\alpha|=k} c_\alpha^2 = \frac{1}{k!} \int\limits_{[0,t]^k} |G(s^k;x)|^2 ds^k$$

$$= \frac{1}{k!} \int\limits_{[0,t]^k} \Big| \sum_{\sigma\in\mathcal{P}^k} T_{t-s_{\sigma(k)}} h \dots T_{s_{\sigma(2)}-s_{\sigma(1)}} h T_{s_{\sigma(1)}} p(x) 1_{s_{\sigma(1)}<\dots<s_{\sigma(k)}} \Big|^2 ds^k$$

$$= \int^{(k)} |F(t;s^k;x)|^2 ds^k,$$

which proves (A.6). □

REMARK A.1. *In this article we needed WCE* (2.8) *only at the final point of the time interval. However, it is readily checked that, due to* $\mathcal{F}_t^y$*-measurability of UFD*

$u(s,x)$ for all $s \le t$, the statement and the proof of Theorem 2.1 *remain virtually unchanged if, in* (2.8) *and* (2.9)*, we replace t by any $s \le t$. This implies in particular that equality* (2.16) *holds not only for grid points t_i but for every $s \in [t_{i-1}, t_i]$.*

REMARK A.2. *If $r = 1$ and $|\alpha| = k$, then, by* [17, *Theorem* 3.1],

$$\xi_\alpha = \frac{1}{\sqrt{\alpha!}} \int_0^t \int_0^{s_k} \dots \int_0^{s_2} E_\alpha(s^k) dy(s_1) \dots dy(s_k).$$

This gives an alternative (but equivalent) form of WCE (2.8) *in terms of multiple Wiener integrals. A similar expansion holds for an arbitrary r.*

Acknowledgments. Very helpful comments and suggestions from Drs. K. Ito and J. Dym and from the reviewers are acknowledged.

REFERENCES

[1] O. AXELSSON AND V. A. BARKER, *Finite Element Solution of Boundary Value Problems*, Academic Press, New York, 1984.
[2] J. BARAS, *Real-time architectures for the Zakai equations and applications*, in Stochastic Analysis, E. Mayer-Wolf et al., eds., Academic Press, Boston, 1991.
[3] V. E. BENESH, *Exact finite-dimensional filters for certain diffusions with nonlinear drift*, Stochastics, 5 (1981), pp. 65–92.
[4] A. BENSOUSSAN, R. GLOWINSKI, AND R. RASCANU, *Approximations of the Zakai equation by splitting up method*, SIAM J. Control Optim., 28 (1990), pp. 1420–1431.
[5] A. BUDHIRAJA AND G. KALLIANPUR, *Approximations to the Solution of the Zakai Equations Using Multiple Wiener and Stratonovich Integral Expansions*, Tech. report 447, Center for Stochastic Processes, University of North Carolina, Chapel Hill, NC, Jan. 1995.
[6] Z. CAI AND WEINAN E, *Hierarchical method for elliptic problems using wavelet*, Comm. Appl. Numer. Methods, 8 (1992), pp. 819–825.
[7] R. H. CAMERON AND W. T. MARTIN, *The orthogonal development of non-linear functionals in a series of Fourier-Hermite functions*, Ann. Math., 48 (1947), pp. 385–392.
[8] J. M. C. CLARK, *The design of robust approximations to the stochastic differential equations of nonlinear filtering*, in Communication Systems and Random Processes Theory, NATO Adv. Sci. Inst. Ser. E Appl. Sci. 25, J. K. Skwirzynski, ed., Springer-Verlag, Berlin, 1977, pp. 721–734.
[9] G. B. DIMASI AND W. J. RUNGGALDIER, *On measure transformations for combined filtering and parameter estimation in discrete time*, Systems Control Lett., 2 (1982), pp. 57–62.
[10] R. J. ELLIOTT AND R. GLOWINSKI, *Approximations to solutions of the Zakai filtering equation*, Stochastic Anal. Appl., 7 (1989), pp. 145–168.
[11] W. H. FLEMING AND E. PARDOUX, *Optimal control for partially observed diffusions*, SIAM J. Control Optim., 20 (1982), pp. 261–285.
[12] P. FLORCHINGER AND F. LEGLAND, *Time discretization of the Zakai equation for diffusion processes observed in correlated noise*, Stochastics Stochastics Rep., 35 (1991), pp. 233–256.
[13] C. P. FUNG, *Solving Hidden Markov Problems by Spectral Methods*, Ph.D. thesis, University of Southern California, Los Angeles, CA, Nov. 1995.
[14] J. GLIMM AND A. JAFFE, *Quantum Physics*, Springer, New York, 1987.
[15] D. GOTTLIEB AND S. A. ORSZAG, *Numerical Analysis of Spectral Methods: Theory and Applications*, CBMS-NSF Regional Conf., Ser. in Appl. Math., 26, SIAM, Philadelphia, 1977.
[16] T. HIDA, H.-H. KUO, J. POTTHOFF, AND L. SREIT, *White Noise*, Kluwer, Boston, 1993.
[17] K. ITO, *Multiple Wiener integral*, J. Math. Soc. Japan, 3 (1951), pp. 157–169.
[18] K. ITO, *Approximation of the Zakai equation for nonlinear filtering*, SIAM J. Control Optim., 34 (1996), pp. 620–634.
[19] G. KALLIANPUR, *Stochastic Filtering Theory*, Springer-Verlag, New York, 1980.
[20] N. V. KRYLOV AND B. L. ROZOVSKII, *On conditional distributions of diffusion processes*, Math. USSR-Izv., 42 (1978), pp. 336–356.
[21] N. V. KRYLOV AND A. VERETENNIKOV, *On explicit formulae for solutions of stochastic equations*, Math. USSR-Sb., 29 (1976), pp. 239–256.
[22] H. KUNITA, *Cauchy problem for stochastic partial differential equations arising in non-linear filtering theory*, Systems Control Lett., 1 (1981), pp. 37–41.

[23] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, Cambridge, UK, 1982.
[24] H. J. Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, 1977.
[25] O. A. Ladyzhenskaia, V. A. Solonikov, and N. N. Ural'tseva, *Linear and Quasi-linear Equations of Parabolic Type*, American Mathematical Society, Providence, RI, 1968.
[26] F. LeGland, *Time discretizations of nonlinear filtering equations*, in Proc. 28th IEEE Conf. on Decision and Control, Tampa, 1989, IEEE Control Systems Society, Piscataway, NJ, 1989, pp. 2601–2606.
[27] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes*, Springer-Verlag, New York, 1992.
[28] J. T.-H. Lo and S.-K. Ng, *Optimal orthogonal expansion for estimation* I*: Signal in white Gaussian noise*, in Nonlinear Stochastic Problems, R. Bucy and J. Moura, eds., D. Reidel, Dordrecht, 1983, pp. 291–309.
[29] R. Mikulevicius and B. L. Rozovskii, *Linear parabolic stochastic PDEs and Wiener chaos*, unpublished.
[30] R. Mikulevicius and B. L. Rozovskii, *Separation of observations and parameters in nonlinear filtering*, in Proc. 32nd IEEE Conf. on Decision and Control, Part 2, San Antonio, 1993, IEEE Control Systems Society, Piscataway, NJ, 1993, pp. 1564–1569.
[31] R. Mikulevicius and B. L. Rozovskii, *Soft solutions of linear parabolic SPDE's and the Wiener chaos expansion*, in Stochastic Analysis on Infinite Dimensional Spaces, Pitman Res. Notes Math. Sec. 310, H. Kunita and H.-H. Kuo, eds., Longman, Harlow, UK, 1994.
[32] D. Ocone, *Multiple integral expansions for nonlinear filtering*, Stochastics, 10 (1983), pp. 1–30.
[33] E. Pardoux, *Filtrage non lineaire et equations aux derivees partielles stochastiques associees*, in Ecole d'été de Probabilités de Saint-Flour, Springer-Verlag, New York, 1989.
[34] B. L. Rozovskii, *Stochastic Evolution Systems*, Kluwer, Amsterdam, 1990.
[35] E. Wong, *Explicit solutions to a class of nonlinear filtering problems*, Stochastics, 5 (1981), pp. 311–321.
[36] D. M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.
[37] M. Zakai, *On the optimal filtering of diffusion processes*, Z. Wahrscheinlichkeitstheorie und verw. Gebiete, 4 (1969), pp. 230–233.
[38] A. K. Zvonkin and N. V. Krylov, *Strong solutions of stochastic differential equations*, in Proc. of the School Seminar on the Theory of Random Processes, Druskinikai, November 25–30, 1974, Vilnius, 1975 (in Russian).