QUANTITATIVE FINANCE RESEARCH CENTRE

Quasi-exact Approximation of Hidden Markov Chain Filters

Eckhard Platen and Renata Rendek

Quasi-exact Approximation

of Hidden Markov Chain Filters

Eckhard Platen [1] and **Renata Rendek**[1]

In honor of Gopinath Kallianpur's 85th birthday.

October 15, 2009

Abstract. This paper studies the application of exact simulation methods for multi-dimensional multiplicative noise stochastic differential equations to filtering. Stochastic differential equations with multiplicative noise naturally occur as Zakai equation in hidden Markov chain filtering. The paper proposes a quasi-exact approximation method for hidden Markov chain filters, which can be applied when discrete time approximations, such as the Euler scheme, may fail in practice.

---

[1]University of Technology Sydney, School of Finance & Economics and Department of Mathematical Sciences, PO Box 123, Broadway, NSW, 2007, Australia

1   Introduction

Filtering is a powerful methodology that aims to extract information about hidden variables from observed data. In particular, the variables of interest can occur as realizations of continuous time hidden Markov chains. The observation of the hidden, time varying variables can be perturbed by other stochastic processes. Therefore, the need arises for filtering methods, which allow to detect in an adaptive manner as much information about the hidden variables as possible.

The systematic construction and investigation of filters for hidden Markov chains goes back to Wonham (1965), Zakai (1969) and Fujisaki, Kallianpur & Kunita (1972). The stochastic filtering theory is the subject of a seminal monograph by Kallianpur (1980). Later the question of finding discrete-time approximations for optimal filters was considered by Clark & Cameron (1980), Newton (1986, 1991) and Kloeden, Platen & Schurz (1993). Moreover, the application of the balanced implicit method to stochastic differential equations (SDEs) in filtering has been considered in Fischer & Platen (1999).

Unfortunatelly, in all the discrete time approximation methods numerical stability issues may arise, which in turn can prevent their use in important practical situations. There is a wide range of literature which deals with the issue of numerical stability. In particular, implicit or predictor-corrector methods are used to control the propagation of errors. We refer for this approach to papers by Talay (1982), Klauder & Petersen (1985), Milstein (1988), Hernandez & Spigler (1993), Saito & Mitsui (1993), Kloeden & Platen (1992), Milstein, Platen & Schurz (1998), Higham (2000), Alcock & Burrage (2006), Bruti-Liberati & Platen (2008) and Platen & Shi (2008). There are various numerical schemes that perform well on some SDEs for certain parameter ranges and sufficiently small step sizes. However, numerical instabilities caused by the presence of multiplicative noise in the Zakai SDE are, in general, not easily overcome by most of these methods. For larger time step sizes numerical instabilities potentially arise, which limits the applicability of the filtering methodology. The issue of numerical stability can be, however, circumvented when it is possible to simulate exact or quasi-exact solutions of the Zakai equation at the observation points. The following study is focused on the application of such exact solution method to the multi-dimensional, multiplicative noise SDEs arising from the Zakai equation of hidden Markov chain filtering. This critical task has to be effectively and reliably performed to make a hidden Markov chain filter successful. Observations are, in practice, only available at discrete-time instants. Therefore, they may occur at any given frequency. If the observations are rare and the standard discrete time numerical schemes fail, then the proposed quasi-exact solution method for hidden Markov chain filters may still result in reliable estimates.

For illustration and comparison we will consider an example involving the discrete-time numerical solution of hidden Markov chain filters which was discussed also in Fischer & Platen (1999). Here the hidden signal process is corrupted by a

Brownian motion. We compare selected discrete time numerical solutions of the Zakai equation to some quasi-exact solution proposed in this paper. This study illustrates the practical usefulness of the quasi-exact solution from the engineering point of view.

This paper is organized as follows: in Section 2 we describe in detail the Wonham filter problem and the resulting Zakai equation for hidden Markov chain filters. Section 3 presents the exact solution of some multi-dimensional, multiplicative noise SDEs with a view towards hidden Markov chain filters, while Section 4 covers discrete time numerical approximations including some of those discussed in Fischer & Platen (1999). In Section 5 we consider an example of hidden Markov chain filters and Section 6 concludes.

2   Hidden Markov Chain Filters

Let us now introduce filters for hidden, continuous time, finite state Markov chains. Let $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ with $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0,T]}$ and $T \in [0, \infty)$ be the underlying filtered probability space and suppose that the *hidden state process* $\xi = \{\xi_t,\, t \in [0,T]\}$ is a continuous time, homogeneous Markov chain on the finite state space $\mathcal{X} = \{a_1, a_2, \ldots, a_d\}$, $d \in \{0, 1, \ldots\}$. Its d-dimensional probability vector $\boldsymbol{p}(t) = (p_1(t), \ldots, p_d(t))^\top$ at time t, with components

$$p_i(t) = P(\xi_t = a_i) \tag{2.1}$$

for each $a_i \in \mathcal{X}$, satisfies then the vector ordinary differential equation (ODE)

$$\frac{d\boldsymbol{p}(t)}{dt} = \boldsymbol{A}\,\boldsymbol{p}(t), \tag{2.2}$$

where $\boldsymbol{A}$ is the intensity matrix and the initial probability vector equals $\boldsymbol{p}(0) = \boldsymbol{p}_0$. The solution of the vector ODE (2.2) is then of the form

$$\boldsymbol{p}(t) = \exp\{\boldsymbol{A}t\}\boldsymbol{p}_0. \tag{2.3}$$

Here $\exp\{\cdot\}$ denotes the matrix exponential defined as

$$\exp\{\boldsymbol{A}\} = \sum_{k=0}^{\infty} \boldsymbol{A}^k \frac{1}{k!}. \tag{2.4}$$

In addition, suppose that the m-dimensional *observation process* $\boldsymbol{W} = \{\boldsymbol{W}_t,\, t \in [0,T]\}$ is the solution of the SDE

$$d\boldsymbol{W}_t = \boldsymbol{h}(\xi_t)\,dt + d\boldsymbol{W}_t^* \tag{2.5}$$

for $t \in [0, T]$ with $\boldsymbol{W}_0 = \boldsymbol{W}_0^* \in \Re^m$, $m \in \{1, 2, \ldots\}$. This type of disturbance of a signal by a Wiener process is called a *Wonham filter problem*, see Wonham (1965).

In the SDE (2.5) the noise process $\boldsymbol{W}^* = \{\boldsymbol{W}_t^*, \, t \in [0, T]\}$ with $\boldsymbol{W}_0^* = 0$ is an m-dimensional standard Wiener process with respect to the real world probability measure P. The Wiener process $\boldsymbol{W}^*$ is assumed to be independent of the hidden state process ξ. Finally, let

$$\mathcal{Y}_t = \sigma \left\{ \boldsymbol{W}_s, \, s \in [0, t] \right\}$$

denote the observation sigma-algebra generated by the observations $\boldsymbol{W}_s$ for $s \in [0, t]$. This means, $\underline{\mathcal{Y}} = (\mathcal{Y}_t)_{t \in [0,T]}$ is the filtration that represents the release of observed information, whereas $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0,T]}$ with $\mathcal{A}_t = \sigma\{\xi_s, \boldsymbol{W}_s : \, s \in [0, t]\}$ expresses the evolution of the total information.

Our task is to filter as much information about the hidden state process ξ as we can from the observation process $\boldsymbol{W}$. With this aim we shall evaluate for a given function $g : \mathcal{X} \to \Re$ the *Wonham filter*, which is the conditional expectation

$$\hat{g}(\xi_T) = E \left(g(\xi_T) \, \middle| \, \mathcal{Y}_T \right)$$

with respect to the real world probability P. The function $g(\cdot)$ could be chosen, for instance, as an indicator function $\mathbf{1}_{\{\xi_T = a_i\}}$, which yields $\hat{g}(\xi_T)$ as a probability. It could also represent a power function $(\xi_T)^q$, which leads to $\hat{g}(\xi_T)$ describing the qth moment of the hidden Markov chain.

By application of the Girsanov transformation one obtains a probability measure $\dot{P}$, where

$$d\dot{P} = L_T^{-1} \, dP \tag{2.6}$$

with

$$L_T = \exp \left\{ -\frac{1}{2} \int_0^T |\boldsymbol{h}(\xi_s)|^2 \, ds + \int_0^T \boldsymbol{h}(\xi_s)^\top \, d\boldsymbol{W}_s \right\} \tag{2.7}$$

such that $\boldsymbol{W}$ is a Wiener process with respect to $\dot{P}$, while L is assumed to be a martingale. Here $L_T = \frac{dP}{d\dot{P}}$ is the corresponding Radon-Nikodym derivative. Note that we express in this situation the real world probability measure P in terms of the new probability measure $\dot{P}$ for which $\boldsymbol{W}$ is a standard vector Wiener process given by (2.5).

Let us introduce the *unnormalized conditional probability* $\sigma(\xi_t)^i$ for the state $a_i \in \mathcal{X}$ at time t by the conditional expectation

$$\sigma(\xi_t)^i = \dot{E} \left(\mathbf{1}_{\{\xi_t = a_i\}} L_t \, \middle| \, \mathcal{Y}_t \right) \tag{2.8}$$

with respect to the new probability measure $\dot{P}$ for $i \in \{1, 2, \ldots, d\}$ and $t \in [0, T]$. It follows from the *Kallianpur-Striebel formula*, see Fujisaki, Kallianpur & Kunita (1972), that the conditional probabilities of ξ_t given in $\mathcal{Y}_t$ are

$$P \left(\xi_t = a_i \, \middle| \, \mathcal{Y}_t \right) = E \left(\mathbf{1}_{\{\xi_t = a_i\}} \, \middle| \, \mathcal{Y}_t \right) = \frac{\sigma(\xi_t)^i}{\sum_{k=1}^d \sigma(\xi_t)^k} \tag{2.9}$$

4

for $a_i \in \mathcal{X}$ and $t \in [0, T]$. Here the d-dimensional process $\boldsymbol{\sigma}(\xi) = \{\boldsymbol{\sigma}(\xi_t) = (\sigma(\xi_t)^1, \ldots, \sigma(\xi_t)^d)^\top, t \in [0, T]\}$ of unnormalized conditional probabilities satisfies the *Zakai equation*

$$\boldsymbol{\sigma}(\xi_t) = \boldsymbol{p}(0) + \int_0^t \boldsymbol{A}\,\boldsymbol{\sigma}(\xi_s)\,ds + \sum_{k=1}^m \int_0^t \boldsymbol{D}^k\,\boldsymbol{\sigma}(\xi_s)\,dW_s^j \tag{2.10}$$

for $t \in [0, T]$. This is a homogeneous linear Itô SDE. In (2.10) $\boldsymbol{D}^k$ is the $d \times d$ diagonal matrix with iith component $h_k(a_i)$ for $i \in \{1, 2, \ldots, d\}$ and $k \in \{1, 2, \ldots, m\}$.

The least-squares estimate at time t for $g(\xi_t)$ with respect to the given observations at time t, that is with respect to the sigma-algebra $\mathcal{Y}_t$, is then the *Wonham filter*, which is given by the conditional expectation

$$\hat{g}(\xi_t) = E\left(g(\xi_t)\,\middle|\,\mathcal{Y}_t\right) = \frac{\sum_{k=1}^d g(a_k)\,\sigma(\xi_t)^k}{\sum_{k=1}^d \sigma(\xi_t)^k} \tag{2.11}$$

for $t \in [0, T]$.

3 Quasi-exact Filters

Let us consider the following d-dimensional multiplicative noise SDE

$$d\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_t dt + \sum_{k=1}^m \boldsymbol{D}^k \boldsymbol{X}_t dW_t^k, \tag{3.1}$$

with a solution that is representing a vector geometric Brownian motion, where $\boldsymbol{X} = \{\boldsymbol{X}_t = (X_t^1, X_t^2, \ldots, X_t^d)^\top, t \in [0, \infty)\}$, $\boldsymbol{A} = [a^{i,j}]_{i,j=1}^d$ and $\boldsymbol{D}^k = [d^{k,i,j}]_{i,j=1}^d$, $k \in \{1, 2, \ldots, m\}$. Here, $W^k, k \in \{1, 2, \ldots, m\}$, are the elements of the vector SDE (2.5), that describes the observation process.

It turns out that if the matrices $\boldsymbol{A}, \boldsymbol{D}^1, \boldsymbol{D}^2, \ldots, \boldsymbol{D}^m$ are constant and commute, that is, if

$$\boldsymbol{A}\boldsymbol{D}^k = \boldsymbol{D}^k\boldsymbol{A} \quad \text{and} \quad \boldsymbol{D}^k\boldsymbol{D}^n = \boldsymbol{D}^n\boldsymbol{D}^k \tag{3.2}$$

for all $k, n \in \{1, 2, \ldots, m\}$, then an explicit solution of the SDE (3.1) can be expressed by

$$\boldsymbol{X}_t = \boldsymbol{\Psi}_t \boldsymbol{X}_0, \tag{3.3}$$

for $t \in [0, \infty)$. Here, $\boldsymbol{\Psi}_t$ is the matrix exponential

$$\boldsymbol{\Psi}_t = \exp\left\{\boldsymbol{A}t - \frac{1}{2}\sum_{l=1}^m \left(\boldsymbol{D}^l\right)^2 t + \sum_{r=1}^m \boldsymbol{D}^r W_t^r\right\}, \tag{3.4}$$

for $t \in [0, \infty)$.

The proof follows from the Itô formula applied to (3.3). Note that

$$
\begin{aligned}
d\boldsymbol{X}_t = d\left[\boldsymbol{\Psi}_t\right]\boldsymbol{X}_0 \;&=\; d\left[\exp\left\{\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t+\sum_{r=1}^{m}\boldsymbol{D}^r W_t^r\right\}\right]\boldsymbol{X}_0 \quad (3.5)\\
&=\; d\left[\exp\left\{\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t\right\}\prod_{r=1}^{m}\exp\left\{\boldsymbol{D}^r W_t^r\right\}\right]\boldsymbol{X}_0,
\end{aligned}
$$

since the matrices

$$
\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t \quad\text{and}\quad \sum_{r=1}^{m}\boldsymbol{D}^r W_t^r \tag{3.6}
$$

commute for all $t \in [0,\infty)$. Therefore,

$$
\begin{aligned}
d\boldsymbol{X}_t \;=\;& d\left[\exp\left\{\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t\right\}\right]\prod_{r=1}^{m}\exp\left\{\boldsymbol{D}^r W_t^r\right\}\boldsymbol{X}_0 \tag{3.7}\\
&+\exp\left\{\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t\right\}d\left[\prod_{r=1}^{m}\exp\left\{\boldsymbol{D}^r W_t^r\right\}\right]\boldsymbol{X}_0\\
\;=\;& \left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)\exp\left\{\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t\right\}\\
&\times\prod_{r=1}^{m}\exp\left\{\boldsymbol{D}^r W_t^r\right\}\boldsymbol{X}_0 dt\\
&+\sum_{k=1}^{m}\boldsymbol{D}^k\exp\left\{\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t\right\}\prod_{r=1}^{m}\exp\left\{\boldsymbol{D}^r W_t^r\right\}\boldsymbol{X}_0 dW_t^k\\
&+\frac{1}{2}\sum_{k=1}^{m}\left(\boldsymbol{D}^k\right)^2\exp\left\{\left(\boldsymbol{A}-\frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\right)t\right\}\\
&\times\prod_{r=1}^{m}\exp\left\{\boldsymbol{D}^r W_t^r\right\}\boldsymbol{X}_0 d[W^k]_t.
\end{aligned}
$$

This equation simplifies to

$$
d\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_t dt + \sum_{k=1}^{m}\boldsymbol{D}^k\boldsymbol{X}_t dW_t^k, \tag{3.8}
$$

since the quadratic variation of the kth observation process equals

$$
[W^k]_t = \int_0^t ds = t, \tag{3.9}
$$

for $k \in \{1,2,\ldots,m\}$.

The above derivation shows that an SDE of the type (2.10) has an explicit solution if the matrices $\boldsymbol{A}, \boldsymbol{D}^1, \ldots, \boldsymbol{D}^m$ commute. Note that $\boldsymbol{D}^1, \ldots, \boldsymbol{D}^m$ in (2.10) are diagonal matrices, and thus commute with each other. However, the matrix $\boldsymbol{A}$ is not commuting with the other matrices. Therefore, we do not have an exact explicit solution of the Zakai equation (2.10). Nevertheless, as we will illustrate later, if we formally take the matrix exponential (3.4) in the product (3.3), then one obtains a proxy of the solution of the corresponding Zakai equation. It turns out that this quasi-exact solution provides in many cases an excellent approximation of the exact solution, as we will confirm numerically. This is a practically very valuable observation. The solution will be exploited to solve approximately and efficiently the Wonham filter problem. What, of course, needs to be done is to show for given matrices $\boldsymbol{A}, \boldsymbol{D}^1, \ldots, \boldsymbol{D}^m$ and initial vector $\boldsymbol{X}_0$ that the quasi-exact solution is indeed close to the exact solution. This can be achieved by comparing the proposed approximation via discrete time simulation with a very accurately obtained numerical approximation using an extremely small time step size, see Kloeden & Platen (1999).

To prepare this type of comparison, let us now introduce the following equidistant time discretization $0 = \tau_0 < \tau_1 < \cdots < \tau_n = T$, such that $\tau_i = i\Delta$, for $i \in \{0, 1, \ldots, \frac{T}{\Delta}\}$. Denote by $\boldsymbol{Y}_{\tau_i}^{\Delta}$ at time τ_i the quasi-exact approximation of the solution $\boldsymbol{\sigma}(\xi_{\tau_i})$ of the Zakai equation (2.10), expressed by the recursive equation

$$\boldsymbol{Y}_{\tau_{i+1}}^{\Delta} = \exp\left\{\boldsymbol{A}\Delta - \frac{1}{2}\sum_{l=1}^{m}\left(\boldsymbol{D}^l\right)^2\Delta + \sum_{r=1}^{m}\boldsymbol{D}^r\Delta W_{\tau_{i+1}}^r\right\}\boldsymbol{Y}_{\tau_i}^{\Delta}, \tag{3.10}$$

at the equidistant discretization points, where the $\Delta W_{\tau_{i+1}}^r$ are increments of the rth element of the vector observation process $\boldsymbol{W}$. That is we have

$$\Delta W_{\tau_{i+1}}^r = W_{\tau_{i+1}}^r - W_{\tau_i}^r, \tag{3.11}$$

for $r \in \{1, 2, \ldots, d\}$ and $i \in \{0, 1, \ldots, \frac{T}{\Delta}\}$.

We will consider in Section 5 an example of the observation processes given by the following SDE

$$dW_t^r = \xi_t dt + dW_t^{r*}, \tag{3.12}$$

for $t \in [0, T]$. We then add for such a scenario simulation all jump times of the continuous time Markov chain to the equidistant time discretization to obtain a jump adapted time discretization $0 = t_0 < t_1 < \cdots < t_{n_T} = T$ with maximum step size $\Delta > 0$, see Platen (1982). Note that n_T is now a random integer. The increments of this observation process at jump adapted discretization points can be obtained by exact simulation given the value of the hidden Markov chain ξ_{t_i} at time t_i. That is

$$W_{t_{i+1}}^r - W_{t_i}^r = \xi_{t_i}(t_{i+1} - t_i) + \sqrt{t_{i+1} - t_i}Z_{i+1}^r, \tag{3.13}$$

for $r \in \{1, 2, \ldots, d\}$ and $i \in \{0, 1, \ldots, n_T - 1\}$. Here $Z_{i+1}^r \sim \mathcal{N}(0, 1)$ is a standard Gaussian random variable, for $r \in \{1, 2, \ldots, d\}$ and $i \in \{0, 1, \ldots, n_T - 1\}$. Note

7

that this process is a drifted Wiener process with piecewise constant random drift. The simulation of the increments of W^r is straightforward and we can easily obtain the increments, since we have via simulation at our disposal the values of W^r at the times t_i for $i \in \{0, 1, \ldots, n_T\}$.

4　Approximate Filters

In practice, it is impossible to detect $\boldsymbol{W}$ continuously on $[0, T]$. One may, however, approximate increments of observations of $\boldsymbol{W}$ in integral form of the type

$$\int_{\tau_0}^{\tau_1} dW_s^j, \ldots, \int_{\tau_n}^{\tau_{n+1}} dW_s^j, \ldots, \int_{\tau_0}^{\tau_1} \int_{\tau_0}^{s_2} dW_{s_1}^j \, dW_{s_2}^k, \ldots$$

for each $j \in \{1, 2, \ldots, m\}$, $\tau_n = n\Delta$ and $n \in \{0, 1, \ldots, \frac{T}{\Delta}\}$. We shall see later on that with such integral observations it is possible to construct strong discrete-time approximations $\boldsymbol{Y}^\Delta$ with time step size Δ of the solution $\boldsymbol{\sigma}(\xi)$ of the Zakai equation (2.10). This allows then for the given function g to form the *approximate Wonham filter*

$$\hat{g}^\Delta(\xi_t) = \frac{\sum_{k=1}^d g(a_k) Y_t^{\Delta,k}}{\sum_{k=1}^d Y_t^{\Delta,k}} \tag{4.1}$$

for $t \in [0, T]$.

We shall say that a discrete-time approximation $\boldsymbol{Y}^\Delta$ with time step size Δ converges on the time interval $[0, T]$ with strong order $\gamma > 0$ to the solution $\boldsymbol{X}$ of the corresponding SDE if there exists a finite constant K, not depending on Δ, and a $\delta_0 \in (0, 1)$ such that

$$\dot{E}\left(\left|\boldsymbol{\sigma}(\xi_{\tau_n}) - \boldsymbol{Y}_{\tau_n}^\Delta\right|\right) \leq K\,\Delta^\gamma \tag{4.2}$$

for all $\Delta \in (0, \delta_0)$ and $\tau_n \in [0, T]$. Note that the expectation in (4.2) is taken with respect to the probability measure $\dot{P}$ under which the observation process $\boldsymbol{W}$ is a Wiener process.

Analogously, we say that an approximate Markov chain filter $\hat{g}^\Delta(\xi_{\tau_{n_t}})$ with time step size Δ converges on the time interval $[0, T]$ with strong order $\gamma > 0$ to the optimal filter $\hat{g}(\xi_{\tau_{n_t}})$ for a given test function g if there exists a finite constant K, not depending on Δ, and a $\delta_0 \in (0, 1)$ such that

$$E\left(\left|\hat{g}(\xi_{\tau_{n_t}}) - \hat{g}^\Delta(\xi_{\tau_{n_t}})\right|\right) \leq K\,\Delta^\gamma \tag{4.3}$$

for all $\Delta \in (0, \delta_0)$ and $t \in [0, T]$. In contrast with (4.2) the expectation in (4.3) is taken with respect to the original probability measure P. In Kloeden, Platen & Schurz (1993) the following convergence result was derived.

Theorem 4.1 (Kloeden-Platen-Schurz) *An approximate Markov chain filter $\hat{g}^{\Delta}(\xi)$ with time step size Δ converges for $t \in [0,T]$ with strong order $\gamma > 0$ to the optimal filter $\hat{g}(\xi)$ for a given bounded function g if the discrete-time approximation $\boldsymbol{Y}^{\Delta}$ used converges on $[0,T]$ to the solution $\boldsymbol{\sigma}(\xi)$ of the Zakai equation (2.10) with the same strong order γ.*

Now, we derive discrete-time strong approximations $\boldsymbol{Y}^{\Delta}$ that are converging with a given strong order $\gamma > 0$ to the solution $\boldsymbol{\sigma}(\xi)$ of the Zakai equation (2.10), which can be used to build a corresponding approximate filter.

Given an equidistant time discretization of the interval $[0,T]$ with step size $\Delta = \frac{T}{N}$ for some $N \in \{1,2,\ldots\}$, we define the partition sigma-algebra

$$\mathcal{P}_N^1 = \sigma\{\Delta W_{i-1}^j : i \in \{1,2,\ldots,N\}, j \in \{1,2,\ldots,m\}\} \tag{4.4}$$

as the sigma-algebra generated by the increments

$$\Delta W_0^j = \int_0^{\Delta} dW_s^j, \ldots, \Delta W_{N-1}^j = \int_{(N-1)\Delta}^{N\Delta} dW_s^j \tag{4.5}$$

for all $j \in \{1,2,\ldots,m\}$. Thus, $\mathcal{P}_N^1$ contains the information about the increments of $\boldsymbol{W}$ for the given time discretization.

The simplest discrete-time approximation is obtained from the Euler scheme. It has for the Zakai equation (2.10) the form

$$\boldsymbol{Y}_{\tau_{n+1}}^{\Delta} = [\boldsymbol{I} + \boldsymbol{A}\,\Delta + \boldsymbol{G}_n]\,\boldsymbol{Y}_{\tau_n}^{\Delta} \tag{4.6}$$

with

$$\boldsymbol{G}_n = \sum_{k=1}^m \boldsymbol{D}^k \,\Delta\, W_n^k \tag{4.7}$$

and initial value $Y_0 = \sigma(\xi_0)$, where $\boldsymbol{I}$ is the $d \times d$ unit matrix. The scheme (4.6) converges with strong order $\gamma = 0.5$ under the given assumptions. For a general SDE this is the maximum order of strong convergence that can be achieved under the partition sigma-algebra $\mathcal{P}_N^1$, as was shown in Clark & Cameron (1980). However, some commutativity property of the Zakai equation (2.10) follows from the diagonal structure of its volatility matrices, see Kloeden & Platen (1999). This allows the strong order $\gamma = 1.0$ to be attained with the information given by $\mathcal{P}_N^1$.

The Milstein scheme, which is of strong order $\gamma = 1.0$ has for the Zakai equation (2.10) the form

$$\boldsymbol{Y}_{\tau_{n+1}}^{\Delta} = \left(\boldsymbol{I} + \boldsymbol{A}\,\Delta + \boldsymbol{G}_n\left(\boldsymbol{I} + \frac{1}{2}\boldsymbol{D}_n\right)\right)\boldsymbol{Y}_{\tau_n}^{\Delta}, \tag{4.8}$$

where

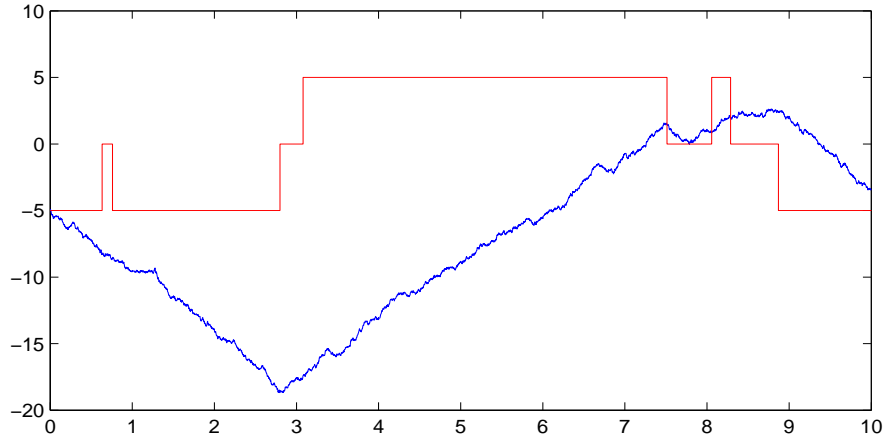$$\boldsymbol{A} = \boldsymbol{A} - \frac{1}{2}\sum_{k=1}^m (\boldsymbol{D}^k)^2. \tag{4.9}$$

9

Figure 5.1: Simulation of the signal and observation processes for $\Delta = \frac{1}{500}$.

Newton (1986) searched for a scheme which is asymptotically the "best" in the class of strong order $\gamma = 1.0$ schemes in the sense that it has the smallest leading error coefficient in an error estimate similar to (4.2), see Kloeden & Platen (1999). In practice, the numerical stability of discrete-time approximations for filters is highly important. Fischer & Platen (1999) studied the application of the balanced implicit method in hidden Markov chain filtering when the time between observations is rather large. They obtained quite reliable results for the balanced implicit scheme even for large step sizes. In the following section we consider an example of a Wonham filter considered also by Fischer & Platen (1999). We show that our quasi-exact approximation of the Zakai equation performs remarkably well, even though the matrix $\boldsymbol{A}$ does not fully commute with the matrices $\boldsymbol{D}^1, \ldots, \boldsymbol{D}^m$ in the Zakai equation.

5 Wonham Filter Example

The unobservable signal process $\xi = \{\xi_t, \, t \in [0, T]\}$ in our example is a time homogeneous, continuous time, real valued Markov chain with the set $\mathcal{X} = \{a_1, a_2, \ldots, a_d\}$ of states. The scalar observation process $W = \{W_t, \, t \in [0, T]\}$ is given by the relation

$$W_t = \int_0^t h(\xi_s)ds + W_t^*, \tag{5.1}$$

for $t \in [0, T]$. One can say that W represents the time integral over the signal process that is corrupted by a Wiener process W^*. We estimate the hidden state of the Markov chain ξ by observing only the values of the process W.

It is convenient to consider a d state continuous time Markov chain $\boldsymbol{X} = \{\boldsymbol{X}_t, \, t \in [0, T]\}$ that is identical to ξ under a transformation of the state space, see Elliott,
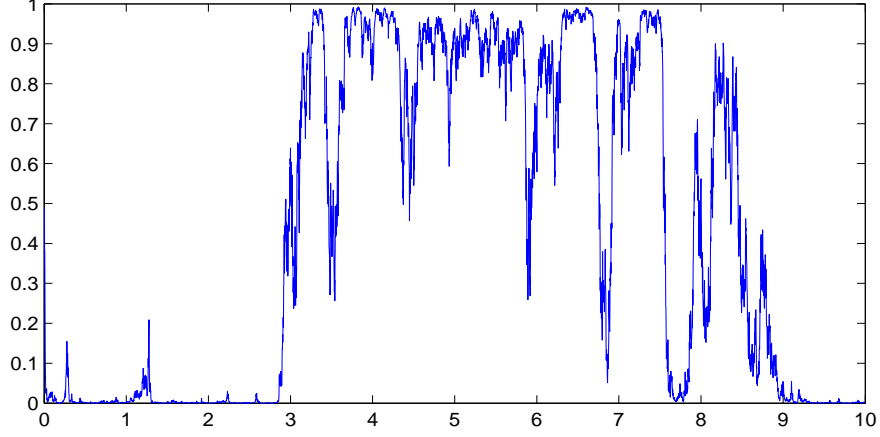
10

Figure 5.2: q_t^1 - obtained by the quasi-exact approximation for $\Delta = \frac{1}{500}$.

Aggoun & Moore (1995). We choose as the state space for $\boldsymbol{X}$ the set $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ of unit vectors in $\Re^d$, with $\boldsymbol{e}_1 = (1, 0, 0, \ldots, 0)^\top$, $\boldsymbol{e}_2 = (0, 1, 0, \ldots, 0)^\top$ and so on. Then we write,

$$\xi_t = \boldsymbol{X}_t^\top \boldsymbol{a} \tag{5.2}$$

with $\boldsymbol{a} = (a_1, a_2, \ldots, a_d)^\top \in \Re^d$. Let $\boldsymbol{A} = [a^{i,j}]_{i,j=1}^d$ be the constant intensity matrix associated with the homogeneous, continuous time Markov chain $\boldsymbol{X}$, so that $\boldsymbol{p}(t) = E(\boldsymbol{X}_t)$ satisfies the vector ordinary differential equation

$$\frac{d\boldsymbol{p}(t)}{dt} = \boldsymbol{A}\,\boldsymbol{p}(t), \tag{5.3}$$

for $t \in [0, T]$, with given initial probability vector $\boldsymbol{p}(0)$. $\boldsymbol{H}$ is the diagonal $d \times d$ matrix that has the elements of the vector $\boldsymbol{a}$ as diagonal elements and is zero elsewhere.

Denote by $\mathcal{Y}_t$ the observation sigma-algebra generated by W up to time t. The Wonham filter for $\boldsymbol{X}$ at time t is then given as

$$\hat{\boldsymbol{X}}_t = E(\boldsymbol{X}_t \,|\, \mathcal{Y}_t).$$

As we have seen in Section 2, the theoretical solution to the problem of calculating $\hat{\boldsymbol{X}}_t$ involves the unnormalized filter for the conditional distribution of $\boldsymbol{X}$, which is denoted by $\sigma(\boldsymbol{X}) = \{\sigma(\boldsymbol{X}_t), t \in [0, T]\}$ and satisfies the Zakai equation

$$d\sigma(\boldsymbol{X}_t) = \boldsymbol{A}\,\sigma(\boldsymbol{X}_t)\,dt + \boldsymbol{D}\,\sigma(\boldsymbol{X}_t)\,dW_t \tag{5.4}$$

for $t \in [0, T]$. The Wonham filter for $\boldsymbol{X}$ is then computed as

$$\hat{\boldsymbol{X}}_t = E(\boldsymbol{X}_t \,|\, \mathcal{Y}_t) = \frac{\sigma(\boldsymbol{X}_t)}{\sigma(\boldsymbol{X}_t)^\top \mathbf{1}} \tag{5.5}$$

11
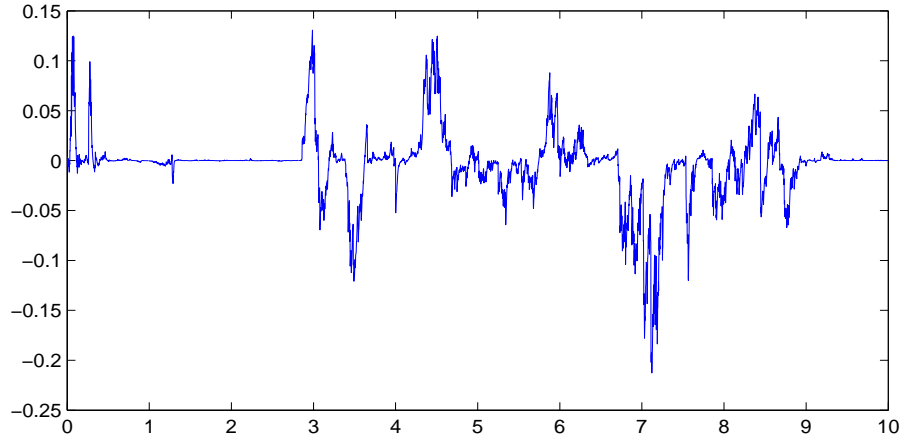
Figure 5.3: Difference between q_t^1 obtained by the quasi-exact approximation and the Euler scheme for $\Delta = \frac{1}{500}$.

for $t \in [0, T]$, see (2.11).

Let us know assume that our observation process W is a one-dimensional stochastic process of the differential form

$$dW_t = \xi_t dt + dW_t^*, \tag{5.6}$$

for $t \in [0, T]$. The exact simulation of this process was already discussed in Section 3.

Moreover, the parameters of the Markov chain X, or equivalently ξ, are chosen to give a realistic multiplicative noise term in the Zakai equation (5.4). The simulated hidden Markov chain X is chosen to have three states, with the vector $\boldsymbol{a}$ taken to be $\boldsymbol{a} = (5, 0, -5)^\top$. Numerical experiments have shown that using more states does not change the nature of the results that we obtain. The intensity matrix $\boldsymbol{A}$ of the hidden Markov chain is chosen to be of the simple form

$$\boldsymbol{A} = \begin{bmatrix} -1.0 & 1.0 & 0 \\ 0.5 & -1.0 & 0.5 \\ 0 & 1.0 & -1.0 \end{bmatrix}. \tag{5.7}$$

This describes how the Markov chain jumps with prescribed intensities to neighboring states. For instance, the intensity to jump from level 0 to level 5 is 0.5 per unit of time.

Now, let us investigate the approximate calculation of the Wonham filter. We simulate the scenario for the signal and observation processes over the time interval $[0, T]$ with $T = 10$ using the exact simulation method (3.13). The simulated output can be seen in Fig. 5.1.

Let us now consider the probability

$$q_t^i = \hat{\boldsymbol{X}}_t^\top \boldsymbol{e}_i = E\left(\boldsymbol{X}_t^\top \boldsymbol{e}_i \,\middle|\, \mathcal{Y}_t\right) \tag{5.8}$$
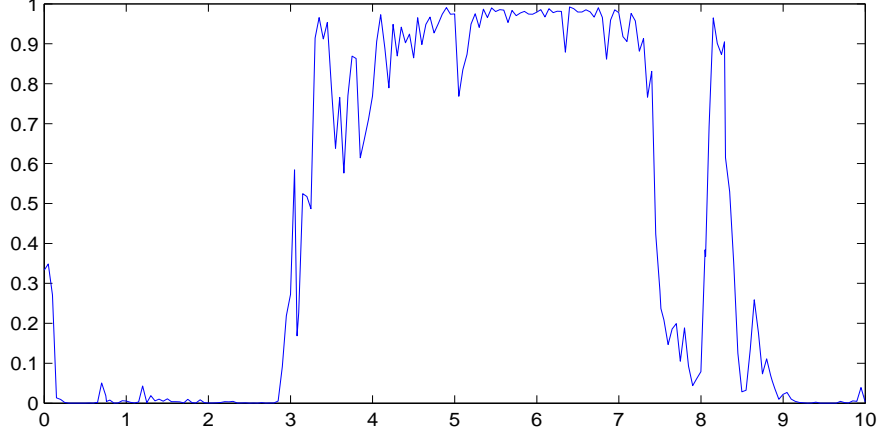
12

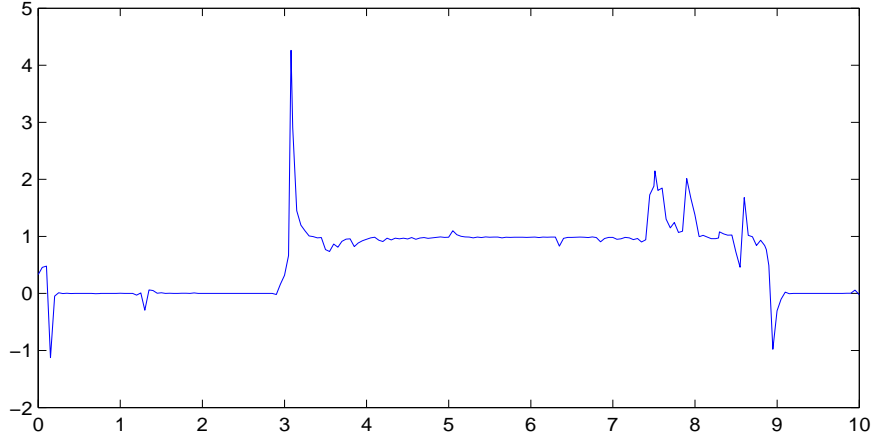Figure 5.4: q_t^1 - obtained by the quasi-exact method with $\Delta = \frac{1}{20}$.

Figure 5.5: q_t^1 - obtained by the Euler method with $\Delta = \frac{1}{20}$.

which, for $i \in \{1, 2, 3\}$, denotes the filtered probability that the hidden Markov chain $\boldsymbol{X}$ is in the ith state. For illustration let us focus on q_t^1, which corresponds to the level $\xi = 5$, that is $i = 1$.

To obtain the quantity q_t^1 we have to solve the SDE (5.4). In order to do so we use our quasi-exact approximation described in Section 3. Applying it to the equation (5.4) and denoting by $\boldsymbol{Y}_{\tau_i}^{\Delta}$ the approximate solution of the Zakai equation at time τ_i, we obtain the following quasi-exact approximation

$$\boldsymbol{Y}_{\tau_{i+1}}^{\Delta} = \exp\left\{\boldsymbol{A}\Delta - \frac{1}{2}\boldsymbol{D}\Delta + \boldsymbol{D}\Delta W_{\tau_{i+1}}\right\}\boldsymbol{Y}_{\tau_i}^{\Delta}, \qquad (5.9)$$

where $\Delta W_{\tau_{i+1}}$ is obtained using (3.13). Additionally, $\boldsymbol{D}$ is a diagonal matrix with

13

elements

$$D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -5 \end{bmatrix}. \tag{5.10}$$

In order to calculate the matrix exponential in (5.9) we use an implementation in Matlab. If a given matrix has a full set of eigenvectors V with corresponding eigenvalues D, then the matrix exponential of this matrix is

$$V \exp\{\text{diag}(D)\}V^{-1}. \tag{5.11}$$

Note also that the matrix exponential of the diagonal matrix is a diagonal matrix, whose diagonal elements are exponents of diagonal elements of the underlying matrix. If this calculation is not possible to apply, Matlab's implementation uses the Padé approximation with scaling and squaring, see Higham (2005).

Additionally, we compare our quasi-exact approximation to the results obtained via two numerical schemes: the Euler scheme and the Milstein scheme, as described in Section 4. When the step size Δ was chosen extremely small with about $\frac{1}{500}$, then all three approximations produced virtually identical results. Fig. 5.2 displays a plot of q_t^1 as calculated by the quasi-exact approximation with $\Delta = \frac{1}{500}$. This is compared to the result obtained by the Euler scheme for the same time step size in Fig. 5.3. In this figure we show the difference between q_t^1 obtained by the quasi-exact approximation and the Euler scheme for $\Delta = \frac{1}{500}$. We remark that small errors occur, most likely, due to the approximate nature of the Euler scheme as well as the Padé approximation of the matrix exponential used in the quasi-exact method. When using the Milstein scheme the results were practically the same. Differences between the schemes become apparent when the time step size Δ is chosen to be larger.

We emphasize that such a fine time discretization for the observation process as employed in our test, however, is often not available in real world filtering problems. Using the same realizations of the observation and signal processes that were given in Fig. 5.1, we display in Figs. 5.4 and 5.5 the plots of the filtered probability q_t^1, see (5.8), calculated by the proposed quasi-exact approximation and the Euler method when using the larger step size $\Delta = \frac{1}{20}$. For this step size we can see that the only acceptable approximation appears to be the proposed quasi-exact approximation of the hidden Markov chain filter. In the other case we even obtain negative "probabilities" and other unrealistic estimates as filter values. This undesirable effect is due to numerical instabilities of the Euler scheme. In the given example also the Milstein scheme does not yield a useful filter.

6   Conclusion

We have proposed a quasi-exact approximation of hidden Markov chain filters, which mimics an exact solution of a multi-dimensional geometric Brownian mo-

tion. It turns out that, even though the drift matrix in the Zakai equation does not perfectly commute, the quasi-exact method is rather successful in approximating hidden Markov chain filters also for large observation time steps.

The results of our experiments demonstrate that discrete time approximations, such as the Euler scheme, cannot be used as reliably as the proposed quasi-exact method. When the available observations are rare the quasi-exact method can provide useful approximations where other standard methods fail.

References

Alcock, J. T. & K. Burrage (2006). A note on the balanced method. *BIT Numerical Mathematics* **46**, 689–710.

Bruti-Liberati, N. & E. Platen (2008). Strong predictor-corrector Euler methods for stochastic differential equations. **8**(3), 561–581.

Clark, J. M. C. & R. J. Cameron (1980). The maximum rate of convergence of discrete approximations for stochastic differential equations. In B. Grigelionis (Ed.), *Stochastic Differential Systems*, Volume 25 of *Lecture Notes in Control and Inform. Sci.*, pp. 162–171. Springer.

Elliott, R. J., L. Aggoun, & J. B. Moore (1995). *Hidden Markov Models: Estimation and Control*, Volume 29 of *Appl. Math.* Springer.

Fischer, P. & E. Platen (1999). Applications of the balanced method to stochastic differential equations in filtering. *Monte Carlo Methods Appl.* **5**(1), 19–38.

Fujisaki, M., G. Kallianpur, & H. Kunita (1972). Stochastic differential equations for the non-linear filtering problem. *Osaka J. Math.* **9**, 19–40.

Hernandez, D. B. & R. Spigler (1993). Convergence and stability of implicit Runge-Kutta methods for systems with multiplicative noise. *BIT* **33**, 654–669.

Higham, D. J. (2000). Mean-square and asymptotic stability of numerical methods for stochastic ordinary differential equations. *SIAM J. Numer. Anal.* **38**, 753–769.

Higham, N. J. (2005). The scaling and squaring method for the matrix exponential revisited. *SIAM J. Matrix Anal. Appl.* **26**(4).

Kallianpur, G. (1980). *Stochastic Filtering Theory.* Springer.

Klauder, J. R. & W. P. Petersen (1985). Numerical integration of multiplicative-noise stochastic differential equations. *SIAM J. Numer. Anal.* **6**, 1153–1166.

Kloeden, P. E. & E. Platen (1992). Higher order implicit strong numerical schemes for stochastic differential equations. *J. Statist. Phys.* **66**(1/2), 283–314.

Kloeden, P. E. & E. Platen (1999). *Numerical Solution of Stochastic Differential Equations*, Volume 23 of *Appl. Math.* Springer. Third corrected printing.

Kloeden, P. E., E. Platen, & H. Schurz (1993). Higher order approximate Markov chain filters. In e. S. Cambanis (Ed.), *Stochastic Processes - A Festschrift in Honour of Gopinath Kallianpur*, pp. 181–190. Springer.

Milstein, G. N. (1988). A theorem of the order of convergence of mean square approximations of systems of stochastic differential equations. *Theory Probab. Appl.* **32**, 738–741.

Milstein, G. N., E. Platen, & H. Schurz (1998). Balanced implicit methods for stiff stochastic systems. *SIAM J. Numer. Anal.* **35**(3), 1010–1019.

Newton, N. J. (1986). An asymptotic efficient difference formula for solving stochastic differential equations. *Stochastics* **19**, 175–206.

Newton, N. J. (1991). Asymptotically efficient Runge-Kutta methods for a class of Ito and Stratonovich equations. *SIAM J. Appl. Math.* **51**, 542–567.

Platen, E. (1982). An approximation method for a class of Itô processes with jump component. *Liet. Mat. Rink.* **22**(2), 124–136.

Platen, E. & L. Shi (2008). On the numerical stability of simulation methods for sdes. Technical report, University of Technology, Sydney. QFRC Research Paper 234.

Saito, Y. & T. Mitsui (1993). T-stability of numerical schemes for stochastic differential equations. *World Sci. Ser. Appl. Anal.* **2**, 333–344.

Talay, D. (1982). Convergence for each trajectory of an approximation scheme of SDE. *Computes Rendus Acad. Sc. Paris, Séries I Math.* **295**(3), 249–252. (in French).

Wonham, W. M. (1965). Some applications of stochastic differential equations to optimal nonlinear filtering. *SIAM J. Control Optim.* **2**, 347–369.

Zakai, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrsch. Verw. Gebiete* **11**, 230–243.