On Galerkin Approximations for the Zakai Equation with Diffusive and Point Process Observations

Rüdiger Frey[1], Thorsten Schmidt[2] and Ling Xu[3]

June 14, 2011

Abstract

This paper studies Galerkin approximations applied to the Zakai equation of stochastic filtering. The basic idea of this approach is to project the infinite-dimensional Zakai equation onto some finite-dimensional subspace; this leads to a finite-dimensional system of stochastic differential equations that can be solved numerically. The contribution of the paper is twofold. On the theoretical side, existing convergence results are extended to filtering models with observations of point-process or mixed type. On the applied side, various issues related to the numerical implementation of the method are considered. The paper closes with a numerical case study.

Keywords Stochastic filtering, Zakai equation, point processes, Galerkin approximation

AMS classification 60G35, 60H15, 65C30, 92E11

1 Introduction

Stochastic filtering deals with the recursive estimation of the conditional distribution of signal process X given some form of noisy observation of X. In the standard continuous time filtering models this noisy observation is generated by a process Z with dynamics of the form

$$Z_t = Z_0 + \int_0^t h(X_s)ds + W_t \tag{1}$$

for some Brownian motion W that is independent of X. In that case $\pi_t(dx)$, the conditional distribution of X_t given $\sigma(Z_s\colon s \leq t)$, can be characterized by a measure-valued stochastic partial differential equation (SPDE) known as *Zakai equation*. This SPDE is in general an infinite-dimensional equation that cannot be solved directly. In view of the practical relevance of filtering, a wide range of methods for the approximation of this equation by finite-dimensional systems and for the numerical solution of filtering problems in general has therefore been developed; a good survey is given in Budhiraja, Chen, and Lee (2007) or in Bain and Crisan (2009). Popular numerical methods for filtering problems include the extended Kalman filter (Jazwinski (1970)); quantization (Gobet, Pagès, Pham, and Printems (2006)); Markov-chain approximation (Dupuis and Kushner (2001), Di Masi and Runggaldier (1982)); spectral methods (Lototsky (2006)) and simulation methods such as particle filtering (Crisan, Moral, and Lyons (1999)).

[1]Department of mathematics, Universität Leipzig, mail: ruediger.frey@math.uni-leipzig.de

[2]Department of mathematics, Chemnitz University of Technology, mail: thorsten.schmidt@mathematik.tu-chemnitz.de

[3]School of Economics and Administration, China University of Petroleum, mail: Ling.Xu@math.uni-leipzig.de.
Part of this work was written while the third author was doing her Ph.D. at Leipzig University. Financial support from the International Max Planck Research School 'Mathematics in the Sciences', Leipzig and from the German Science foundation (DFG) is gratefully acknowledged.

If the signal X is a diffusion process with uniformly parabolic generator the conditional distribution $\pi_t(dx)$ admits a Lebesgue density that solves a SPDE in a suitable function space, the so-called Zakai equation for the unnormalized conditional density; see for instance Pardoux (1979b). Galerkin approximations are an important numerical technique for dealing with this SPDE. The basic idea of this approach is to project the Zakai equation for the conditional density onto some finite-dimensional subspace H_n generated by basis functions $e_1, \ldots, e_n$. This leads to an n-dimensional SDE system for the Fourier coefficients of the solution of the projected equation; this SDE system can then be solved by numerical methods for "ordinary" SDEs.

Theoretical and numerical aspects of Galerkin approximations are well understood for the case of pure diffusion observation as in (1); see for instance Germani and Piccioni (1984) and Germani and Piccioni (1987) for convergence results for Galerkin approximations and Ahmed and Radaideh (1997) for a case study and a discussion of numerical aspects. Much less is known for the case of mixed observations of diffusion *and* point-process type. In this paper we therefore consider a model where a doubly stochastic point process N with intensity $\lambda(X_t)$ is observable in addition to the process Z. Models of this type arise naturally in credit risk modelling (see Example 2.1 below) or in the modelling of high frequency data in finance (Frey and Runggaldier (2001), Cvitanic, Liptser, and Rozovski (2006)). Outside the field of financial mathematics point-process information plays among others a crucial role in the analysis of queueing systems (Brémaud (1981)).

Our contribution is twofold. On the theoretical side we generalize the convergence results of Germani and Piccioni (1987) to the case of mixed observations. On the applied side we extend the numerical analysis of Ahmed and Radaideh (1997) in various ways: to begin with, we propose to use Hermite polynomials as basis functions (instead of Gaussian basis functions); we explain how to change the basis adaptively in order to deal with sudden shifts in location and scale of the conditional density caused for instance by jumps in the observation, and we compare several methods for the numerical implementation of the SDE-system that results from the Galerkin approximation. An extensive simulation study shows that the Galerkin approximation works well for systems with mixed observation provided that the necessary care is taken in the implementation of the method.

The paper is organized as follows. The model and the various versions of the Zakai equation are described in Section 2. In that section we moreover introduce the basic form of the Galerkin approximation. Convergence results for the Galerkin approximation are given in Section 3. Section 4 deals with the numerical implementation of the model; results from numerical experiments are finally reported in Section 5.

2 Zakai equation and Galerkin approximation

In this section we introduce the nonlinear filtering problem studied in this paper. Moreover, we present different versions of the Zakai equation that describe the solution of the filtering problem. Finally we introduce the Galerkin approximation for the Zakai equation for the unnormalized conditional density and we derive an SDE system for the Fourier coefficients.

2.1 Model and notation

We consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ where the filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}$ satisfies the usual conditions and where T is an arbitrary but fixed horizon date. The nonlinear filtering problem we study consists of an unobserved state process X and observations Z and N. Z is a nonlinear continuous transformation of X with additional Gaussian noise, while N is a doubly stochastic Poisson process whose intensity is a nonlinear function of X.

The state process. We consider an unobserved *state process* X on $\mathbb{R}^d$ which is the solution of the SDE

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dV_s, \quad 0 \le t \le T, \tag{2}$$

for a m-dimensional $\mathbb{F}$-Brownian motion V. Moreover, we assume that X_0 has finite second moments and a density $p_0 \in L^2(\mathbb{R}^d)$. Set $a(x) = \sigma(x)\sigma(x)^\top$. The components of $a(x)$ and $b(x)$ are denoted by $a_{ij}(x)$ and $b_i(x)$, respectively. The restriction of the generator $\mathscr{L}$ of the Markov process X to $C_b^2(\mathbb{R}^d)$, the set of all bounded and twice continuously differentiable functions on $\mathbb{R}^d$, is given by the second order differential operator

$$\mathscr{L} = \sum_{i=1}^d b_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}. \tag{3}$$

Note that the Itô-formula implies that for $f \in C_b^2(\mathbb{R}^d)$, $M_t^f := f(X_t) - f(X_0) - \int_0^t \mathscr{L}f(X_s)ds$ is an $\mathbb{F}$-martingale.

The observation processes. The observation is given by the two processes Z and N. The process Z satisfies

$$Z_t = \int_0^t h(X_s)ds + W_t, \quad 0 \le t < \infty\,, \tag{4}$$

where $h : \mathbb{R}^d \to \mathbb{R}^l$ is a measurable function and W is an l-dimensional standard Brownian motion, independent of X. Moreover, the process N is a doubly stochastic Poisson process with intensity $\lambda(X_t)$ where λ is a positive, continuous and bounded function, so that the process $N_t - \int_0^t \lambda(X_s)ds$ is an $\mathbb{F}$-martingale. We denote the jump times of N by $\tau_1, \tau_2, \ldots$.

The objective of nonlinear filtering is to find suitable ways for computing $\pi_t(dx)$, the conditional distribution of the state X_t given the observation history in a recursive way. More formally, let $\mathcal{F}_t^{Z,N} := \sigma(Z_u, N_u : 0 \le u \le t)$, so that the associated filtration $\mathbb{F}^{Z,N}$ represents the information given by the observation. The conditional distribution of X_t given the observation until time t is determined by

$$\pi_t(f) := \mathbb{E}\big(f(X_t)|\mathcal{F}_t^{Z,N}\big), \quad f \in L^\infty(\mathbb{R}^d).$$

The following regularity assumptions on the data of the problem will be used throughout the paper

(A1) Assume that the following three conditions hold:

(i) $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, and $h : \mathbb{R}^d \to \mathbb{R}^l$ are bounded on $\mathbb{R}^d$. Moreover, b is C^1 with bounded derivatives and σ is C^2 with bounded first and second order derivatives.

(ii) There exists $\alpha > 0$, such that $z^\top a(x)z \ge \alpha z^\top z$, $\forall x, z \in \mathbb{R}^d$.

(iii) $\lambda : \mathbb{R}^d \to [\varpi_1, \varpi_2]$ is a continuous function for constants $0 < \varpi_1 < \varpi_2$.

Example 2.1. Filtering problems with diffusive and point process observations arise naturally in credit risk modeling. This connection was studied systematically in Frey and Runggaldier (2010) and Frey and Schmidt (2010), among others. In these papers reduced-form portfolio credit risk models are considered where default times are doubly stochastic random times with intensity driven by some economic factor process X. In a large homogeneous portfolio the number of default events is thus given by some doubly stochastic Poisson process N with intensity $\lambda(X_t)$. In line with reality, it is assumed that investors cannot observe the process

X directly, but are confined to noisy observations of X, modelled by a process Z as in (4). Moreover, they obviously observe the occurrence of default events and hence the process N.

In this context the pricing of credit derivatives naturally leads to a filtering problem, as we now explain. In abstract terms a credit derivative with maturity T can be described in terms of some $\mathcal{F}_T^N$-measurable payoff H. Denote by $\mathbb{Q}$ the risk neutral measure used for pricing. The price of the credit derivative at time $t \leq T$ is then given by $H_t = \mathbb{E}^{\mathbb{Q}}(H \mid \mathcal{F}_t^{Z,N})$ (assuming zero interest rates for simplicity). Using iterated conditional expectations we get

$$H_t = \mathbb{E}^{\mathbb{Q}}\Big(\mathbb{E}^{\mathbb{Q}}(H \mid \mathcal{F}_t) \mid \mathcal{F}_t^{Z,N}\Big).$$

It is well-known that the pair (X, N) is an $\mathbb{F}$-Markov process. Hence for typical claims H one has the equality $\mathbb{E}^{\mathbb{Q}}(H|\mathcal{F}_t) = h(t, X_t, N_t)$ for a suitable function h, and we get that $H_t = \mathbb{E}^{\mathbb{Q}}\big(h(t, X_t, N_t)|\mathcal{F}_t^{Z,N}\big)$. The computation of this conditional expectation is a nonlinear filtering problem of the type considered in the present paper.

For further information on incomplete-information models in credit risk we refer to the to the survey article Frey and Schmidt (2011).

2.2 The measure-valued Zakai equation

The evolution equation for the measure $\pi_t(dx)$ is usually deduced using a change of measure method. Define

$$\Lambda_t := \prod_{\tau_n \leq t} \lambda(X_{\tau_n -}) \cdot \exp\left(\int_0^t h(X_s)^\top dW_s + \frac{1}{2}\int_0^t \|h(X_s)\|^2 ds - \int_0^t (\lambda(X_s) - 1) ds \right)$$

for $t \in [0, T]$. Then the regularity assumptions in **(A1)** imply that $(\Lambda_t^{-1})_{t\in[0,T]}$ is a nonnegative martingale. We define the measure $\mathbb{P}^0$ by its Radon-Nikodym derivative $d\mathbb{P}^0 = \Lambda_T^{-1} d\mathbb{P}$. The Girsanov theorem yields that, under $\mathbb{P}^0$, Z is a standard Brownian motion, that N is a Poisson process with intensity equal to one, and that X, Z and N are independent. Denote by $Y_t := N_t - t$ the compensated Poisson process, such that under $\mathbb{P}^0$, Y is a martingale. Then the conditional distribution $\pi_t(dx)$ has a representation in terms of an associated unnormalized version ρ: denoting by $\mathbb{E}^0$ the expectation w.r.t. $\mathbb{P}^0$, we obtain by the abstract Bayes rule for any $f \in L^\infty(\mathbb{R}^d)$

$$\pi_t(f) = \frac{\mathbb{E}^0(f(X_t)\Lambda_t|\mathcal{F}_t^{Z,N})}{\mathbb{E}^0(\Lambda_t|\mathcal{F}_t^{Z,N})} =: \frac{\rho_t(f)}{\rho_t(1)}. \tag{5}$$

It is well-known that the measure-valued process ρ_t satisfies the *classical Zakai equation*: let $\rho_0(f) := \mathbb{E}[f(X_0)|\mathcal{F}_0^{Z,N}]$. Then, for any $f \in C_b^2(\mathbb{R}^d)$, $t \in [0, T]$,

$$\rho_t(f) = \rho_0(f) + \int_0^t \rho_s(\mathscr{L} f) ds + \int_0^t \rho_s(f h^\top) dZ_s + \int_0^t \rho_{s-}\Big(f(\lambda - 1)\Big) dY_s, \tag{6}$$

$\mathbb{P}^0 - a.s.$, see for instance Theorem 3.24 in Bain and Crisan (2009) (only continuous observations). A formal proof that under **(A1)**, (6) holds in the setup of the present paper is given in Xu (2010), Theorem 2.9.

2.3 The Zakai equation for the conditional density

Our aim is to determine the dynamics of the Lebesgue-density of the unnormalized conditional distribution $\rho_t(dx)$. Consider the separable Hilbert space $H = L^2(\mathbb{R}^d)$ with norm $\|\cdot\|_H$ and scalar product $(\cdot, \cdot)$. To obtain intuition, suppose that

$$\rho_t(f) = (q_t, f)$$

for all $f \in C^2(\mathbb{R}^d)$ with compact support and for some H-valued process $q = (q_t)_{0\le t\le T}$ such that $q_t(\cdot)$ can be identified with a smooth function. Denote by the differential operator $\mathscr{L}^*$ the formal adjoint of the generator $\mathscr{L}$. As $(q_t, \mathscr{L}f) = (\mathscr{L}^* q_t, f)$ the measure valued equation (6) simplifies to

$$(q_t, f) = (q_0, f) + \int_0^t (\mathscr{L}^* q_s, f) ds + \int_0^t (h^\top q_s, f) dZ_s + \int_0^t ((\lambda - 1) q_{s-}, f) dY_s. \tag{7}$$

This suggest that q solves the stochastic partial differential equation (SPDE)

$$dq_t = \mathscr{L}^* q_t dt + h^\top q_t dZ_t + (\lambda - 1) q_{t-} dY_t$$

in an appropriate sense. The next step is to give this equation a precise mathematical meaning using the theory for mild and weak solutions for SPDEs as in Peszat and Zabczyk (2007). Besides the Hilbert space $H = L^2(\mathbb{R}^d)$ we consider the Sobolev space $V = H^1(\mathbb{R}^d) \subset H$. We define an extension $\mathcal{A}^*$ of $\mathscr{L}^*$ with domain $D(\mathcal{A}^*) \subset V$ as follows: $u \in V$ is an element of $D(\mathcal{A}^*)$ if there exists $f \in H$ such that for all $v \in V$

$$-\frac{1}{2} \sum_{i,j=1}^d \int_{\mathbb{R}^d} a_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} \, dx + \sum_{i=1}^d \int_{\mathbb{R}^d} \Big(b_i - \frac{1}{2} \sum_{j=1}^d \frac{\partial a_{ij}(x)}{\partial x_j} \Big) \frac{\partial v}{\partial x_i} u \, dx = (f, v),$$

and we set $\mathcal{A}^* u = f$ in that case. If $u \in C_0^2(\mathbb{R}^d)$, we obtain that $f = \mathscr{L}^* u$ by checking that $(f, v) = (u, \mathscr{L} v)$ with integration by parts. It is well-known that $\mathcal{A}^*$ generates an C_0-semigroup G^*, i.e. G^* is a map from $[0, T]$ into H^* such that $G^*(0) = 0$, $G^*(t + s) = G^*(t) G^*(s)$ and G^* is continuous in the strong operator topology, see Prato and Zabczyk (1992, Proposition A.10). Moreover, G^* is analytic.

Mild and weak solutions. Let $\mathcal{N}^2(0, T; H)$ denote the set of all $\mathbb{F}^{Z,N}$-adapted, H-valued processes $\xi = (\xi_t)_{0\le t\le T}$, continuous in the mean square norm, which are such that

$$|\xi|_T := \Big(\sup_{t\in[0,T]} \mathbb{E}^0 \Big(\|\xi(t)\|_H^2 \Big) \Big)^{1/2} < \infty. \tag{8}$$

It is well-known that $\mathcal{N}^2(0, T; H)$ is a Banach space with norm $|\cdot|_T$, see Germani and Piccioni (1987).

Define the multiplication-operators $\mathcal{B}\colon H \to H^l$, $\mathcal{B}f = f h^\top$ and $\mathcal{C}\colon H \to H$, $\mathcal{C}f = (\lambda - 1) f$. A *mild solution* of the SPDE

$$dq_t = \mathcal{A}^* q_t dt + \mathcal{B} q_t dZ_t + \mathcal{C} q_t dY_t. \tag{9}$$

is a process $q \in \mathcal{N}^2(0, T; H)$ such that

$$q_t = G_t^* q_0 + \int_0^t G_{t-s}^* \mathcal{B} q_s dZ_s + \int_0^t G_{t-s}^* \mathcal{C} q_{s-}, dY_s \,, \quad t \le T. \tag{10}$$

Denote by $\mathcal{A} := (\mathcal{A}^*)^*$ the adjoint operator of $\mathcal{A}^*$ and note that on $C_0^2(\mathbb{R}^d)$ the operator $\mathcal{A}$ coincides with the generator $\mathscr{L}$ of X. A *weak solution* of the SPDE (9) is a process $q \in \mathcal{N}^2(0, T; H)$ such that for all $v \in D(\mathcal{A})$

$$(q_t, v) = (q_0, v) + \int_0^t (q_s, \mathcal{A} v) \, ds + \int_0^t (q_s, \mathcal{B} v) dZ_s + \int_0^t (q_{s-}, \mathcal{C} v) dY_s, \quad t \le T \,. \tag{11}$$

In our context q is a weak solution of (9) if and only if it is a mild solution of that equation; this follows immediately from Theorem 9.15 in Peszat and Zabczyk (2007).

The Zakai equation. The following result describes the evolution of the density of the unnormalized conditional distribution $\rho_t(dx)$.

Theorem 2.2. *Assume that* **(A1)** *holds. Then for all $q_0 \in V$ there is a unique mild solution q of the SPDE* (9)*. Moreover, $q_t \in H^1(\mathbb{R}^d)$ and for all $f \in L^2(\mathbb{R}^d)$ we have that*

$$\rho_t(f) = (q_t, f).$$

In view of this result, equation (9) will be called the Zakai equation for the unnormalized conditional density.

Theorem 2.2 has been obtained in Pardoux (1979b) and in Germani and Piccioni (1987) for the case of pure diffusion information and in Pardoux (1979a) for the pure Poisson case ($h \equiv 0$). The extension to the case of mixed observations may be found in Xu (2010).

2.4 The Galerkin approximation

The Galerkin approximation for a (stochastic) PDE essentially projects the equation to a finite-dimensional subspace. In the case of the Zakai equation for the unnormalized conditional density the solution of the projected equation can be characterized in terms of a finite-dimensional system of ordinary stochastic differential equations (SDEs), as we now explain.

Formally the Galerkin approximation is defined as follows: Let $\{e_1, e_2, \ldots\} \subset D(\mathcal{A}^*) \cap D(\mathcal{A})$ be a basis of the Hilbert-space H. Let H_n be the linear subspace spanned by $\{e_1, \ldots, e_n\}$ and denote by P_n the projection from H to H_n. We define the projection of the operator $\mathcal{A}^*$ by

$$(\mathcal{A}^*)^{(n)} := P_n \mathcal{A}^* P_n$$

and analogously for the operators $\mathcal{B}$ and $\mathcal{C}$.

Definition 2.3. The n-dimensional *Galerkin approximation* of (9) is the solution of

$$\begin{aligned} dq_t^{(n)} &= (\mathcal{A}^*)^{(n)} q_t^{(n)} dt + \mathcal{B}^{(n)} q_t^{(n)} dZ_t + \mathcal{C}^{(n)} q_{t-}^{(n)} dY_t, \\ q_0^{(n)} &= P_n q_0. \end{aligned} \tag{12}$$

As previously, there are two equivalent concepts of solutions. The mild solution of (12) is obtained with $(G^*)^{(n)} := \exp(\mathcal{A}^*)^{(n)}$. On the other side, the weak form is obtained using the adjoint operator $\mathcal{A}^{(n)} := \big((\mathcal{A}^*)^{(n)}\big)^*$. Since for $u, v \in H$ one has $(P_n \mathcal{A}^* P_n u, v) = (u, P_n \mathcal{A} P_n v)$ the weak form of the Galerkin approximation (12) becomes

$$d(q_t^{(n)}, v) = (q_t^{(n)}, P_n \mathcal{A} P_n v) dt + (q_t, P_n \mathcal{B} P_n v) dZ_t + (q_{t-}, P_n \mathcal{C} P_n v) dY_t\,, \quad v \in H. \tag{13}$$

Note that for $v \in H_n^{\perp}$ we obtain $d(q_t^{(n)}, v) = 0$. Since moreover $q_0^{(n)} = P_n q_0 \in H_n$ it follows that $q_t^{(n)} \in H_n$ for $t \in [0, T]$ $\mathbb{P}$-a.s. Hence, $q_t^{(n)}$ can be written as

$$q_t^{(n)}(x) = \sum_{i=1}^{n} \psi_i^{(n)}(t) e_i(x), \qquad t \in [0, T], \tag{14}$$

where $\psi_i^{(n)}$, $1 \le i \le n$ are the *Fourier coefficients*. Plugging (14) into the weak form of the Galerkin approximation (13), we get that the Fourier coefficients satisfy the following system of ordinary SDEs:

$$\begin{aligned} \sum_{i=1}^{n} (e_i, e_j) d\psi_i^{(n)}(t) = \Big(\sum_{i=1}^{n} \psi_i^{(n)}(t)(e_i, \mathcal{A} e_j) \Big) dt + \sum_{\ell=1}^{l} \Big(\sum_{i=1}^{n} \psi_i^{(n)}(t)(e_i, h^\ell e_j) \Big) dZ_t^\ell \\ + \Big(\sum_{i=1}^{n} \psi_i^{(n)}(t-) \Big(e_i, (\lambda - 1) e_j \Big) \Big) dY_t. \end{aligned}$$

Define the $n \times n$ matrices A, C, D and $B^\ell, \ell = 1, \dots, l$ by their components:

$$a_{ji} := (e_i, \mathcal{A}e_j),\ b^\ell_{ji} := (e_i, h^\ell e_j),\ c_{ji} := (e_i, (\lambda - 1)e_j), d_{ji} := (e_i, e_j), \tag{15}$$

and let $q_0^{(n)} = P_n q_0$. As $\{e_1, e_2, \dots\}$ is a basis of H, the matrix D has full rank and is invertible. Using matrix notation we obtain the following SDE system for the vector-valued process $\Upsilon^{(n)} := (\psi_1^{(n)}, \dots \psi_n^{(n)})^\top$

$$\begin{aligned} d\Upsilon_t^{(n)} &= D^{-1}\Big(A\Upsilon_t^{(n)}dt + \sum_{\ell=1}^{l} B^\ell \Upsilon_t^{(n)} dZ_t^\ell + C\Upsilon_{t-}^{(n)} dY_t\Big), \\ \Upsilon_0^{(n)} &= D^{-1} q_0^{(n)}\,. \end{aligned} \tag{16}$$

This SDE system will be the starting point for our numerical analysis in Section 4. Note that for $\{e_1, e_2, \dots\}$ smooth, one has $a_{ji} = (e_i, \mathscr{L}e_j)$ which is more convenient for computing the coefficients of the system (16).

Moments of the conditional distribution. Obviously, the (normalized) conditional density of $\pi_t(dx)$ can be approximated via

$$p_t := \frac{q_t}{\int_{\mathbb{R}^d} q_t(x)dx} \approx \frac{q_t^{(n)}}{\int_{\mathbb{R}^d} q_t^{(n)}(x)dx} =: p_t^{(n)}\,; \tag{17}$$

here $\approx$ means that we approximate the term on the left side by the Galerkin approximation on the right side. In this case we have that $\mathbb{E}(f(X_t)|\mathcal{F}_t^{Z,N}) \approx (p_t^n, f)$. On the other side, we can represent some characteristics of the conditional distribution directly via q_t. Consider for simplicity the case $d = 1$. Denote by $\hat{x}_t$ and $\hat{\sigma}_t^2$ be the conditional mean and variance of the state process at time $t \in [0, T]$. Then

$$\hat{x}_t = \mathbb{E}(X_t|\mathcal{F}_t^{Z,N}) = \frac{\int x q_t(x)dx}{\int q_t(x)dx} \approx \frac{\int x q_t^{(n)}(x)dx}{\int q_t^{(n)}(x)dx} = \frac{\sum_{i=1}^n \psi_i^{(n)}(t)(x, e_i)}{\sum_{i=1}^n \psi_i^{(n)}(t)(1, e_i)}. \tag{18}$$

Note that the second equality follows from the definition of the unnormalized distribution, see (5). For the last equality we used (14). In a similar way we approximate in $\hat{\sigma}_t^2 = \mathbb{E}\big((X_t - \hat{x}_t)^2|\mathcal{F}_t^{Z,N}\big) = \mathbb{E}(X_t^2|\mathcal{F}_t^{Z,N}) - (\hat{x}_t)^2$ the conditional second moment by

$$\mathbb{E}(X_t^2|\mathcal{F}_t^{Z,N}) \approx \frac{\sum_{i=1}^n \psi_i^{(n)}(t)(x^2, e_i)}{\sum_{i=1}^n \psi_i^{(n)}(t)(1, e_i)}. \tag{19}$$

Analogously all moments of the conditional distribution can be represented by the Fourier coefficients. Notice that $(1, e_i)$, (x, e_i) and (x^2, e_i) are independent of the observation and can be computed off-line (we implicitly assume that these integrals exist for the chosen basis functions).

3 Convergence results

This section gives sufficient conditions for the convergence of the Galerkin approximation $q^{(n)}$ defined in (14) to the solution of the Zakai equation q from (9) in an appropriate sense. The following theorem is the main theoretical result of the paper:

Theorem 3.1. *Assume that* **(A1)** *holds. Let q be the solution of the Zakai equation in* (9) *and $q^{(n)}$ be the corresponding Galerkin approximation. Then, for any $q_0 \in H$,*

$$\sup_{t\in[0,T]} \mathbb{E}^0(\|q_t^{(n)} - q_t\|_H^2) \to 0, \quad as \quad n \to \infty,$$

if and only if, for any $x \in H$,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \left\| \left(\exp(P_n \mathcal{A}^* P_n t) - G_t^* \right) x \right\|_H = 0. \tag{20}$$

Note that $G_t^* x$ is the solution of the Kolmogorov forward PDE with initial condition x (the PDE describing the evolution of the transition density of X) and $\exp(P_n \mathcal{A}^* P_n t)x$ is the Galerkin approximation to this (deterministic) PDE. Hence Theorem 3.1 shows that the Galerkin approximation for the Zakai equation converges if and only if the Galerkin approximation for the deterministic forward equation converges.

Necessary and sufficient conditions for (20) to hold can be obtained by means of the Trotter-Kato theorem. A convenient condition that ensures (20) under **(A1)** is that

$$\bigcup_{n\in\mathbb{N}} H_n \text{ is dense in } V; \tag{21}$$

see Theorem 4, Germani and Piccioni (1984).

Proof of Theorem 3.1

The remainder of this section is devoted to the proof of Theorem 3.1. The essential part of the proof is a continuity result for the mild form of the Zakai equation, see Proposition 3.4 below. This result is an extension of a result from Germani and Piccioni (1987) where the case of continuous observation is treated. We recall the mild form of the Zakai equation in the Banach space $\mathcal{N}^2(0,T;H)$, $q_t = G_t^* q_0 + \int_0^t G_{t-s}^* \mathcal{B} q_s dZ_s + \int_0^t G_{t-s}^* \mathcal{C} q_{s-}, dY_s$, $t \le T$ where for $f \in H$, $\mathcal{B}f = h^\top f$ and $\mathcal{C}f = (\lambda - 1)f$.

We start by introducing some necessary operator spaces. By $\mathcal{S}$ we denote the space of all C_0-semigroups of linear bounded operators from H to H such that there exists $\bar{S} \in \mathbb{R}^+$ with for all $S \in \mathcal{S}$

$$\sup_{t\in[0,T]} \|S_t\| \le \bar{S}. \tag{22}$$

We endow $\mathcal{S}$ with the topology of uniform strong convergence on $[0,T]$, i.e. a sequence $(S^{(n)})$ in $\mathcal{S}$ converges to $S \in \mathcal{S}$ if for all $x \in H$

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \left\| (S_t^{(n)} - S_t)x \right\|_H = 0.$$

For any $l \in \mathbb{N}$ denote by $\mathcal{U}^l$ the space of linear bounded operators from H to H^l (l-fold product of H). In the special case $l = 1$ we write $\mathcal{U} = \mathcal{U}^1$. An operator $A \in \mathcal{U}^l$ can be written by its component-wise: for all $x \in H$,

$$Ax = (A^1 x, \ldots, A^l x)^\top$$

with $A^i \in \mathcal{U}$. The space $\mathcal{U}^l$ is endowed with the strong topology, that is a sequence $(A^{(n)})$ in $\mathcal{U}^l$ converges to $A \in \mathcal{U}^l$, if for all $x \in H$

$$\lim_{n\to\infty} \left\| (A^{(n)} - A)x \right\|_{H^l} = 0.$$

The studied SPDEs. For the proof we study a more general class of linear stochastic partial differential equations that includes the Zakai equation (10) as a special case. Consider a generic semigroup $S \in \mathcal{S}$ and linear operators $B \in \mathcal{U}^l$, $C \in \mathcal{U}$ and some $f \in H$. In the sequel we study the following equation in $\mathcal{N}^2(0,T;H)$:

$$\xi_t = S_t f + \int_0^t S_{t-s} B \xi_s dZ_s + \int_0^t S_{t-s} C \xi_{s-} dY_s, \quad t \in [0,T]. \tag{23}$$

The following decomposition of this equation is the starting point for our analysis: define the linear operator L on $\mathcal{N}^2(0,T;H)$ by

$$(L\xi)(t) := \int_0^t S_{t-s} B \xi_s dZ_s + \int_0^t S_{t-s} C \xi_{s-} dY_s \tag{24}$$

for all $t \in [0,T]$ and $\xi \in H$. Furthermore, set $\xi_t^{[0]} := S_t f$ such that $\xi^{[0]} \in \mathcal{N}^2(0,T;H)$. We obtain that (23) can be rewritten as the following equation in $\mathcal{N}^2(0,T;H)$

$$\xi = \xi^{[0]} + L\xi. \tag{25}$$

The operator L is a bounded linear operator and it is moreover quasinilpotent, as the following estimate shows.

Lemma 3.2. *Set* $\gamma := \sqrt{T}\bar{S}(\|B\|^2 + \|C\|^2)^{\frac{1}{2}}$. *Then, for all* $n \in \mathbb{N}$

$$\|L^n\|^{\frac{1}{n}} \leq \frac{\gamma}{(n!)^{\frac{1}{2n}}}. \tag{26}$$

The proof is given in the appendix A.1.

Lemma 3.3. *Equation* (25) *has a unique solution in* $\mathcal{N}^2(0,T;H)$,

$$\xi = (I-L)^{-1}\xi^{[0]} := \sum_{i=0}^{\infty} L^i \xi^{[0]}, \tag{27}$$

and $(I-L)^{-1} : \mathcal{N}^2(0,T;H) \to \mathcal{N}^2(0,T;H)$ *is a bounded linear operator:* $\|(I-L)^{-1}\| < \kappa$ *with* $\kappa = \frac{2}{\sqrt{3}} e^{2\gamma^2}$.

Proof. The crucial part in the proof of the lemma is the estimate

$$\sum_{n=0}^{\infty} \|L\|^n \leq \sum_{n=0}^{\infty} \frac{\gamma^n}{(n!)^{\frac{1}{2}}} = \sum_{n=0}^{\infty} 2^{-n} \frac{(2\gamma)^n}{(n!)^{\frac{1}{2}}} \leq \left(\Big(\sum_{n=0}^{\infty} 2^{-2n} \Big) \Big(\sum_{n=0}^{\infty} \frac{(2\gamma)^{2n}}{n!} \Big) \right)^{\frac{1}{2}} = \kappa,$$

which shows that the Volterra series $\sum_{i=0}^{n} L^i$ does in fact converge as $n \to \infty$. □

In view of Lemma 3.3 we can define the mapping $F : H \times \mathcal{U}^l \times \mathcal{U} \times \mathcal{S} \to \mathcal{N}^2(0,T;H)$ by

$$F(f,B,C,S) := \xi,$$

where ξ is the unique solution in $\mathcal{N}^2(0,T;H)$ of (23) with coefficients (f,B,C,S). The following result shows that F is continuous.

Proposition 3.4. *Consider sequences* $(f^{(n)})$, $(B^{(n)})$, $(C^{(n)})$ *and* $(S^{(n)})$ *in* H, $\mathcal{U}^l$, $\mathcal{U}$ *and* $\mathcal{S}$, *converging to* $f \in H$, $B \in \mathcal{U}^l$, $C \in \mathcal{U}$ *and* $S \in \mathcal{S}$, *respectively. Then,*

$$\Big| F(f^{(n)}, B^{(n)}\, C^{(n)}, S^{(n)}) - F(f,B,C,S) \Big|_T \to 0, \quad \textit{as} \quad n \to \infty.$$

Proof of Proposition 3.4. Since $S^{(n)} \to S$, $B^{(n)} \to B$ and $C^{(n)} \to C$, by the uniform boundedness principle there exist $\bar{N}$ and a constant $\bar{\gamma}$ such that

$$\sup_{t\in[0,T],n\geq\bar{N}} \Big\{ \|S_t\| \vee \|S_t^{(n)}\| \vee \|B\| \vee \|B^{(n)}\| \vee \|C\| \vee \|C^{(n)}\| \Big\} \leq \bar{\gamma}. \tag{28}$$

In the following, we only consider sufficiently large $n > \bar{N}$. Set

$$\xi := F(f, B, C, S), \quad \xi^{(n)} := F(f^{(n)}, B^{(n)}, C^{(n)}, S^{(n)}).$$

Together with $\xi_t^{[0,(n)]} := S_t^{(n)} f^{(n)}$ we define $L^{(n)}$ by

$$(L^{(n)}\xi)(t) := \int_0^t S_{t-s}^{(n)} B^{(n)}(\xi_s) dZ_s + \int_0^t S_{t-s}^{(n)} C^{(n)}(\xi_{s-}) dY_s$$

for all $\xi \in \mathcal{N}^2(0,T;H)$. Then, by the very definition of F,

$$\xi^{(n)} = \xi^{[0,(n)]} + L^{(n)}\xi^{(n)}, \quad \xi = \xi^{[0]} + L\xi.$$

Hence

$$\begin{aligned} \xi^{(n)} - \xi &= (\xi^{[0,(n)]} - \xi^{[0]}) + L^{(n)}(\xi^{(n)} - \xi) + (L^{(n)} - L)\xi \\ &= (I - L^{(n)})^{-1}\Big((\xi^{[0,(n)]} - \xi^{[0]}) + (L^{(n)} - L)\xi\Big). \end{aligned} \tag{29}$$

By Lemma 3.3 there exists a constant $\kappa = \kappa(\bar{\gamma})$, such that

$$\big\|(I - L^{(n)})^{-1}\big\| \leq \kappa. \tag{30}$$

Furthermore, as $S^{(n)} \in \mathcal{S}$,

$$\begin{aligned} \big|\xi^{[0,(n)]} - \xi^{[0]}\big|_T^2 &= \sup_{t\in[0,T]} \Big\| S_t^{(n)} f^{(n)} - S_t f \Big\|_H^2 \\ &\leq 2 \sup_{t\in[0,T]} \Big\| S_t^{(n)}(f^{(n)} - f) \Big\|_H^2 + 2 \sup_{t\in[0,T]} \Big\| (S_t^{(n)} - S_t) f \Big\|_H^2 \\ &\leq 2\bar{\gamma}^2 \|f^{(n)} - f\|_H^2 + 2 \sup_{t\in[0,T]} \Big\| (S_t^{(n)} - S_t) f \Big\|_H^2. \end{aligned}$$

The last term converges to zero as $(f^{(n)})$ and $(S^{(n)})$ converge to f and S, respectively.

Finally, we show that $(L^{(n)} - L)\xi$ converges to zero. From the definition of L and $L^{(n)}$ we obtain by the Itô-isometry that

$$\begin{aligned} &|(L^{(n)} - L)\xi|_T^2 = \sup_{t\in[0,T]} \mathbb{E}^0\Big(|((L^{(n)} - L)\xi)(t)|_H^2 \Big) \\ &= \sup_{t\in[0,T]} \mathbb{E}^0\Big(\int_0^t \Big\| (S_{t-s}^{(n)} B^{(n)} - S_{t-s} B)\xi_s \Big\|_{H^l}^2 ds + \int_0^t \Big\| (S_{t-s}^{(n)} C^{(n)} - S_{t-s} C)\xi_s \Big\|_H^2 ds \Big) \\ &\leq 2\Big[\sup_{t\in[0,T]} \mathbb{E}^0\Big(\int_0^t \Big\| S_{t-s}^{(n)}(B^{(n)} - B)\xi_s \Big\|_{H^l}^2 ds \Big) + \sup_{t\in[0,T]} \mathbb{E}^0\Big(\int_0^t \Big\| (S_{t-s}^{(n)} - S_{t-s}) B\xi_s \Big\|_{H^l}^2 ds \Big) \\ &+ \sup_{t\in[0,T]} \mathbb{E}^0\Big(\int_0^t \Big\| S_{t-s}^{(n)}(C^{(n)} - C)\xi_s \Big\|_H^2 ds \Big) + \sup_{t\in[0,T]} \mathbb{E}^0\Big(\int_0^t \Big\| (S_{t-s}^{(n)} - S_{t-s}) C\xi_s \Big\|_H^2 ds \Big) \Big] \\ &:= 2(E_1 + E_2 + E_3 + E_4). \end{aligned}$$

We consider the terms E_1 to E_4 separatly. Observe that by (22),

$$E_1 \leq \bar{\gamma}^2 \mathbb{E}^0\Big(\int_0^T \Big\| \Big(B^{(n)} - B \Big)(\xi_\tau) \Big\|_{H^l}^2 d\tau \Big).$$

As $B^{(n)}$ converges to B, we have for all $t \in [0,T]$ and $\omega \in \Omega$

$$\Big\| \Big(B^{(n)} - B \Big)(\xi_s(\omega)) \Big\|_{H^l}^2 \to 0.$$

In order to show that $E_1 \to 0$ as $n \to \infty$ we apply dominated convergence. Since $\|B^{(n)}\|, \|B\| \leq \bar{\gamma}$ we get

$$\Big\| (B^{(n)} - B)(\xi_s) \Big\|_{H^l}^2 \leq 4\bar{\gamma}^2 \|\xi_s\|_H^2$$

and the last term is integrable since $\mathbb{E}^0\Big(\int_0^T \|\xi_s\|_H^2 ds \Big) \leq T|\xi|_T^2 < \infty$.

In a similar way

$$\begin{aligned} E_2 \leq & \sup_{t\in[0,T]} \mathbb{E}^0\Big(\int_0^t \sup_{s\leq\tau\leq T} \Big\| (S^{(n)}_{\tau-s} - S_{\tau-s}) B\xi_s \Big\|_{H^l}^2 ds \Big) \\ = & \mathbb{E}^0\Big(\int_0^T \sup_{s\leq\tau\leq T} \Big\| (S^{(n)}_{\tau-s} - S_{\tau-s}) B\xi_s \Big\|_{H^l}^2 ds \Big) \end{aligned} \tag{31}$$

while uniform strong convergence of $S^{(n)}$ gives $\sup_{s\leq\tau\leq T} \Big\| (S^{(n)}_{\tau-s} - S_{\tau-s}) B\xi_s(\omega) \Big\|_{H^l}^2 \to 0$ for all $\omega \in \Omega$. As

$$\mathbb{E}^0\Big(\int_0^T \sup_{s\leq\tau\leq T} \Big\| (S^{(n)}_{\tau-s} - S_{\tau-s}) B\xi_s \Big\|_{H^l}^2 ds \Big) \leq 4\bar{\gamma}^4 \mathbb{E}^0\Big(\int_0^T \|\xi_s\|_H^2 ds \Big) \leq 4\bar{\gamma}^4 T|\xi|_T^2$$

and $|\xi|_T < \infty$ by Lemma 3.3 we obtain again by dominated convergence that $E_2 \to 0$. Analogously we obtain $E_3 \to 0$ and $E_4 \to 0$ and we conclude. □

Finally we turn to the

Proof of Theorem 3.1. Under Condition (20) the assumptions of Proposition 3.4 are clearly satisfied for the Galerkin approximation of the Zakai equation, as $P_n x \to x$, $P_n B P_n x \to Bx$ and $P_n C P_n x \to Cx$ for all $x \in H$. For the proof of the converse statement (the fact that (20) is also necessary for the convergence of the Galerkin approximation) we refer to the proof of Theorem 6.1 in Germani and Piccioni (1987). □

4 Numerical methods

In this section we discuss various aspects of the practical implementation of the Galerkin approximation of the Zakai equation. We begin with a few algorithms for the numerical solution of the SDE system (16). In Section 4.2 we consider the special class of basis functions constructed from Hermite polynomials. It turns out that the efficiency of the Galerkin approximation can be improved substantially if the scale and the location of the bases are changed adaptively, an issue which we discuss in Section 4.3.

4.1 Numerical solution of the Zakai equation

In order to solve the SDE system in (16) numerically, we discretize the system in time. As numerical schemes we consider the Euler-Maruyama method and the splitting-up method. While the Euler-Maruyama method is fast to implement it can become quite unstable if the time step is relatively large (see Figure 3). This difficulty can be overcome with the splitting-up method. Note that in practical filtering problems the observation often comes at discrete time points, so that the time-discretization step can not be chosen arbitrarily small.

Our aim is to approximate Equation (16). It will be convenient to use N instead of $Y = N - t$ as driver, so that the equation becomes

$$d\Upsilon_t^{(n)} = (A - C)\Upsilon_t^{(n)} dt + \sum_{\ell=1}^{l} B^\ell \Upsilon_t^{(n)} dZ_t + C\Upsilon_{t-}^{(n)} dN_t, \quad \Upsilon_0^{(n)} = q_0^{(n)}. \tag{32}$$

Consider the equidistant partition $0 = t_0 < t_1 < \cdots < t_K = T$ with step size $\Delta := T/K$. The approximation at time points $t_1, \ldots, t_K$ is denoted by $\Upsilon_1, \ldots, \Upsilon_K$ with $\Upsilon_0 := \Upsilon_0^{(n)}$ and $\Upsilon_k = (\psi_{k,1}, \ldots, \psi_{k,n})'$.

Euler-Maruyama method. The Euler-Maruyama method (EM) generalizes the Euler method to stochastic differential equations, see e.g. McLachlan and Krishnan (1997). It is described in the following algorithm:

Algorithm 4.1 (EM method)**.** For $k = 1, \ldots, K$, compute Υ_k from Υ_{k-1} by

$$\Upsilon_k = \Upsilon_{k-1} + D^{-1}\Big((A - C)\Upsilon_{k-1}\Delta + \sum_{\ell=1}^{l} B^\ell \Upsilon_{k-1}(Z_{t_k}^\ell - Z_{t_{k-1}}^\ell) + C\Upsilon_{k-1}(N_{t_k} - N_{t_{k-1}})\Big).$$

Splitting-up method. The splitting-up method (SU) is a numerical method based on semigroup theory. It decomposes the original SDE into stochastic and a deterministic equations which are easier to handle. We refer to Bensoussan, Glowinski, and Rascanu (1990) and Le Gland (1992) for further details in the case of continuous observations. Here we propose an extension of the method to the case with mixed observations. For simplicity, we assume $D = D^{-1} = I_n$, i.e. the basis $\{e_1, e_2, \ldots, e_n\}$ consists of orthonormal functions.

Intuitively, the SU method computes Υ_k from Υ_{k-1} in three steps: the first step uses only the dt-part of equation (32) and returns the solution of the SDE $d\Upsilon_t^1 = (A - C)\Upsilon_t^1 dt$. The solution of this equation on $[t_{k-1}, t_k]$ is the matrix exponential $\Upsilon_{t_k}^1 = \exp\big((A - C)\Delta\big)\Upsilon_{t_{k-1}}^1$. Step 2 incorporates the new information from Z via the linear SDE $d\Upsilon_t^2 = B\Upsilon_t^2 dZ_t$ with initial condition $\Upsilon_{t_{k-1}}^2 = \Upsilon_{t_k}^1$. The solution of this SDE is given by the matrix exponential

$$\Upsilon_{t_k}^2 = \exp\left(\sum_{\ell=1}^{l}\Big(B^\ell(Z_{t_k} - Z_{t_{k-1}}) - \frac{1}{2}(B^\ell)^2\Delta\Big)\right)\Upsilon_{t_k}^1.$$

The new jump information is incorporated via the linear equation $d\Upsilon_t^3 = C\Upsilon_{t-}^3 dN_t$, this time with initial condition $\Upsilon_{t_{k-1}}^3 = \Upsilon_{t_k}^2$, which gives

$$\Upsilon_{t_k}^3 = (I_n + C)^{(N_{t_k} - N_{t_{k-1}})}\Upsilon_{t_k}^2.$$

These steps lead to the following algorithm:

Algorithm 4.2 (SU method)**.** For $k = 1, \ldots, K$, compute Υ_k from Υ_{k-1} by

(1) Compute $\Upsilon_k^1 := \exp\big((A - C)\Delta\big)\Upsilon_{k-1}$.

(2) Compute $\Upsilon_k^2 := \exp\big(\sum_{\ell=1}^{l}(B^\ell(Z_{t_k} - Z_{t_{k-1}}) - \frac{1}{2}(B^\ell)^2\Delta)\big)\Upsilon_k^1$.

(3) Return $\Upsilon_k := (I_n + C)^{(N_{t_k} - N_{t_{k-1}})}\Upsilon_k^2$

4.2 Galerkin approximation based on Hermite polynomials

The choice of the basis functions has a large impact on the quality of the Galerkin approximation. Ahmed and Radaideh (1997) propose to use Gaussian series, i.e. a series build by densities of n-dimensional Gaussian distributions with different means and arbitrary positive, symmetric covariance matrices. Ahmed and Radaideh (1997) show that these are linearly independent and complete and hence they can be used to construct Galerkin approximations as described above.

In this paper we instead consider a basis computed from Hermite polynomials. This basis has a number of computational advantages over Gaussian series as will become clear below. We start by recalling some properties of Hermite polynomials, see e.g. Courant and Hilbert (1968). Define the Hermite polynomials by

$$f_i(x) = (-1)^i e^{x^2/2}\frac{d^i}{dx^i}e^{-x^2/2}, \quad x \in \mathbb{R} \tag{33}$$

with $i = 0, 1, 2, \ldots$. These polynomials are orthogonal with respect to the weighting function $\phi(x) := (2\pi)^{-1/2}\, e^{-x^2/2}$, as $\int_{\mathbb{R}} f_i(x) f_j(x)\phi(x)dx = i!\, 1_{\{i=j\}}$. Consequently, the functions $e_1, e_2, \ldots$ given by

$$e_i(x) := \sqrt{\frac{\phi(x)}{(i-1)!}} f_{i-1}(x), \quad x \in \mathbb{R}, \tag{34}$$

constitute an orthonormal basis of $L^2(\mathbb{R})$, which we call *Hermite basis*. In the following result we deduce the convergence of the Galerkin approximation with the use of (21).

Proposition 4.3. *Assume that* **(A1)** *holds and that* $q_t^{(n)}$ *is as in* (14) *with respect to the Hermite basis. Then, for any* $q_0 \in H$*,*

$$\sup_{t\in[0,T]} \mathbb{E}^0(\|q_t^{(n)} - q_t\|_H^2) \to 0, \quad \text{as} \quad n \to \infty.$$

The proof is given in Appendix A.1.

To actually obtain the Galerkin approximation under the Hermite basis one computes the coefficient matrices $A, B^1, \ldots, B^l, C, D$ as in (15) with respect to the Hermite basis and then solves (16) numerically by one of the methods described beforehand. In some special cases one obtains explicit formulas for the entries of the coefficient matrices as in the following example.

Example 4.4 (Kalman filter with point process observations)**.** Consider $d = m = l = 1$ and assume that $b(x) = bx$ and $\sigma(x) = \sigma$, such that

$$X_t = X_0 + \int_0^t bX_s ds + \sigma V_t$$

with $X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$. Moreover, consider $h(x) = hx$ and $\lambda(x) = \lambda x^2$ with $\lambda > 0$ such that the observation is given by $Z_t = \int_0^t hX_s ds + W_t$ and the doubly stochastic Poisson process N with

intensity $(\lambda X_t^2)_{t\geq 0}$. From (15) we compute that $d_{ji} = (e_i, e_j) = 1_{\{i=j\}}$ and

$$a_{ji} = (e_i, \mathcal{A}e_j) = \begin{cases} (-\frac{b}{2} + \frac{\sigma^2}{8})\sqrt{(j+2)(j+1)} & \text{if } i = j+2, \\ -\frac{b}{2} - \frac{\sigma^2(2j+1)}{8} & \text{if } i = j, \\ (\frac{b}{2} + \frac{\sigma^2}{8})\sqrt{j(j-1)} & \text{if } i = j-2; \end{cases} \tag{35}$$

$$b^1_{ji} = (e_i, h(\cdot)e_j) = \begin{cases} h\sqrt{j} & \text{if } i+1 = j, \\ h\sqrt{j+1} & \text{if } i-1 = j; \end{cases} \tag{36}$$

$$c_{ji} = (e_i, (\lambda(\cdot) - 1)e_j) = \begin{cases} \lambda\sqrt{j(j-1)} & \text{if } i = j-2, \\ 2\lambda j & \text{if } i = j, \\ \lambda\sqrt{(j+2)(j+1)} & \text{if } i = j+2. \end{cases} \tag{37}$$

Finally, one needs to express q_0 in terms of Hermite polynomials, which we do with some additional notation in (38).

Computation of moments. The i-th Hermite function is an polynomial of order i, and we denote by $\vartheta^i_0, \ldots, \vartheta^i_i$ the coefficients in the representation $f_i(x) = \sum_{k=0}^i \vartheta^i_k x^k$. We state the first number of coefficients in the appendix, section A.2. Conversely, any power of x can be represented as linear combination of Hermite polynomials and we write $x^i = \sum_{k=0}^i \iota^i_k f_k(x)$. With this we are able to compute the initial density in the previous example.

Example 4.5 (Example 4.4 continued)**.** With the above notation we are able to compute the projection of the initial density $q_0(x) = (2\pi\sigma_0^2)^{-1/2} \exp\big(-\frac{(x-\mu_2)^2}{2\sigma_0^2}\big)$ and obtain

$$q_{0j} = (q_0, e_j) = \frac{1}{\sqrt{j!}} \sum_{k=0}^{j} \vartheta^j_k \sum_{i=0}^{k} C^i_k c^i_1 c_2^{(k-i)} \iota^i_0, \tag{38}$$

where $c_1 = \sqrt{\frac{2\sigma_0^2}{2+\sigma_0^2}}$ and $c_2 = \frac{2\mu_0}{2+\sigma_0^2}$.

Recall from (18) that in order to compute mean and variance of the filter distribution via Galerkin approximation one needs to determine be integrals (x^j, e_i). For the Hermite basis this is done in the following lemma.

Lemma 4.6. *With respect to the Hermite basis we have that*

$$(x^j, e_i) = \frac{\sqrt{2}(2\pi)^{\frac{1}{4}}}{\sqrt{i!}} \sum_{k=0}^{i} \vartheta^i_k 2^{\frac{k}{2}} \iota_0^{k+j}, \quad j = 0, 1, \ldots \tag{39}$$

4.3 The adaptive Galerkin approximation

During the filtering process the conditional distribution $\pi_t(dx)$ typically changes location and scale. This can create problems for the Galerkin approximation with a fixed basis. For instance, the graphs in Figure 1 show that while the standard Hermite polynomials do approximate the density of a normal distribution well if the mean is close to zero and if the variance σ^2 lies between one and two, the fit becomes substantially worse if μ is substantially different from zero or if σ^2 is outside of the interval $[1, 2]$. Hence we propose an adaptive scheme, called *adaptive Galerkin approximation* (AGA), which improves the numerical performance of the Galerkin approach significantly.

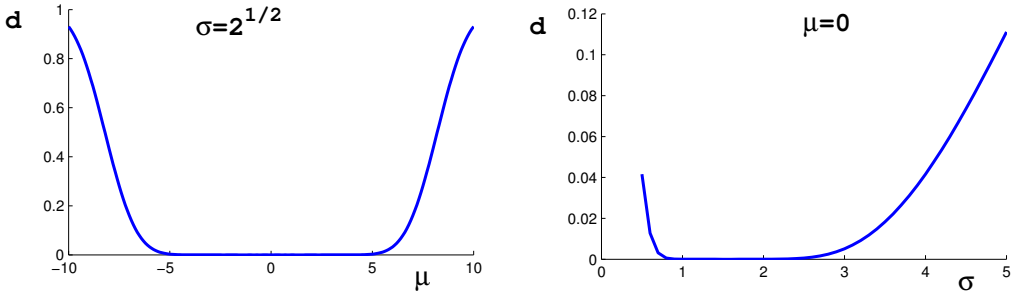

Figure 1: Comparison of the density p of a normal distribution with mean μ and variance σ^2 with its approximation $\hat{p} = \sum_{i=1}^{n}(p, e_i)e_i$ for different choices of μ and σ: the graphs show the distance $d := (\int(\hat{p} - p)^2 dx)^{1/2}$ as function of μ and as function of σ with fixed $\sigma = \sqrt{2}$ (left) and $\mu = 0$ (right). The approximation is bad if $\mu \notin (-5, 5)$ (left) or $\sigma \notin (0.9, 2)$ (right). The adaptive Galerkin method overcomes this difficulty.

Assume for simplicity that $l = 1$ and that the basis $\{e_i\} \subset D(\mathcal{A}^*)$ of H consists of orthonormal functions. We consider the equidistant time discretisation given by $t_k = T/K$, $k = 0, \dots, K$. The standard Galerkin approximation computes Υ_k at each time t_k. The AGA additionally adapts the location μ_k and the scale $\sigma_k > 0$ of the basis by choosing appropriate values for these parameters at every time step. Hence the method works with the adapted basis $\{e_1^k, e_2^k, \dots\}$ given by

$$e_i^k(x) := \frac{1}{\sqrt{\sigma_k}} e_i\Big(\frac{x - \mu_k}{\sigma_k}\Big), \quad x \in \mathbb{R}. \tag{40}$$

Similar to (15), we denote by A^k, B^k, C^k and D^k the matrices given by

$$a_{ji}^k = (e_i^k, \mathcal{A}e_j^k), \quad b_{ji}^k = (e_i^k, he_j^k), \quad c_{ji}^k = (e_i^k, (\lambda - 1)e_j^k), \quad d_{ji}^k = (e_i^k, e_j^k). \tag{41}$$

In algorithmic form the AGA can be described as follows:

Algorithm 4.7 (AGA)**. 1. Initialization**:

i) Set μ_0 and σ_0 using the initial density: $\mu_0 = \int x p_0(x) dx$ and $\sigma_0 = \big(\int (x - \mu_0)^2 p_0(x) dx\big)^{1/2}$, and define the basis functions e_i^0, $1 \le i \le n$, as in (40).

ii) Compute A^0, B^0, C^0 and D^0 according to (41).

iii) Compute $\Upsilon_0 = (\psi_{0,1}, \dots, \psi_{0,n})$ by (14): $\psi_{0,i} = (p_0, e_i^0)$.

2. Iteration: For $k = 1, \dots, K - 1$ do the following steps.

i) Compute $\tilde{\Upsilon}_k = (\tilde{\psi}_{k,1}, \dots, \tilde{\psi}_{k,n})'$ from Υ_{k-1} applying Algorithm 4.1 or 4.2, using the basis functions e_i^{k-1}, $1 \le i \le n$.

ii) Compute the following estimates of the conditional mean and standard deviation:

$$\hat{x}_k^{(n)} = \frac{\sum_{i=1}^n \tilde{\psi}_{k,i}(x, e_i^{k-1})}{\sum_{i=1}^n \tilde{\psi}_{k,i}^{(n)}(1, e_i^{k-1})} \quad \text{and} \quad \hat{\sigma}_k^{(n)} = \left(\frac{\sum_{i=1}^n \tilde{\psi}_{k,i}(x^2, e_i^{k-1})}{\sum_{i=1}^n \tilde{\psi}_{k,i}(1, e_i^{k-1})} - (\hat{x}_k^{(n)})^2\right)^{\frac{1}{2}}$$

If $|\hat{x}_k^{(n)} - \mu_{k-1}|$ and $|\hat{\sigma}_k^{(n)} - \sigma_{k-1}|$ are smaller than a given threshold, set $\Upsilon_k = \tilde{\Upsilon}_k$, $\mu_k = \mu_{k-1}$, $\sigma_k = \sigma_{k-1}$, $e_i^k := e_i^{k-1}$, $1 \le i \le n$. Let $k = k + 1$ and continue with the iteration (Step 2).

iii) Otherwise do a *transition of the basis* as follows: let $\mu_k := \hat{x}_k^{(n)}$, $\sigma_k := \hat{\sigma}_k^{(n)}$, define the new basis functions as in (40) and compute the matrices A^k, B^k, C^k and D^k according to (41). Finally, compute Υ_k by projecting $\tilde{q}_{t_k} = \sum_{i=1}^n \tilde{\psi}_{k,i} e_i^{k-1}$ on the new basis: let $\psi_{k,i} = (\tilde{q}_{t_k}, e_i^k)$ and set $\Upsilon_k = (\psi_{k,1}, \dots, \psi_{k,n})'$. Let $k = k+1$ and continue with Step 2.

The AGA provides better results compared to the standard Galerkin approximation (see the numerical experiments in Section 5 below) while it is typically more time consuming since the coefficient matrices in (41) need to be recomputed at every transition of the basis. However, in the AGA with respect to the Hermite basis the corresponding terms can be computed explicitly which leads to an efficient implementation of the AGA. In particular, when the coefficients b, σ^2, h and λ are of polynomial type, the corresponding coefficients can be computed explicitly with the aid of Lemma 4.6.

4.4 The multi-dimensional case

In this section we shortly sketch the extension to the multi-dimensional case. For an introduction of multi-dimensional Hermite polynomials, we refer to Berkowitz and Garner (1970). Here, we proceed by the following method: let $\{e_1, e_2, \dots\}$ denote the Hermite bases defined in Equation (34). This constitutes a basis of $L^2(\mathbb{R})$. Hence,

$$\Big\{ e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_d} : i_1, i_2, \dots i_d \in \mathbb{N}_0 \Big\}$$

is a Hilbert basis of $L^2(\mathbb{R}^d)$, where the tensor product is defined by $(e \otimes f)(x_1, x_2) := e(x_1) f(x_2)$. The n^d-dimensional subspace V_n turns out to be

$$V_n = \operatorname{span}\Big\{ e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_d} : i_1, i_2, \dots i_d \in \{0, 1, \dots, n-1\} \Big\}.$$

5 Numerical Experiments

General description. In this section we present results from a number of numerical case studies. The aim is to assess the performance of the Galerkin approximation relative to other methods (mostly particle filters) and to illustrate various practical aspects of the method. The basic setup of each numerical experiment is as follows. In Step 1 a trajectory $x = (x_t)_{0 \le t \le T}$ of the signal process (2) was generated using the Euler-Maryuama method. In Step 2 we generated for the given trajectory x from Step 1 a trajectory z of the continuous observation (4) and a trajectory n of the point process observation N. In Step 3 various variants of the Galerkin approximation were used to solve the corresponding Zakai equation for the conditional filter density. For comparison purposes the filter problem was also solved using a particle filter.

The performance of the numerical filtering algorithms was assessed in different ways:

- By design, the mean $\hat{x}_t$ of the filter distribution at time t minimizes the L^2-distance between the unobserved state X_t and the set $L^2(\Omega, \mathcal{F}_t^{Z,N}, \mathbb{P})$. This suggests the following performance criterion: Generate m independent trajectories x^j, z^j, n^j, $1 \le j \le m$ and solve numerically the ensuing filter problem for discrete time points $t_1, \dots, t_K$. Compute the so-called root mean square error (abbreviated RMSE) given by

$$\text{RMSE} = \Big(\frac{1}{mK} \sum_{j=1}^{m} \sum_{k=1}^{K} \| X^j(t_k) - \hat{x}^j(t_k) \|^2 \Big)^{\frac{1}{2}},$$

  Obviously, a filtering method that leads to a smaller RMSE can be considered to be more accurate.

- We can plot individual trajectories x, $\hat{x}$ and $\hat{\sigma}^2$ of the signal, and of the conditional mean and variance of the filter distribution. This allows for a pathwise comparison of different numerical methods.
- Finally, in some cases (e.g. Kalman-filtering for linear Gaussian models) the filter density p_t is known explicitly. In those cases we can compare the filter density $p_{t_k}(\cdot)$ and the approximation $p^k(\cdot)$ that is obtained by normalizing the numerical solution of the Zakai equation (see (17)).

Our numerical experiments with a one-dimensional signal process use the setup of Example 4.4. In this case X is a one-dimensional Ornstein-Uhlenbeck process with mean-reversion parameter b and volatility σ, and $h(\cdot)$ and $\lambda(\cdot)$ are of the form $h(x) = hx$ and $\lambda(x) = \lambda x^2$. The parameter values are as follows: $b = 0.5$, $\sigma = 2$, and the initial distribution of X is normal with $\mu = 5$, $\sigma^2 = 0.01$. Unless stated otherwise we took a time step $\Delta = 10^{-6}$ (essentially continuous observations). The values of h and λ vary with the experiments and are hence given in the captions of the graphs and tables.

We also considered the case of a multidimensional signal process of dimension $d = 5$. We assumed that the signal process has dynamics $dX_t = bX_t dt + \sigma dV_t$ for a 3-dimensional Brownian motion V. The observation process is three-dimensional and $h(x) = \tilde{h}x$. The matrices b, σ and $\tilde{h}$ are as follows:

$$b = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 1 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & -1 & -1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 1 \end{pmatrix}, \ \sigma = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \ h = \begin{pmatrix} 0.2 & 0.3 & 0.2 & 0.3 & 0.4 \\ 0.2 & 0.1 & 0.2 & 0.1 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 & 0.2 \end{pmatrix}.$$

and N is a one-dimensional Poisson process with intensity $0.1(X_t^1)^2 + 0.2(X_t^2)^2 + 0.3(X_t^3)^2 + 0.1(X_t^4)^2 + 0.1(X_t^5)^2$. The basis functions were chosen as indicated in Section 4.4.

Results. In the following we summarize the key findings from our numerical experiments; the outcome is described in detail in the captions of Figures 2 – Figure 7 below.

(i) For a one-dimensional state variable process the adaptive Galerkin approximation performs very well: given a sufficient number of basis functions the precision is equal to the precision of a particle filter, but the computation time is significantly lower. This can be seen from inspection of Table 1, where we give the RMSE and the computation time for various filtering algorithms and parameter values. The performance of the Galerkin approximation is further illustrated in Figure 2. In Figure 3 we consider the special case of the Kalman Bucy filter (no point process observations). Here the filter density is known explicitly and we can compare the approximation obtained via Galerkin approximation to the correct density. The figure clearly shows that the Galerkin approximation provides a good approximation to the overall density (and not just to the conditional mean $\hat{x}_t$).

(ii) Figure 7 for the case $d = 5$ indicates that the Galerkin approximation works reasonably well also for a higher dimensional signal process. However, the number of basis functions increases exponentially in d (at least for the basis chosen as in Section 4.4). It would be interesting to see if a further performance enhancement is possible if we choose a different basis, but this is left for further research.

(iii) The adaptive Galerkin method can bring a substantial performance enhancement if we consider examples with small observation noise and hence with rapidly moving scale and location of the filter distribution. This is clearly illustrated in Figure 5.

(iv) While computationally more involved, the splitting-up approximation is significantly more stable if the time-discretisation step Δ is moderately large, as is clearly shown in Figure 6.

N_G/N_P	5/20	10/50	15/100	20/1000
AGA(EM)	0.63 (0.1s)	0.42 (0.1s)	0.42 (0.1s)	0.42 (0.1s)
AGA(SU)	0.65 (2.4s)	0.43 (3.1s)	0.43 (3.9s)	0.43 (4.3s)
PF	0.46 (9s)	0.46 (22s)	0.42 (46s)	0.43 (472s)

Table 1: Performance comparison for different filter algorithms: we plot the RMSE and in brackets the computation time for two Galerkin filters and a particle filter. Here N_G represents the number of basis functions in the Galerkin approximation and N_P the number of particle in the particle filter. AGA(EM) respectively AGA(SU) stands for the adaptive Galerkin approximation with Euler-Maruyama approximation, respectively with splitting-up approximation for Equation (16). We used the values $\tilde{h} = \tilde{\lambda} = 0.1$ which corresponds to a relatively uninformative observation filtration; in computing the RMSE we used $m = 100$.

At this point we would like to stress that in many applications of filtering, observations arrive at discrete time points such as daily observations, and that one resorts to continuous-time filtering methods merely for convenience. This implies that Δ cannot be freely chosen by the analyst and it is important to have numerical methods that are robust with respect to the choice of Δ.

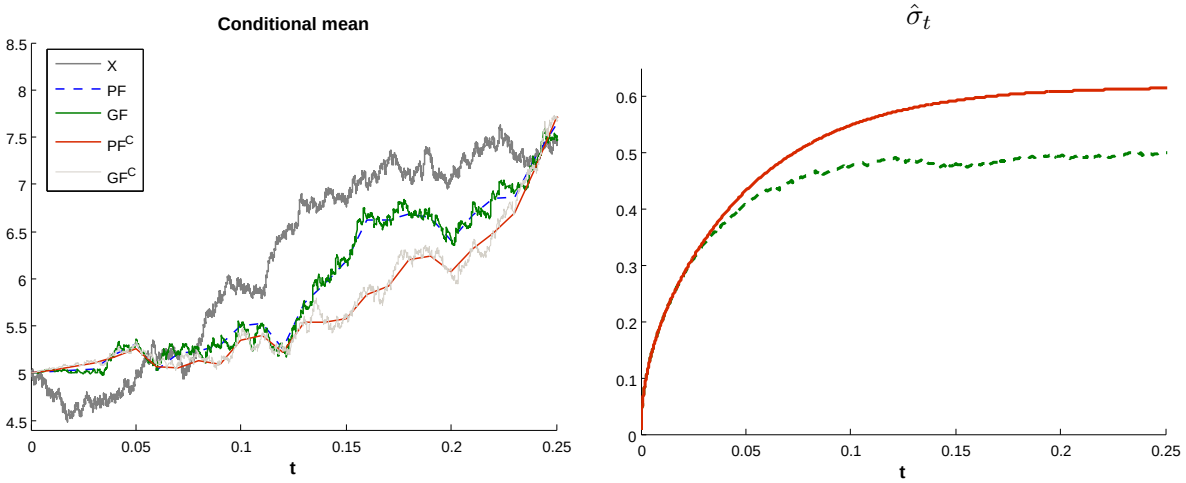

Figure 2: *Illustration of filtering and the value of point process information.* We choose $h = 5.5$, $\lambda = 10$. *Left:* trajectories of X and of $\hat{x}$ using the Galerkin method and particle filtering. Both methods perform well. We additionally plot both methods in the case where the point process information is neglected (PF^C and GF^C). In the *right* graph we illustrate the gain of using *point process observations*: we plot the trajectory of the conditional standard deviation $\hat{\sigma}_t$ for the case with only continuous observation $\tilde{\lambda} = 0$ and with continuous and point process observations ($\lambda = 10$, lower trajectory). The approximation by the two methods are very close in that case. Clearly, including point process information reduces the conditional standard deviation significantly.

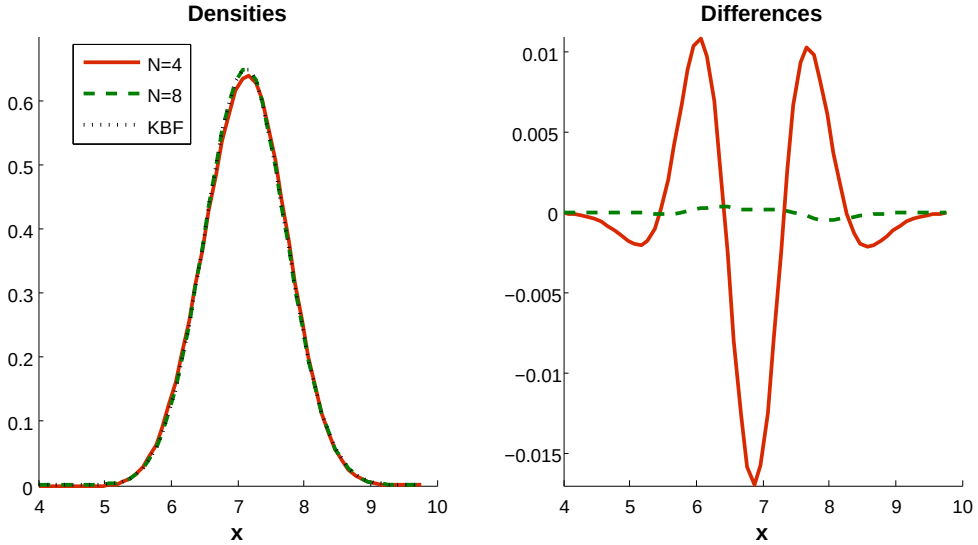

Figure 3: *Comparison of the theoretical filter density and the AGA.* We consider a Hermite basis and the case of purely continuous observations ($\lambda = 0$). In this case, the filter problem has an explicit solution that can be computed with the Kalman-Bucy filter (KBF). *Left:* filter densities; *right:* approximation error. The adaptive Galerkin approximation (AGA) is very close to the explicit solution for $N \geq 8$.

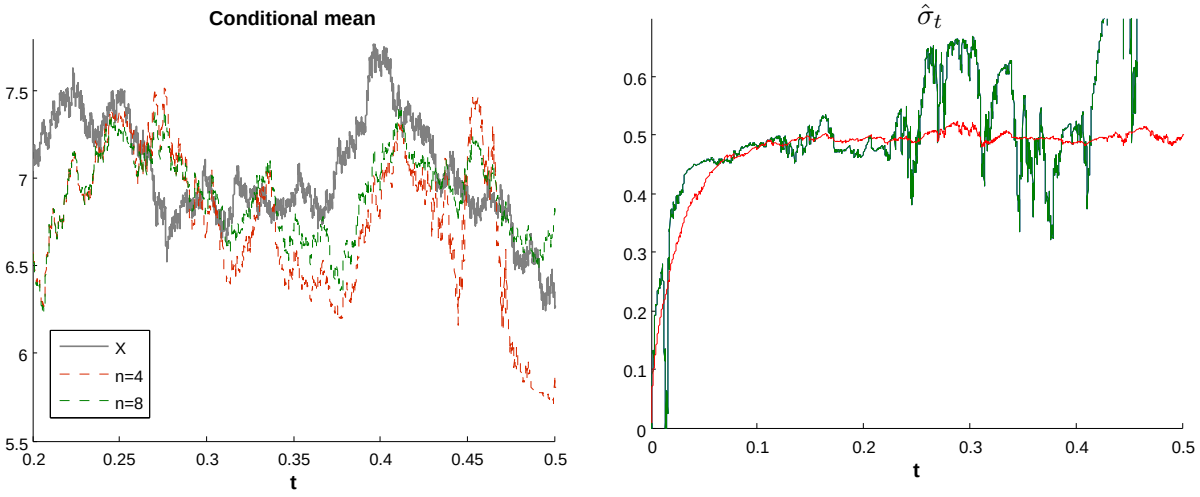

Figure 4: Filter estimate (left) and conditional standard deviation (right) for a varying number $N_G = n$ of basis functions in the adaptive Galerkin approximation (AGA). Here $h = 5.5$, $\lambda = 10$ and we use the AGA with $n = 4, 8, 16$ basis functions. The case $n = 4$ shows a bad performance; the filter with 8 and 16 basis functions performs reasonably well. The right plot indicates that plots of the conditional variance can be a useful tool for determining if the number of basis functions used is appropriate.

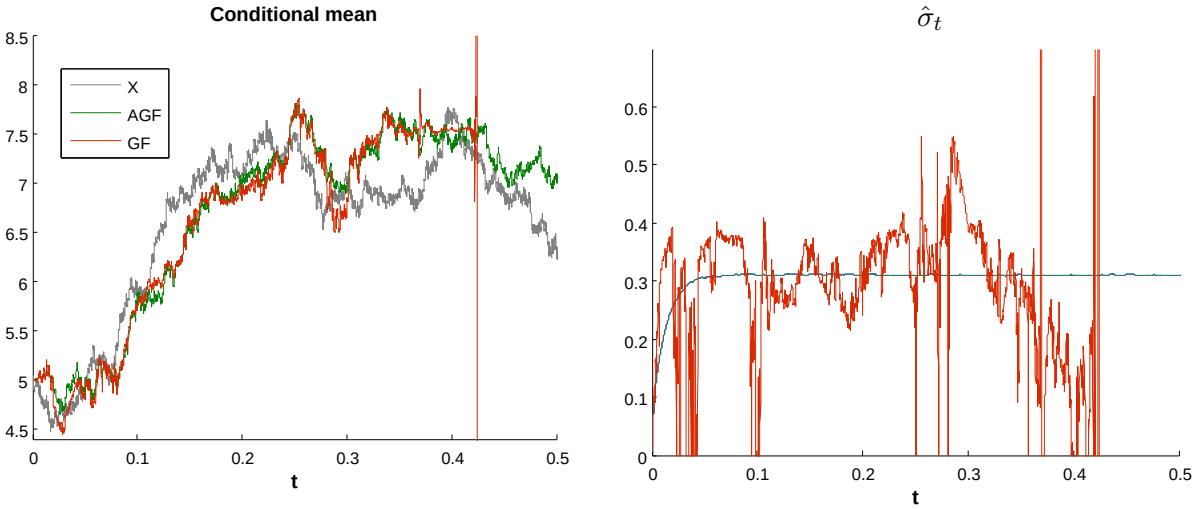

Figure 5: *Comparison of the ordinary Galerkin approximation (GF) with the adapted Galerkin approximation (AGF).* Both approximations work with $N = 20$ Hermite basis functions. In this example $h = 20$, $\lambda = 10$ (these parameter values correspond to a very low observation noise). Note that the ordinary Galerkin filter performs poorly, whereas the AGA performs reasonably well.

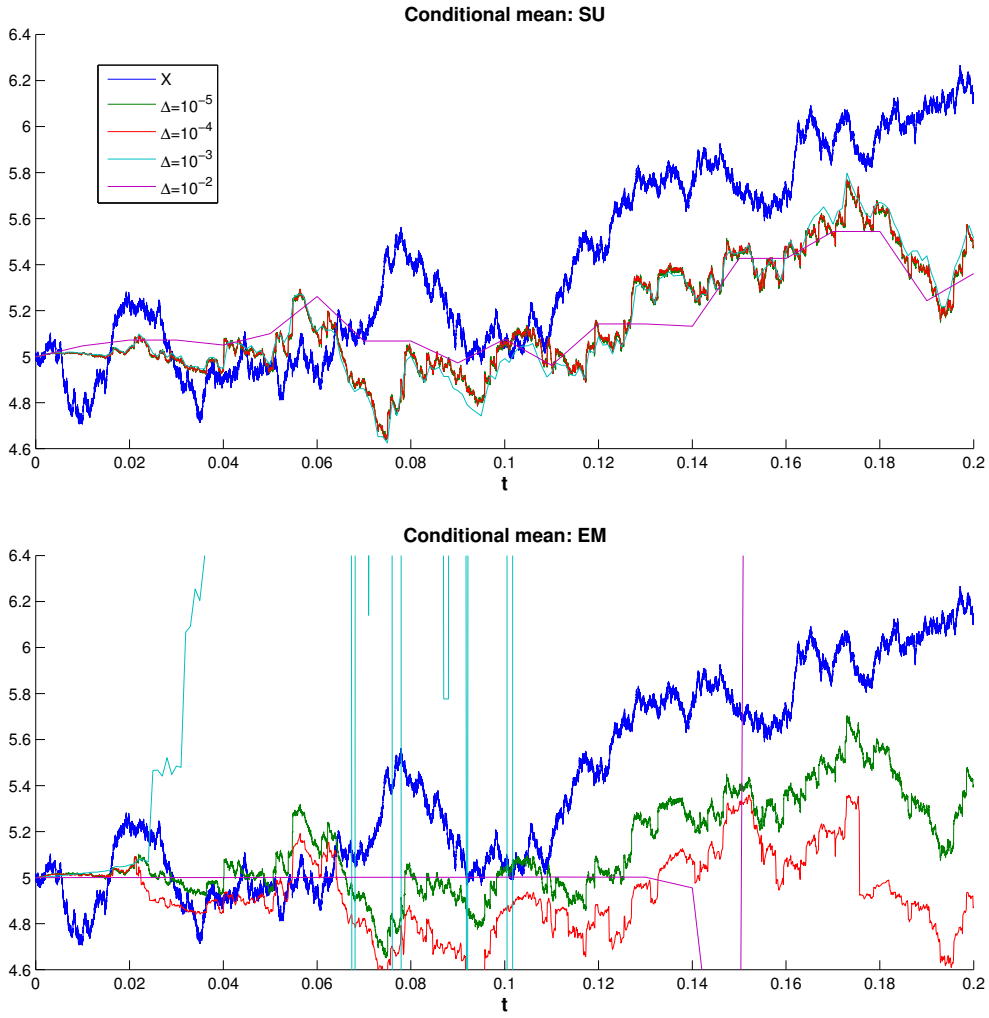

Figure 6: *Comparison of splitting-up approximation and Euler-Maruyama approximation.* In this figure, we compare the results obtained by these two methods for different Δ. The results obtained coincide for $\Delta = 10^{-5}$, but the splitting-up approximation is more stable when Δ is large. In particular, the splitting up method provides a good estimate even for $\Delta = 10^{-2}$.

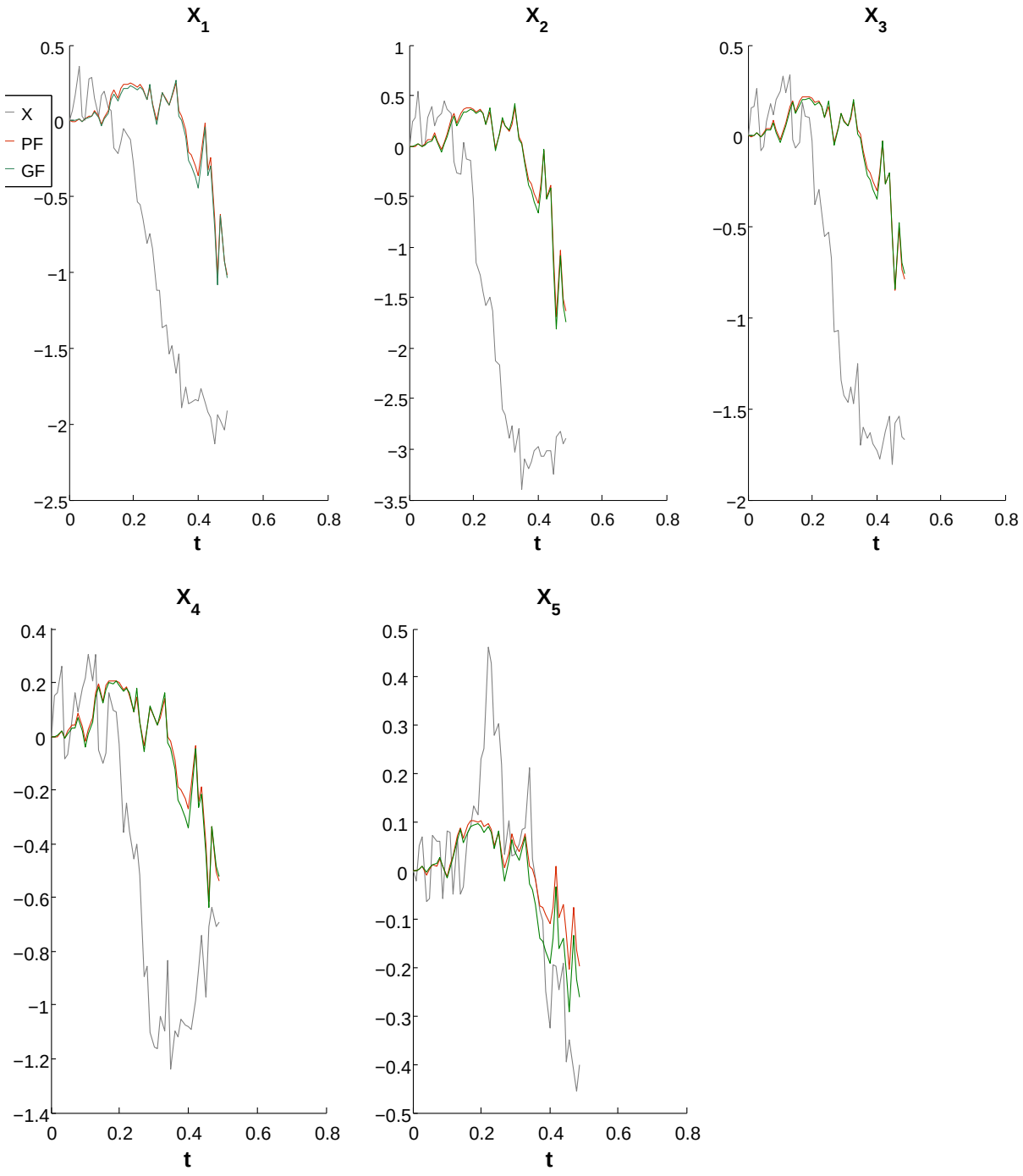

Figure 7: *The multidimensional case.* comparison of the adaptive Galerkin approximation (GF) with the particle filter (PF) for the case of a multi-dimensional signal process X. We show plots of the conditional mean for a basis of size $n = 4^5 = 1024$. For the particle filter we took 10^3 particles. The computation time was 12 seconds (particle filter) and 9 seconds (AGA) while the obtained results are very close.

A Appendix

A.1 Proofs

This appendix contains the proofs of Lemma 3.2 and Proposition 4.3.

Proof of Lemma 3.2. Our aim is to show that for all $n \in \mathbb{N}$ it holds that

$$\|L^n\|^{\frac{1}{n}} \leq \frac{\sqrt{T}\bar{S}(\|B\|^2 + \|C\|^2)^{\frac{1}{2}}}{(n!)^{\frac{1}{2n}}}. \tag{42}$$

The operator L was defined in (24). We rewrite $L\xi$ as

$$(L\xi)(t) =: \int_0^t E(t,s)\xi_{s-}dM_s$$

with the $l+1$-dimensional martingale $M := (Z^\top, Y)^\top$ and $E(t,s) := S_{t-s}(B^\top, C)^\top$. Iterative application of L gives that

$$(L^n\xi)(t) = \int_0^t E(t,t_1)\Big(\int_0^{t_1} E(t_1,t_2)\Big(\dots\int_0^{t_{n-1}} E(t_{n-1},t_n)\xi_{t_n}dM_{t_n}\dots\Big)dM_{t_2}\Big)dM_{t_1}.$$

To compute $|L^n\xi|_T = \sup_{t\in[0,T]}\left(\mathbb{E}^0\big(\|(L^n\xi)(t)\|_H^2\big)\right)^{1/2}$, note that the quadratic variation of M is $< M >_t = I_{l+1}t$ where I_{l+1} is the identity matrix on $\mathbb{R}^{l+1}$. The Itô-isometry therefore yields

$$\begin{aligned}
&\mathbb{E}^0\big(\|(L^n\xi)(t)\|_H^2\big)\\
&\quad= \mathbb{E}^0\Big(\int_0^t \Big\|E(t,t_1)\Big(\int_0^{t_1} E(t_1,t_2)\Big(\dots\int_0^{t_{n-1}} E(t_{n-1},t_n)\xi_{t_n}dM_{t_n}\dots\Big)dM_{t_2}\Big)\Big\|^2 dt_1\Big)\\
&\quad\leq \bar{S}^2(\|B\|^2+\|C\|^2)\int_0^T \mathbb{E}^0\Big(\int_0^{t_1}\Big\|E(t_1,t_2)\Big(\dots\int_0^{t_{n-1}} E(t_{n-1},t_n)\xi_{t_n}dM_{t_n}\dots\Big)dM_{t_2}\Big\|_H^2\Big)dt_1\\
&\quad\leq \bar{S}^{2n}(\|B\|^2+\|C\|^2)^n|\xi|_T^2\cdot\int_0^T\int_0^{t_1}\dots\int_0^{t_{n-1}} dt_n,\dots,dt_1\\
&\quad= \bar{S}^{2n}(\|B\|^2+\|C\|^2)^n|\xi|_T^2\cdot\frac{T^n}{n!}
\end{aligned}$$

and we obtain (42). □

Proof of Proposition 4.3. As remarked after Theorem 3.1, in (21) the claim is proved if we can show that $\cup_n V_n$ is dense in $V = H^1(\mathbb{R}^d)$. Here $V_n = \text{span}\{e_1,\dots,e_n\}$ and $V = H^1(\mathbb{R}^d)$. Let C_0^∞ be the set of smooth functions with compact support. Then, by Proposition 1 and Theorem 4 in Bongioanni and Torrea (2006), there exist for all $u \in C_0^\infty$ a sequence $u_n \in \cup_n V_n$ such that $\|u-u_n\|_V \to 0$ as $n\to\infty$. Since C_0^∞ is dense in V the claim follows. □

A.2 Hermite polynomials

For convenience of the reader we state some Hermite polynomials: $f_0(x) = 1$, $f_1(x) = x$, $f_2(x) = x^2 - 1$ and $f_3(x) = x^3 - 3x$. Often it is useful to exploit the following reccurence relation:

$$f_{i+1}(x) = xf_i(x) - if_{i-1}(x), \quad f_i^{'}(x) = if_{i-1}(x). \tag{43}$$

Proof of Lemma 4.6. First, note that,

$$\begin{aligned}
(1, e_i) =& \frac{1}{\sqrt{i!}} \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{4}} e^{-\frac{x^2}{4}} f_i(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{-\frac{1}{4}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} f_i(\sqrt{2}x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{-\frac{1}{4}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} x^k dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k f_j(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} f_j(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} f_j(x) f_0(x) dx \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \sum_{j=0}^{k} \iota_j^k \delta_{0,j} \\
=& \frac{1}{\sqrt{i!}} \sqrt{2} (2\pi)^{\frac{1}{4}} \sum_{k=0}^{i} \vartheta_k^i 2^{\frac{k}{2}} \iota_0^k.
\end{aligned}$$

since any power of x can be represented by the Hermite polynomials; we use the notation from page 14. An analogous argument with the constant function 1 replaced by x^j gives the result. □

References

Ahmed, N. and S. Radaideh (1997). A powerful numerical technique solving the Zakai equation for nonlinear filtering. *Dynamics and Control 7*(3), 293–308.

Bain, A. and D. Crisan (2009). *Fundamentals of Stochastic Filtering*. New York: Springer.

Bensoussan, A., R. Glowinski, and A. Rascanu (1990). Approximation of the Zakai equation by the splitting up method. *SIAM J. Control Optim. 28*(6), 1420–1431.

Berkowitz, S. and F. J. Garner (1970). The calculation of multidimensional Hermite polynomials and Gram-Charlier coefficients. *Mathematics of Computation 24*, 537–545.

Bongioanni, B. and J. L. Torrea (2006). Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci. 116*(3), 337–360.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer, New York.

Budhiraja, A., L. Chen, and C. Lee (2007). A survey of numerical methods for nonlinear filtering problems. *Physica D 230*, 27–36.

Courant, R. and D. Hilbert (1968). *Methoden der Mathematischen Physik*. Springer.

Crisan, D., P. D. Moral, and T. Lyons (1999). Discrete filtering using branching and interacting particle systems. *Markov Processes and Related Fields 5*(3), 293–318.

Cvitanic, J., R. Liptser, and B. Rozovski (2006). A filtering approach to tracking volatility from prices observed at random times. *The Annals of Applied Probability 16*, 1633–1652.

Di Masi, G. and W. Runggaldier (1982). On approximation methods for nonlinear filtering. In *Nonlinear filtering and stochastic control*, Volume 972 of *Lecture-Notes in Mathematics*, Berlin, pp. 249–259. Springer.

Dupuis, P. G. and H. J. Kushner (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer.

Frey, R. and W. Runggaldier (2001). A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *International Journal of Theoretical and Applied Finance 4*, 199–210.

Frey, R. and W. Runggaldier (2010). Pricing credit derivatives under incomplete information: a nonlinear filtering approach. *Finance and Stochastics 14*, 495–526.

Frey, R. and T. Schmidt (2010). Pricing and hedging of credit derivatives via the innovations approach to nonlinear filtering. to appear in *Finance and Stochastics*.

Frey, R. and T. Schmidt (2011). Filtering and incomplete information in credit risk. In T. Bielecki, D. Brigo, and F. Patras (Eds.), *Recent Advancements in the Theory and Practice of Credit Derivatives*, pp. 185–218. Bloomberg Press.

Germani, A. and M. Piccioni (1984). A Galerkin approximation for the Zakai equation. *Lecture Notes in Control and Information Sciences 59*, 415–423.

Germani, A. and M. Piccioni (1987). Finite-dimensional approximations for the equations of nonlinear filtering derived in mild form. *Applied Mathematics and Optimization 16*(1), 51–72.

Gobet, E., G. Pagès, H. Pham, and J. Printems (2006). Discretization and simulation of Zakai equation. *SIAM Journal on Numerical Analysis 44*, 2505–2538.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*. Academic Press.

Le Gland, F. (1992). Splitting-up approximation for SPDEs and SDEs with application to nonlinear filtering. *Lecture Notes in Control and Information Sciences 176*, 177–187.

Lototsky, S. V. (2006). Wiener chaos and nonlinear filtering. *Appl. Math. Optim. 54*(3), 265–291.

McLachlan, G. J. and T. Krishnan (1997). *The EM Algorithm and Extensions* (2nd ed.). John Wiley & Sons. New York.

Pardoux, E. (1979a). Filtering of a diffusion process with Poisson-type observation. In *Stochastic Control Theory and Stochastic Differential Systems*, Volume 16 of *Lecture Notes in Control and Information Science*, pp. 510–518. Berlin- Heidelberg: Springer.

Pardoux, E. (1979b). Stochastic partial differential equations and filtering of diffusion processes. *Stochastics 3*, 127–167.

Peszat, P. and J. Zabczyk (2007). *Stochastic Partial Differential Equations with Lévy Noise*. Cambridge University Press.

Prato, G. D. and J. Zabczyk (1992). *Stochastic Equations in Infinte Dimensions*. Cambridge University Press.

Xu, L. (2010). *On Galerkin Approximations for the Zakai Equation with Diffusive and Point Process Observations*. Ph. D. thesis, Leipzig University.