

Nonlinear filtering for stochastic systems with fixed delay: approximation by a modified Milstein scheme*

Antonella Calzolari [†] Patrick Florchinger [‡] Giovanna Nappo [§]

Abstract

In this paper we study approximation schemes for a nonlinear filtering problem of a partially observed diffusive system when the state process \mathbf{X} is the solution of a stochastic delay diffusion equation with a constant time lag τ and the observation process is a noisy function of the state. The approximating state is the linear interpolation of a modified Milstein scheme, which is asymptotically optimal with respect to the mean square \mathcal{L}^2 -error within the class of all pathwise approximations based on equidistant observations of the driving Brownian motion. Upper bounds for the error of the filter approximations are computed. Some other discretization schemes for the state process are also considered.

Introduction

On the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, P)$, let $\mathbf{X} = (X(t))_{t \in [-\tau, T]}$ be the state process governed by the following stochastic delay differential equation with constant time lag

$$\begin{cases} X(t) = \eta(t), & -\tau \leq t \leq 0, \\ X(t) = \eta(0) + \int_0^t a(u, X(u), X(u - \tau)) du + \int_0^t b(u, X(u), X(u - \tau)) d\tilde{W}_u, & 0 \leq t \leq T, \end{cases} \quad (1)$$

where τ is a positive constant, $\boldsymbol{\eta} = (\eta(s))_{s \in [-\tau, 0]}$ is the initial path and $\tilde{\mathbf{W}} = (\tilde{W}(t))_{t \in [0, T]}$ is a standard Brownian motion.

Let $\mathbf{Y} = (Y(t))_{t \in [0, T]}$ be the observation process given by

$$Y(t) = \int_0^t h(u, X(u), X(u - \tau)) du + W(t), \quad 0 \leq t \leq T, \quad (2)$$

where $\mathbf{W} = (W(t))_{t \in [0, T]}$ is a standard Brownian motion, independent of $\tilde{\mathbf{W}}$, and $h : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is a Borel measurable function.

*Partially supported by "Sapienza" Università di Roma Anno: 2006 - prot. C26A06KWKF "Strutture di dipendenza in modelli stocastici e applicazioni " and PRIN project Anno 2008 - prot. 2008YYYBE4_004 "Modellizzazione e metodi numerici in finanza"

[†]Dipartimento di Matematica - Università di Roma "Tor Vergata", via della Ricerca Scientifica 1, I 00133 Roma, Italy

[‡]Département de Mathématiques, Université de Metz, 23 Allée des Oeillets F 57160 Moulins les Metz, France

[§]Dipartimento di Matematica - "Sapienza" Università di Roma, piazzale A. Moro 2, I 00185 Roma, Italy

Partially observed stochastic delay systems appear in many applications, for instance in population growth (see, e.g., [8]) and in mathematical finance (see, e.g., [1]).

We assume the following hypotheses on the initial path η , the drift coefficient a and the diffusion coefficient b of equation (1) stated above:

(A1) η is a \mathcal{F}_0 -measurable random variable with values in $C([-\tau, 0], \mathbb{R})$ and such that

$$E\left(\sup_{s \in [-\tau, 0]} |\eta(s)|^2\right) < \infty;$$

(A2) the drift $a(t, x_1, x_2)$ and the diffusion coefficient $b(t, x_1, x_2)$ satisfy

$$a, b \in C^{1,2}([0, T] \times \mathbb{R}^2)$$

with bounded spatial derivatives $a^{(0,1,0)}$, $a^{(0,0,1)}$, $b^{(0,1,0)}$ and $b^{(0,0,1)}$.

Properties (A1) and (A2) implies that a unique strong solution of equation (1) with initial condition $X(s) = \eta(s)$, for $-\tau \leq s \leq 0$, exists and satisfies

$$E\left(\sup_{s \in [-\tau, T]} |X(s)|^2\right) < \infty. \quad (3)$$

Moreover we assume that

(B1) $h : [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}$ is jointly continuous and with less than linear growth, i.e. for any $t \in [0, T]$ and $x = (x_1, x_2)$ in \mathbb{R}^2

$$|h(t, x_1, x_2)|^2 \leq K(1 + x_1^2 + x_2^2);$$

(B2) h is globally Lipschitz with respect to (x_1, x_2) , uniformly in time, i.e. for any $t \in [0, T]$, $x = (x_1, x_2)$ and $x' = (x'_1, x'_2)$ in \mathbb{R}^2

$$|h(t, x_1, x_2) - h(t, x'_1, x'_2)| \leq L_h |(x_1, x_2) - (x'_1, x'_2)|.$$

In this paper we present an approximation scheme for the conditional law π_t of the state process at time t , given the observation process up to time t , i.e., an approximation scheme for the so-called filter

$$\pi_t(\varphi) = E[\varphi(X(t))/\mathcal{F}_t^Y], \quad 0 \leq t \leq T \quad (4)$$

for all functions φ belonging to a determining class, i.e., the best estimate of $\varphi(X(t))$ given the σ -algebra of the observations up to time t , $\mathcal{F}_t^Y = \sigma\{Y(s), s \leq t\}$.

Approximation in nonlinear filtering of partially observed stochastic differential systems without delays has been widely studied by many authors, see e.g. Kushner [13], Le Gland [15], Crisan [6], Del Moral [7] and the references therein. The case of stochastic delay systems is much less investigated. In [5], Chang gives a computable and weakly convergent approximation for the filter associated to partially observed delay systems by applying an Euler discretization scheme to the state process. Other weak approximation schemes, such as those developed by Kushner [14] in a stochastic control framework, could also be applied to obtain weak convergence of the filters. Nevertheless, it seems difficult to compute the rate of convergence with these methods. In [4], Calzolari, Florchinger and Nappo construct a strong approximation scheme which depends on the actual observation process and converges in probability to

the original filter as measure valued process. As in [5], the starting point is the Euler approximation of the state process. Moreover in [4], a nonlinear filtering problem concerning functionals of the trajectory $(X(t+s), s \in [-\tau, 0])$ and a more general model of stochastic delay differential equations is considered and, under appropriate conditions on the initial path, the drift coefficients and the diffusion coefficient, the order of convergence at any fixed time turns out to be $(\log n/n)^{1/2}$, where n is the number of points in the discretization grid.

The aim of this paper is to get a new approximation scheme combining the techniques presented by Calzolari, Florchinger and Nappo in [3] (see also Proposition 4.1 of [4], and Theorem 1.1 in the sequel) with the modified Milstein scheme for pathwise approximation of stochastic delay differential equations with constant time lag introduced by Hofmann and Mueller-Gronbach in [9]. The choice of the modified Milstein approximation is motivated by the fact that this approximation is asymptotically optimal with respect to the mean square \mathcal{L}^2 -error within the class of all pathwise approximations based on the knowledge of $\tilde{W}(t_\ell)$ for a finite number of equidistant times $t_\ell \in [0, T]$ (see Theorem 2 in [9]). For the ease of the reader we have summarized the results of [9] in Proposition 2.3.

The main features of the method we propose are the following. The approximation of the filter is given by a deterministic functional which depends on an approximation $(\hat{\mathbf{X}}^n, \mathbf{Y}^n)$ of the original system (\mathbf{X}, \mathbf{Y}) evaluated in the actual observation process \mathbf{Y} (see (7)). In this paper two different approximations \mathbf{Y}^n of the observation process \mathbf{Y} are considered (see (8) and (18)). Though the first one is the most natural, the corresponding filter approximation is not feasible, since in order to be computed, the whole trajectory of the observation process \mathbf{Y} has to be known. On the other hand the second approximation leads to a feasible filter approximation, i.e. depending on $Y(t)$ and $Y(t_\ell)$ for a finite number of times $t_\ell \in [0, t]$. With the modified Milstein approximation for the state process (see (24)), the approximations of the filter converge in probability in bounded Lipschitz metric to the original filter, the rate of convergence being of order $n^{-1/2}$. Moreover, the error can be bounded from above by an explicit constant, and the error bound for the filter is asymptotically the best that can be achieved by combining the general upper bounds of [4] and the results in [9], when restricting to state process approximations depending on $\tilde{W}(t_\ell)$ for a finite number of equidistant times $t_\ell \in [0, T]$.

Note that, as discussed in the last section, by using different continuous approximation schemes for the state process, such as the Euler-Maruyama or the Milstein approximations, the technique exposed in this paper would also allow us to compute explicit upper bounds for the error, even with a faster (or asymptotically faster) rate of convergence, as for the Milstein approximation. Nevertheless from a practical point of view, continuous approximation methods are not feasible in the sense that the whole trajectory of the process $\tilde{\mathbf{W}}$ has to be known. Furthermore, as discussed in Section 4 (see also Remark 1.4), with \mathbf{Y}^n given by (18), the gain in the rate of convergence obtained by using the Milstein approximation is lost and the rate remains equal to $n^{-1/2}$. Finally, though from a theoretical point of view, the asymptotical upper bound that we can get with our method by using the Milstein approximation is better than that obtained by the modified Milstein approximation, it turns out that, from a practical point of view, the best asymptotical upper bound is obtained by the latter approximation scheme.

This paper is divided into four sections and is organized as follows. In Section 1 we recall a general method to define an approximation of the filter and to compute an upper bound for the bounded Lipschitz distance between the original filter and its approximation for two different approximations of the observation process. In Section 2 we describe the modified Milstein approximation for the state process given in [9], and recall the asymptotic estimates for the rate of convergence. In Section 3, we construct the approximation schemes for the filter and prove the main convergence results (Theorem 3.2

and Theorem 3.4). Finally in Section 4, we briefly present some other approximation schemes for the filter that can be derived by the same techniques using various continuous approximation schemes for the state process, and discuss about the rate of convergence for the filter approximations, both from a theoretical and a practical point of view.

1 The approximation of the filter

In this section we assume that $\hat{\mathbf{X}}^n = (\hat{X}^n(t))_{t \in [-\tau, T]}$ is a sequence of processes, defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, P)$, converging to \mathbf{X} , independent of \mathbf{W} , such that

$$\sup_{t \in [-\tau, T]} E[|\hat{X}^n(t)|^2] < \infty, \quad (5)$$

and we construct a partially observed stochastic system $(\hat{\mathbf{X}}^n, \mathbf{Y}^n)$, with the property that \mathbf{Y}^n converges to \mathbf{Y} , and such that the associated filter $\hat{\pi}^n$ has a robust version. This means that

$$\hat{\pi}_t^n(dx) = \hat{U}^n(t, \mathbf{Y}^n; dx) \quad (6)$$

for a suitable deterministic measure-valued functionals \hat{U}^n , with paths in the Skorohod space $D_{\mathcal{P}(\mathbb{R})}([0, T])$, where $\mathcal{P}(\mathbb{R})$ denotes the space of probability measures on \mathbb{R} , and such that $\hat{U}^n(t, \mathbf{y}) = \hat{U}^n(t, \mathbf{y}(\cdot \wedge t))$. While the existence of such a functional is straightforward (see, e.g. [12]), the explicit form of \hat{U}^n may be difficult to obtain. Then we consider

$$\tilde{\pi}_t^n = \hat{U}^n(t, \mathbf{Y}) \quad (7)$$

as an approximation of the original filter π_t , depending on the trajectory actually observed. Though different approximations \mathbf{Y}^n can be considered, as a natural choice we define

$$\mathbf{Y}^n(t) = \int_0^t h(s, \hat{X}^n(s), \hat{X}^n(s - \tau)) ds + W(t), \quad 0 \leq t \leq T. \quad (8)$$

Then, since the observation noise \mathbf{W} is independent of both \mathbf{X} and $\hat{\mathbf{X}}^n$, thanks to hypothesis **(B1)**, and estimates (3) and (5) we can compute the filters π_t and $\hat{\pi}_t^n$ associated respectively with the original and the approximating delay systems (\mathbf{X}, \mathbf{Y}) and $(\hat{\mathbf{X}}^n, \mathbf{Y}^n)$ by the classical Kallianpur-Striebel formula (see [2]). The filter π_t is then given, for any bounded measurable function φ , by

$$\pi_t(\varphi) = \frac{\sigma_t(\varphi)}{\sigma_t(\mathbf{1})} = \frac{E^0 [\varphi(X(t)) \mathcal{L}_t | \mathcal{F}_t^Y]}{E^0 [\mathcal{L}_t | \mathcal{F}_t^Y]}, \quad (9)$$

where E^0 denotes the expectation with respect to the reference probability measure P^0 , defined by the Radon-Nikodym derivative

$$\frac{dP^0}{dP} = (\mathcal{L}_T)^{-1},$$

with

$$\mathcal{L}_t = \exp \left\{ \int_0^t h(s, X(s), X(s - \tau)) dY(s) - \frac{1}{2} \int_0^t |h(s, X(s), X(s - \tau))|^2 ds \right\}, \quad (10)$$

which is a martingale thanks to the sublinearity of h and the bound (3). Similarly

$$\hat{\pi}_t^n(\varphi) = \frac{\sigma_t^n(\varphi)}{\sigma_t^n(\mathbf{1})} = \frac{E^{0,n} \left[\varphi(\hat{X}^n(t)) \hat{\mathcal{L}}_t^n | \mathcal{F}_t^{Y^n} \right]}{E^{0,n} \left[\hat{\mathcal{L}}_t^n | \mathcal{F}_t^{Y^n} \right]}, \quad (11)$$

where $E^{0,n}$ denotes the expectation with respect to the reference probability measure $P^{0,n}$, defined by the Radon-Nikodym derivative

$$\frac{dP^{0,n}}{dP} = (\hat{\mathcal{L}}_T^n)^{-1},$$

with

$$\hat{\mathcal{L}}_t^n = \exp \left\{ \int_0^t h(s, \hat{X}^n(s), \hat{X}^n(s-\tau)) dY^n(s) - \frac{1}{2} \int_0^t |h(s, \hat{X}^n(s), \hat{X}^n(s-\tau))|^2 ds \right\}, \quad (12)$$

which is a martingale thanks to the sublinearity of h and the bound (5).

Furthermore note that the independence of $\hat{\mathbf{X}}^n$ and \mathbf{W} under P implies the independence of $\hat{\mathbf{X}}^n$ and \mathbf{Y}^n under the reference probability measure $P^{0,n}$, and in addition the law of $\hat{\mathbf{X}}^n$ is the same under the probability measures P and $P^{0,n}$.

Before stating our convergence result, we recall the definition of the bounded Lipschitz metric. For any metric space S , and probability measures ν_1 and ν_2 on S ,

$$d_{BL}(\nu_1, \nu_2) = \sup \left\{ \frac{|\nu_1(\varphi) - \nu_2(\varphi)|}{\|\varphi\| \vee L_\varphi}; \varphi \text{ bounded and Lipschitz} \right\}$$

where $\|\varphi\|$ denotes the sup-norm, and L_φ is the Lipschitz constant of φ .

Theorem 1.1. *Assume that conditions (A1), (A2), and (B1), (B2) for the system (1)-(2) are satisfied. Assume that the processes $\hat{\mathbf{X}}^n$ are independent of \mathbf{W} , and satisfy condition (5), and that the approximation processes \mathbf{Y}^n are given by (8). Then there exists a computable positive constant \mathcal{H} , independent of t and n , such that*

$$E[d_{BL}(\pi_t, \tilde{\pi}_t^n)] \leq \mathcal{H} \left[E \left[\int_{-\tau}^t |\hat{X}^n(s) - X(s)|^2 ds \right] + \left(E \left[\int_{-\tau}^t |\hat{X}^n(s) - X(s)|^2 ds \right] \right)^{1/2} \right] + (E[|\hat{X}^n(t) - X(t)|^2])^{1/2}. \quad (13)$$

Moreover, if the right hand side of the above inequality converges to zero, then the sequence $\tilde{\pi}_t^n$ defined in (7) converges in probability in bounded Lipschitz metric to the original filter π_t .

Proof. Let P^n be the probability measure defined by the Radon-Nikodym derivative

$$\frac{dP^n}{dP^0} = \exp \left\{ \int_0^T h(s, \hat{X}^n(s), \hat{X}^n(s-\tau)) dY(s) - \frac{1}{2} \int_0^T |h(s, \hat{X}^n(s), \hat{X}^n(s-\tau))|^2 ds \right\}, \quad (14)$$

where P^0 is the reference probability measure defined above.

Then the law of $(\hat{\mathbf{X}}^n, \mathbf{Y})$ under P^n is the same as the law of $(\hat{\mathbf{X}}^n, \mathbf{Y}^n)$ under P , and $(\mathbf{X}, \hat{\mathbf{X}}^n, \mathbf{Y})$, and the probability measures P^0 , P , and P^n satisfy conditions (a), (a_n), (b1), and (b2) of Calzolari, Florchinger, and Nappo [3]. Therefore, using (32) in Theorem 2.3 of [3], we have

$$E[d_{BL}(\pi_t, \tilde{\pi}_t^n)] \leq 2E^0\left[|(dP^n/dP^0)|_{\tilde{\mathcal{F}}_t} - (dP/dP^0)|_{\tilde{\mathcal{F}}_t}|\right] + E[|X(t) - \hat{X}^n(t)|], \quad (15)$$

where $\tilde{\mathcal{F}}_t = \mathcal{F}_t^{X, \hat{X}^n, Y}$. Moreover, with slight modifications in the proof of Proposition 4.1 of [4], we get for all $t \leq T$

$$\begin{aligned} E^0\left[|(dP^n/dP^0)|_{\tilde{\mathcal{F}}_t} - (dP/dP^0)|_{\tilde{\mathcal{F}}_t}|\right] \\ \leq 2\left(E\left[\int_0^t |h(s, \hat{X}^n(s), \hat{X}^n(s-\tau)) - h(s, X(s), X(s-\tau))|^2 ds\right]\right)^{1/2} \\ + E\left[\int_0^t |h(s, \hat{X}^n(s), \hat{X}^n(s-\tau)) - h(s, X(s), X(s-\tau))|^2 ds\right]. \end{aligned}$$

Finally, substituting the previous bound in (15), using the Lipschitz assumption on h and Cauchy-Schwarz inequality

$$\begin{aligned} E[d_{BL}(\pi_t, \tilde{\pi}_t^n)] &\leq 4L_h \left(E\left[\int_{-\tau}^t |\hat{X}^n(s) - X(s)|^2 ds\right]\right)^{1/2} \\ &\quad + 2L_h^2 E\left[\int_{-\tau}^t |\hat{X}^n(s) - X(s)|^2 ds\right] + (E[|\hat{X}^n(t) - X(t)|^2])^{1/2}, \end{aligned} \quad (16)$$

and \mathcal{H} can be taken to be equal to $\max(4L_h, 2L_h^2)$. \square

Remark 1.2. *The same arguments apply when the function h appearing in the definition of \mathbf{Y} is of the form*

$$h(u, X(u), X(u-\tau_1), \dots, X(u-\tau_m)),$$

where the delays τ_i , $i = 1, \dots, m$, are in $[0, \tau]$, for a fixed $\tau > 0$, the only difference being that the constant \mathcal{H} depends linearly on m . Furthermore, the result extends when the delays τ_i depend on time.

On the other hand, when instead h is of the form

$$h(u, X(u)),$$

then inequality (16) reduces to

$$\begin{aligned} E[d_{BL}(\pi_t, \tilde{\pi}_t^n)] &\leq 2L_h \left(E\left[\int_0^t |\hat{X}^n(s) - X(s)|^2 ds\right]\right)^{1/2} \\ &\quad + L_h^2 E\left[\int_0^t |\hat{X}^n(s) - X(s)|^2 ds\right] + (E[|\hat{X}^n(t) - X(t)|^2])^{1/2}. \end{aligned} \quad (17)$$

Among all the possible approximations for the observation process that one can consider instead of (8) the following one

$$Y^n(t) = \int_0^t h(\delta \lfloor s/\delta \rfloor, \hat{X}^n(\delta \lfloor s/\delta \rfloor), \hat{X}^n(\delta \lfloor (s-\tau)/\delta \rfloor)) ds + W(t), \quad 0 \leq t \leq T, \quad (18)$$

has the property, as shown in the sequel, that the functional $\hat{U}^n(t, y(\cdot \wedge t))$ depends on $y(t)$ and on $y(t_\ell)$ for a finite number of times $t_\ell \in [0, t]$, so that the approximation of the filter given by (7) is feasible, i.e. it depends only on $Y(t)$ and on $Y(t_\ell)$ for a finite number of times $t_\ell \in [0, t]$ (see (22)).

Moreover, if, for the sake of simplicity, we assume that h does not depend on time (otherwise we need to assume some regularity of h with respect to time) then the following result holds.

Theorem 1.3. *Assume that conditions (A1),(A2), and (B1),(B2) for the system (1)-(2) are satisfied. Assume that the processes \hat{X}^n are independent of \mathbf{W} , and satisfy condition (5), and that the approximation processes \mathbf{Y}^n are given by (18) with $h(t, x) = h(x)$. Then there exists a computable positive constant \mathcal{H} , independent of t and n , such that*

$$\begin{aligned} E[d_{BL}(\pi_t, \tilde{\pi}_t^n)] &\leq \mathcal{H} \left[\left(E \left[\int_{-\tau}^t |\hat{X}^n(\delta \lfloor s/\delta \rfloor) - X(\delta \lfloor s/\delta \rfloor)|^2 ds \right] \right)^{1/2} \right. \\ &\quad + E \left[\int_{-\tau}^t |\hat{X}^n(\delta \lfloor s/\delta \rfloor) - X(\delta \lfloor s/\delta \rfloor)|^2 ds \right] \\ &\quad + \left(E \left[\int_{-\tau}^t |X(\delta \lfloor s/\delta \rfloor) - X(s)|^2 ds \right] \right)^{1/2} + E \left[\int_{-\tau}^t |X(\delta \lfloor s/\delta \rfloor) - X(s)|^2 ds \right] \\ &\quad \left. + (2E[|\hat{X}^n(\delta \lfloor t/\delta \rfloor) - X(\delta \lfloor t/\delta \rfloor)|^2])^{1/2} + (2E[|X(\delta \lfloor t/\delta \rfloor) - X(t)|^2])^{1/2} \right]. \quad (19) \end{aligned}$$

Moreover, if the right hand side of the above inequality converges to zero, then the sequence $\tilde{\pi}_t^n$ defined in (7) converges in probability in bounded Lipschitz metric to the original filter π_t .

Proof. Arguing as in the proof of Theorem 1.1, instead of inequality (16) we get

$$\begin{aligned} E[d_{BL}(\pi_t, \tilde{\pi}_t^n)] &\leq 4L_h \left(E \left[\int_{-\tau}^t |\hat{X}^n(\delta \lfloor s/\delta \rfloor) - X(s)|^2 ds \right] \right)^{1/2} \\ &\quad + 2L_h^2 E \left[\int_{-\tau}^t |\hat{X}^n(\delta \lfloor s/\delta \rfloor) - X(s)|^2 ds \right] + (E[|\hat{X}^n(\delta \lfloor t/\delta \rfloor) - X(t)|^2])^{1/2}. \end{aligned}$$

Then, the previous inequality leads to the following upper bound

$$\begin{aligned} E[d_{BL}(\pi_t, \tilde{\pi}_t^n)] &\leq 4L_h \left(2E \left[\int_{-\tau}^t |\hat{X}^n(\delta \lfloor s/\delta \rfloor) - X(\delta \lfloor s/\delta \rfloor)|^2 ds \right] \right)^{1/2} \\ &\quad + 4L_h \left(2E \left[\int_{-\tau}^t |X(\delta \lfloor s/\delta \rfloor) - X(s)|^2 ds \right] \right)^{1/2} \\ &\quad + 4L_h^2 E \left[\int_{-\tau}^t |\hat{X}^n(\delta \lfloor s/\delta \rfloor) - X(\delta \lfloor s/\delta \rfloor)|^2 ds \right] + 4L_h^2 E \left[\int_{-\tau}^t |X(\delta \lfloor s/\delta \rfloor) - X(s)|^2 ds \right] \\ &\quad + (2E[|\hat{X}^n(\delta \lfloor t/\delta \rfloor) - X(\delta \lfloor t/\delta \rfloor)|^2])^{1/2} + (2E[|X(\delta \lfloor t/\delta \rfloor) - X(t)|^2])^{1/2}, \quad (20) \end{aligned}$$

and \mathcal{H} can be taken to be equal to $\max(4\sqrt{2}L_h, 4L_h^2)$. \square

Remark 1.4. *From the previous result, we observe that, for the above feasible approximation filter, we cannot obtain a rate of convergence faster than $n^{-1/2}$. Indeed, even when the initial path is a deterministic constant function, and the coefficients a and b are uniformly bounded above, then $E[|X(t) - X(s)|^2] = O(|t - s|)$.*

With the choice (18) as an approximation of the observation process the reference probability measure $P^{0,n}$ is defined by the Radon-Nikodym derivative

$$\frac{dP^{0,n}}{dP} = (\hat{\mathcal{L}}_T^n)^{-1},$$

with

$$\begin{aligned} \hat{\mathcal{L}}_t^n = \exp \left\{ \int_0^t h(\delta \lfloor s/\delta \rfloor, \hat{X}^n(\delta \lfloor s/\delta \rfloor), \hat{X}^n(\delta \lfloor (s-\tau)/\delta \rfloor)) dY_s^n \right. \\ \left. - \frac{1}{2} \int_0^t |h(\delta \lfloor s/\delta \rfloor, \hat{X}^n(\delta \lfloor s/\delta \rfloor), \hat{X}^n(\delta \lfloor (s-\tau)/\delta \rfloor))|^2 ds \right\}, \end{aligned} \quad (21)$$

which is a martingale thanks to the sublinearity of h and the upper bound (5).

The advantage of this choice for \mathbf{Y}^n is that the Radon-Nikodym derivative can be easily computed since the stochastic integrals reduce to sums involving $Y^n(t)$ and $Y^n(t_\ell)$ for a finite number of times $t_\ell \in [0, t]$. Indeed, taking into account that $s \mapsto h(\delta \lfloor s/\delta \rfloor, \hat{X}^n(\delta \lfloor s/\delta \rfloor), \hat{X}^n(\delta \lfloor (s-\tau)/\delta \rfloor))$ is piecewise constant, we have that (we assume that $\tau = m\delta$ and $\delta = T/n$)

$$\hat{\mathcal{L}}_t^n = L_t^n(\hat{X}^n(\cdot), Y_0^n, Y_\delta^n, \dots, Y_{\lfloor t/\delta \rfloor}^n, Y_t^n),$$

where, for $0 \leq \ell \leq n$,

$$\begin{aligned} \log L_{\ell\delta}^n(x(\cdot), y_0, y_1, \dots, y_\ell) &= \sum_{k=0}^{\ell-1} h(k\delta, x(k\delta), x((k-m)\delta)) (y_{k+1} - y_k) \\ &\quad - \frac{1}{2} \sum_{k=0}^{\ell-1} |h(k\delta, x(k\delta), x((k-m)\delta))|^2 \delta, \end{aligned}$$

and, for $t \in (\ell\delta, (\ell+1)\delta)$, $0 \leq \ell \leq n-1$,

$$\begin{aligned} \log L_t^n(x(\cdot), y_0, y_1, \dots, y_\ell, y) &= \log L_{\ell\delta}^n(x(\cdot), y_0, y_1, \dots, y_\ell) \\ &\quad + h(\delta \lfloor t/\delta \rfloor, x(\delta \lfloor t/\delta \rfloor), x(\delta \lfloor (t-\tau)/\delta \rfloor)) (y - y_\ell) \\ &\quad - \frac{1}{2} |h(\delta \lfloor t/\delta \rfloor, x(\delta \lfloor t/\delta \rfloor), x(\delta \lfloor (t-\tau)/\delta \rfloor))|^2 (t - \ell\delta). \end{aligned}$$

Moreover, under $P^{0,n}$, the processes \hat{X}^n and \mathbf{Y}^n are independent and the law of the approximated state process is invariant under P and $P^{0,n}$, and hence, for $t \in [\ell\delta, (\ell+1)\delta)$, $0 \leq \ell \leq n-1$,

$$\tilde{\sigma}_t^n(\varphi) = E \left[\varphi(\hat{X}^n(t)) L_t^n(\hat{X}^n(\cdot), y_0, y_1, \dots, y_\ell, y) \right] \Bigg|_{y_0=Y_0^n, y_1=Y_\delta^n, \dots, y_\ell=Y_{\ell\delta}^n, y=Y_t^n}.$$

Therefore, when \mathbf{Y}^n is given by (18), the functional $\hat{U}^n(t, y(\cdot \wedge t))$ depends on $y(t)$ and on $y(\ell\delta)$ for $\ell \leq \lfloor t/\delta \rfloor$, and the approximation of the filter is feasible. Indeed the filter approximation can be computed as

$$\tilde{\pi}_t^n(\varphi) = \frac{E[\varphi(\hat{X}^n(t)) L_t^n(\hat{X}^n(\cdot), y_0, y_1, \dots, y_\ell, y)]}{E[L_t^n(\hat{X}^n(\cdot), y_0, y_1, \dots, y_\ell, y)]} \Bigg|_{y_0=Y_0, y_1=Y_\delta, \dots, y_\ell=Y_{\ell\delta}, y=Y_t}. \quad (22)$$

Remark 1.5. When the initial condition is not approximated, i.e. when $\hat{X}^n(s) = \eta(s)$ for all $s \in [-\tau, 0]$, then in all the previous upper bounds (13), (16), (19) and (20) for $E[d_{BL}(\pi_t, \tilde{\pi}_t^n)]$ the integrals between $-\tau$ and t reduce to integrals between 0 and t .

2 The Milstein type approximation of the state

In this section we are dealing with the Milstein type approximation $\hat{\mathbf{X}}^n = (\hat{X}^n(t))_{t \in [-\tau, T]}$, introduced in [9], for the state process $\mathbf{X} = (X(t))_{t \in [-\tau, T]}$.

In addition to the previous hypotheses we assume that the following conditions on the initial path $\boldsymbol{\eta}$, the drift coefficient a and the diffusion coefficient b of equation (1) are satisfied:

(A3) there exists a constant c , such that, for all $s, t \in [-\tau, 0]$,

$$E[|\eta(t) - \eta(s)|^2] \leq c|t - s|;$$

(A4) there exists $K > 0$ such that

$$|a^{(1,0,0)}(t, x_1, x_2)| + |b^{(1,0,0)}(t, x_1, x_2)| \leq K(1 + |x_1| + |x_2|),$$

where $a^{(1,0,0)}$ and $b^{(1,0,0)}$ denote the partial derivatives with respect to time;

(A5) equation (1) is non-deterministic, i.e.

$$\int_0^T E[b^2(t, X(t), X(t - \tau))] dt > 0.$$

Properties (A3) and (A4), together with (A1) and (A2), imply that the unique strong solution of equation (1) satisfies

$$E[|X(t) - X(s)|^2] \leq C|t - s| \quad (23)$$

with the constant C dependent on $\boldsymbol{\eta}$, a and b .

Let us consider the Milstein type discretization scheme with step $\delta = \delta_n = T/n$, with $\tau = m\delta$ (for the sake of simplicity, T/τ is rational) :

$$\left\{ \begin{array}{ll} \hat{X}^n(\ell\delta) = \eta(\ell\delta), & -m \leq \ell \leq 0, \\ \hat{X}^n((\ell + 1)\delta) = \hat{X}^n(\ell\delta) + a(\ell\delta, \hat{X}^n(\ell\delta), \hat{X}^n((\ell - m)\delta))\delta \\ \quad + b(\ell\delta, \hat{X}^n(\ell\delta), \hat{X}^n((\ell - m)\delta))[\tilde{W}((\ell + 1)\delta) - \tilde{W}(\ell\delta)] \\ \quad + \frac{1}{2}(bb^{(0,1,0)})(\ell\delta, \hat{X}^n(\ell\delta), \hat{X}^n((\ell - m)\delta))[(\tilde{W}((\ell + 1)\delta) - \tilde{W}(\ell\delta))^2 - \delta] & 0 \leq \ell \leq m - 1, \\ \hat{X}^n((\ell + 1)\delta) = \hat{X}^n(\ell\delta) + a(\ell\delta, \hat{X}^n(\ell\delta), \hat{X}^n((\ell - m)\delta))\delta \\ \quad + b(\ell\delta, \hat{X}^n(\ell\delta), \hat{X}^n((\ell - m)\delta))[\tilde{W}((\ell + 1)\delta) - \tilde{W}(\ell\delta)] \\ \quad + \frac{1}{2}(bb^{(0,1,0)})(\ell\delta, \hat{X}^n(\ell\delta), \hat{X}^n((\ell - m)\delta))[(\tilde{W}((\ell + 1)\delta) - \tilde{W}(\ell\delta))^2 - \delta] \\ \quad + \frac{1}{2}b((\ell - m)\delta, \hat{X}^n((\ell - m)\delta), \hat{X}^n((\ell - 2m)\delta))b^{(0,0,1)}(\ell\delta, \hat{X}^n(\ell\delta), \hat{X}^n((\ell - m)\delta)) \\ \quad \times (\tilde{W}((\ell + 1 - m)\delta) - \tilde{W}((\ell - m)\delta))(\tilde{W}((\ell + 1)\delta) - \tilde{W}(\ell\delta)), & m \leq \ell \leq n - 1, \end{array} \right.$$

and let $\hat{\mathbf{X}}^n = (\hat{X}^n(t))_{t \in [-\tau, T]}$ be the linear interpolation of the previous discretization scheme, i.e. for $t \in [\ell\delta, (\ell + 1)\delta]$, $-m \leq \ell \leq n - 1$

$$\hat{X}^n(t) = \hat{X}^n(\ell\delta) + (\hat{X}^n((\ell + 1)\delta) - \hat{X}^n(\ell\delta))(t - \ell\delta)/\delta. \quad (24)$$

Remark 2.1. Note that $\hat{X}^n(t)$ depends on n only through δ and T , since by our assumption $n = T/\delta$. Therefore, for $t \in [0, T']$, with $T' = n' \delta$, $n' < n$, the processes $\hat{X}^{n'}(t)$ and $\hat{X}^n(t)$ coincide.

In [9] the authors have proved that the above approximation scheme is asymptotically optimal in a sense that we are going to explain. To this end we need to introduce the following constants

$$\mathcal{C}_T = \left(\frac{T}{4} \int_0^T E[\theta^2(u, T)] du \right)^{1/2}$$

and

$$\bar{\mathcal{C}}_T = \left(\frac{T}{6} \int_0^T E[b^2(u, X(u), X(u - \tau))] du + \frac{T}{4} \int_0^T \int_0^T E[\theta^2(u, v)] dv du \right)^{1/2}$$

where

$$\theta(u, v) = \begin{cases} \Phi(u, v) b(u - \tau, X(u - \tau), X(u - 2\tau)) b^{(0,0,1)}(u, X(u), X(u - \tau)), & \text{for } \tau \leq u \leq v, \\ 0, & \text{otherwise} \end{cases} \quad (25)$$

and $\Phi(u, v) = 0$, for $u > v$, while, for $v > u$

$$d\Phi(u, v) = a^{(0,1,0)}(v, X(v), X(v - \tau)) \cdot \Phi(u, v) dv + a^{(0,0,1)}(v, X(v), X(v - \tau)) \cdot \Phi(u, v - \tau) dv \\ + b^{(0,1,0)}(v, X(v), X(v - \tau)) \cdot \Phi(u, v) d\tilde{W}_v + b^{(0,0,1)}(v, X(v), X(v - \tau)) \cdot \Phi(u, v - \tau) d\tilde{W}_v,$$

with initial condition $\Phi(u, u) = 1$.

Remark 2.2. Note that under condition (A5) the constant $\bar{\mathcal{C}}_T$ is strictly positive, whereas, when the diffusion coefficient does not depend on the delay state variable x_2 , i.e. when $b^{(0,0,1)} = 0$, the constant \mathcal{C}_T equals zero, and the constant $\bar{\mathcal{C}}_T$ is finite.

Furthermore observe that when either η is a deterministic function (and, by (A3), Hölder continuous of order 1/2) or is a random continuous function with

$$E\left[\sup_{s \in [-\tau, 0]} |\eta(s)|^q \right] < \infty$$

for q sufficiently large, then (see Remarks 4 and 5 in [9])

$$E\left[\sup_{0 \leq u, v \leq T} \theta^2(u, v) \right] < \infty,$$

and therefore \mathcal{C}_T and $\bar{\mathcal{C}}_T$ are both finite.

Then we can state the optimality result for the approximation scheme proved in [9].

Proposition 2.3. Assume that \mathcal{C}_T and $\bar{\mathcal{C}}_T$ are finite. Then, for any pair of non-negative numbers α and β , such that $\alpha \bar{\mathcal{C}}_T + \beta \mathcal{C}_T > 0$, the approximation \hat{X}^n defined by (24) satisfies

$$\left(E\left[\int_0^T |\hat{X}^n(s) - X(s)|^2 ds \right] \right)^{1/2} + \beta \left(E\left[|\hat{X}^n(T) - X(T)|^2 \right] \right)^{1/2} \approx (\alpha \bar{\mathcal{C}}_T + \beta \mathcal{C}_T) n^{-1/2}, \quad (26)$$

where \approx denotes the strong asymptotic equivalence of sequence of real numbers, i.e. $a_n \approx b_n$ if $\lim a_n/b_n = 1$.

Furthermore, the approximation $\hat{\mathbf{X}}^n$ is asymptotically the best among the approximation schemes which are functions of $\{W(\ell\delta)\}_{\ell=0,\dots,n}$, in the following sense: if $\hat{\mathbf{X}}_*^n$ denotes the best mean square approximation of \mathbf{X} , i.e. for all $t \leq T$

$$\hat{X}_*^n(t) = E[X(t) \mid \tilde{W}(t_1), \dots, \tilde{W}(t_n)],$$

where $t_\ell = \ell\delta$, (so that $t_n = T$), and if \mathcal{C}_T is strictly positive, then the analogous of (26) obtained by replacing $\hat{\mathbf{X}}^n$ with $\hat{\mathbf{X}}_*^n$ holds.

Proof. The result follows immediately by Theorem 1 and Theorem 2 in [9] which guarantee that

$$\left(E \left[\int_0^T |\hat{X}^n(s) - X(s)|^2 ds \right] \right)^{1/2} \approx \bar{\mathcal{C}}_T n^{-1/2}. \quad (27)$$

and, when $\mathcal{C}_T > 0$,

$$\left(E [|\hat{X}^n(T) - X(T)|^2] \right)^{1/2} \approx \mathcal{C}_T n^{-1/2} \quad (28)$$

while, when $\mathcal{C}_T = 0$,

$$\left(E [|\hat{X}^n(T) - X(T)|^2] \right)^{1/2} \leq \gamma_T n^{-3/4}$$

for some constant $\gamma_T > 0$, depending on a , b and $\boldsymbol{\eta}$. Furthermore, when $\mathcal{C}_T > 0$, the analogous of (27) and (28) obtained by replacing $\hat{\mathbf{X}}^n$ with $\hat{\mathbf{X}}_*^n$ hold. \square

Remark 2.4. In Proposition 2.3 we have implicitly assumed that $\tau < T$. As observed at the end of Section 6 in [9], page 104, the above assumption is not necessary, and the result holds in much more generality, without any restriction on the time lag and the time horizon T .

As a consequence of the previous Remarks 2.1 and 2.4, denoting by $\mathcal{C}_{t,T}$ and $\bar{\mathcal{C}}_{t,T}$ the quantities defined by

$$\mathcal{C}_{t,T} = \left(\frac{T}{4} \int_0^t E[\theta^2(u, t)] du \right)^{1/2}$$

and

$$\bar{\mathcal{C}}_{t,T} = \left(\frac{T}{6} \int_0^t E[b^2(u, X(u), X(u - \tau))] du + \frac{T}{4} \int_0^t \int_0^t E[\theta^2(u, v)] dv du \right)^{1/2},$$

the above Proposition 2.3 can be extended as follows.

Proposition 2.5. Let $\bar{t} = \frac{i}{2^j} T$, for some integers i and j , with $i \leq 2^j$ and assume that $\mathcal{C}_{\bar{t},T}$ and $\bar{\mathcal{C}}_{\bar{t},T}$ are finite. Then, for any pair of non-negative numbers α and β , such that $\alpha \bar{\mathcal{C}}_{\bar{t},T} + \beta \mathcal{C}_{\bar{t},T} > 0$, and for $n = 2^k$, with $k \geq j$, the approximation $\hat{\mathbf{X}}^n$ defined by (24) satisfies

$$\alpha \left(E \left[\int_0^{\bar{t}} |\hat{X}^n(s) - X(s)|^2 ds \right] \right)^{1/2} + \beta \left(E [|\hat{X}^n(\bar{t}) - X(\bar{t})|^2] \right)^{1/2} \approx (\alpha \bar{\mathcal{C}}_{\bar{t},T} + \beta \mathcal{C}_{\bar{t},T}) n^{-1/2},$$

as k goes to infinity.

Proof. The proof can be achieved by applying Proposition 2.3 with \bar{t} instead of T , and taking into account Remark 2.1 with $n = 2^k$, $\delta = T/2^k$, $T' = \bar{t}$, and $n' = i2^{k-j}$. \square

Remark 2.6. When the times t_ℓ in the grid are not equidistant, one cannot assert that the modified Milstein scheme is asymptotically the best. Nevertheless the rate of convergence for both schemes $\hat{\mathbf{X}}^n$ and $\hat{\mathbf{X}}_*^n$ is still of order $n^{-1/2}$ (see [9]).

3 The approximation of the filter

In this section we assume that the approximating sequence for the state process considered in Section 1 is the Milstein type approximation sequence given by (24).

Theorem 3.1. Assume conditions (A1)–(A5), together with (B1) and (B2) for the partially observed system (1)-(2) are satisfied and that \mathcal{C}_T and $\bar{\mathcal{C}}_T$ are finite.

Then the sequence of approximating filters $\tilde{\pi}_T^n$ defined in (7), with the approximation processes \mathbf{Y}^n given by (8), converges in probability in bounded Lipschitz metric to the original filter in the following sense

$$\limsup_{n \rightarrow \infty} \sqrt{n} E[d_{BL}(\pi_T, \tilde{\pi}_T^n)] \leq 4L_h (2c\tau T + \bar{\mathcal{C}}_T^2)^{1/2} + \mathcal{C}_T, \quad (29)$$

where c and L_h are defined in (A3) and (B2), respectively.

Proof. In the proof of Theorem 1.1 we have shown that the bound (16) for $E[d_{BL}(\pi_T, \tilde{\pi}_T^n)]$ holds. Taking into account that for $s \in [-\tau, 0]$ the process $\hat{X}^n(s)$ is the linear interpolation of $\eta(\ell\delta) = X(\ell\delta)$, for $\ell \in \{-m, -m+1, \dots, 0\}$, by condition (A3) we get that

$$E\left[\int_{-\tau}^0 |\hat{X}^n(s) - X(s)|^2 ds\right] \leq 2c \frac{T}{n}.$$

Then the result follows by Proposition 2.3. □

We now assume that $n = 2^k$, so that the grid $\{t_0, t_1, \dots, t_n\}$ with $t_\ell = \ell T/n$ is increasing with n .

Theorem 3.2. Assume conditions (A1)–(A5), together with (B1) and (B2) for the partially observed system (1)-(2) are satisfied, that

$$\sup_{t \in [0, T]} \mathcal{C}_{t, T} \leq \mathcal{C} < \infty, \quad (30)$$

and finally that $\bar{\mathcal{C}}_T$ is finite.

Then the sequence of approximating filters defined in (7), with the approximation processes \mathbf{Y}^n given by (8), converges in probability in bounded Lipschitz metric to the original filter in the following sense

$$\limsup_{\substack{n \rightarrow \infty \\ n = 2^k}} \sqrt{n} \max_{\ell=0, \dots, n} E[d_{BL}(\pi_{t_\ell}, \tilde{\pi}_{t_\ell}^n)] \leq 4L_h (2c\tau T + \bar{\mathcal{C}}_T^2)^{1/2} + \mathcal{C}, \quad (31)$$

where c and L_h are defined in (A3) and (B2), respectively.

Proof. The proof is similar to that of Theorem 3.1, invoking Proposition 2.5 instead of Proposition 2.3. □

Observe that condition (30) is satisfied, provided that one of the conditions considered in Remark 2.2 is satisfied.

Remark 3.3. *The result of Theorem 3.1 can be extended to the case of h depending on a different number of fixed delays, as in Remark 1.2.*

On the other hand, when the observation function h does not depend on x_2 , i.e.

$$h(t, x_1, x_2) = h(t, x_1),$$

then the contribution due to the integral on $[-\tau, 0]$ disappears, and one can get the inequality

$$\limsup_{\substack{n \rightarrow \infty \\ n = 2^k}} \sqrt{n} \max_{\ell=0, \dots, n} E[d_{BL}(\pi_{t_\ell}, \tilde{\pi}_{t_\ell}^n)] \leq 2L_h \bar{\mathcal{C}}_T + \mathcal{C}. \quad (32)$$

Finally, when both the observation function h and the diffusion coefficient b do not depend on x_2 , then one can get the following inequality

$$\limsup_{\substack{n \rightarrow \infty \\ n = 2^k}} \sqrt{n} \max_{\ell=0, \dots, n} E[d_{BL}(\pi_{t_\ell}, \tilde{\pi}_{t_\ell}^n)] \leq 2L_h \left(\frac{T}{6} \int_0^T E[b^2(u, X(u))] du \right)^{1/2}. \quad (33)$$

Similar results can be obtained when considering the approximation processes \mathbf{Y}^n given by (18) instead of (8). In particular the following result holds.

Theorem 3.4. *Assume conditions (A1)—(A5), together with (B1) and (B2) for the partially observed system (1)-(2) are satisfied, that condition (30) holds and finally that $\bar{\mathcal{C}}_T$ is finite.*

Then the sequence of approximating filters defined in (7), with the approximation processes \mathbf{Y}^n given by (18), converges in probability in bounded Lipschitz metric to the original filter in the following sense

$$\limsup_{\substack{n \rightarrow \infty \\ n = 2^k}} \sqrt{n} \max_{\ell=0, \dots, n} E[d_{BL}(\pi_{t_\ell}, \tilde{\pi}_{t_\ell}^n)] \leq 4\sqrt{2}L_h \left[(2c\tau T + \bar{\mathcal{C}}_T^2)^{1/2} + (CT)^{1/2} \right] + \sqrt{2}(\mathcal{C}_T + 1), \quad (34)$$

where c , L_h and C are defined in (A3), (B2) and (23), respectively.

Proof. The proof is similar to that of Theorem 3.2, invoking inequality (20) in the proof of Theorem 1.3 instead of inequality (16) in the proof of Theorem 1.1. \square

4 Final remarks

In this section we discuss briefly how by the same techniques different approximations schemes for the filter at every time $t \leq T$ can be easily constructed.

Let us consider the Euler-Maruyama scheme $\mathbf{X}_n^E = (X_n^E(t))_{t \in [-\tau, T]}$, the truncated Milstein scheme $\mathbf{X}_n^M = (X_n^M(t))_{t \in [-\tau, T]}$ and the Milstein scheme $\mathbf{X}_n^{\bar{M}} = (X_n^{\bar{M}}(t))_{t \in [-\tau, T]}$ for stochastic delay differential equations, defined as follows

$$X_n^E(t) = X_n^{\overline{M}}(t) = \eta(t), \quad -\tau \leq t \leq 0,$$

and for $t \in (\ell\delta, (\ell+1)\delta]$, with $0 \leq \ell \leq n-1$,

$$\begin{aligned} X_n^E(t) &= X_n^E(\ell\delta) + a(\ell\delta, X_n^E(\ell\delta), X_n^E((\ell-m)\delta))(t - \ell\delta) \\ &\quad + b(\ell\delta, X_n^E(\ell\delta), X_n^E((\ell-m)\delta))[\tilde{W}(t) - \tilde{W}(\ell\delta)] \end{aligned}$$

$$\begin{aligned} X_n^{\overline{M}}(t) &= X_n^{\overline{M}}(\ell\delta) + a(\ell\delta, X_n^{\overline{M}}(\ell\delta), X_n^{\overline{M}}((\ell-m)\delta))(t - \ell\delta) \\ &\quad + b(\ell\delta, X_n^{\overline{M}}(\ell\delta), X_n^{\overline{M}}((\ell-m)\delta))[\tilde{W}(t) - \tilde{W}(\ell\delta)] \\ &\quad + \frac{1}{2}(bb^{(0,1,0)})(\ell\delta, X_n^{\overline{M}}(\ell\delta), X_n^{\overline{M}}((\ell-m)\delta))[(\tilde{W}(t) - \tilde{W}(\ell\delta))^2 - (t - \ell\delta)], \end{aligned}$$

as far as the Euler-Maruyama and Milstein truncated scheme are concerned, while, for the Milstein scheme

$$X_n^M(t) = X_n^{\overline{M}}(t), \quad \text{for } -\tau \leq t \leq \tau = m\delta,$$

and for $t \in (\ell\delta, (\ell+1)\delta]$, with $m \leq \ell \leq n-1$

$$\begin{aligned} X_n^M(t) &= X_n^M(\ell\delta) + a(\ell\delta, X_n^M(\ell\delta), X_n^M((\ell-m)\delta))(t - \ell\delta) \\ &\quad + b(\ell\delta, X_n^M(\ell\delta), X_n^M((\ell-m)\delta))[\tilde{W}(t) - \tilde{W}(\ell\delta)] \\ &\quad + \frac{1}{2}(bb^{(0,1,0)})(\ell\delta, X_n^M(\ell\delta), X_n^M((\ell-m)\delta))[(\tilde{W}(t) - \tilde{W}(\ell\delta))^2 - (t - \ell\delta)] \\ &\quad + \frac{1}{2}B((\ell-m)\delta, \ell\delta, X_n^M(\ell\delta), X_n^M((\ell-m)\delta), X_n^M((\ell-2m)\delta))J_n(t), \end{aligned}$$

where $B(u, v, x_1, x_2, x_3) = b(u, x_2, x_3)b^{(0,0,1)}(v, x_1, x_2)$ and

$$J_n(t) = \int_{\ell\delta}^t [\tilde{W}(s - m\delta) - \tilde{W}((\ell-m)\delta)]d\tilde{W}(s) = \int_{\ell\delta}^t [\tilde{W}(s - \tau) - \tilde{W}(\ell\delta - \tau)]d\tilde{W}(s),$$

for $t \geq \ell\delta$.

The rate of convergence of these approximations is related to the following upper bounds for the error (see [9], Proposition 1, and for more general results [10] and [11]):

$$\sup_{t \in [0, T]} \left(E[|X(t) - X_n^E(t)|^q] \right)^{1/q} \leq c^E n^{-1/2} \quad (35)$$

$$\sup_{t \in [0, T]} \left(E[|X(t) - X_n^{\overline{M}}(t)|^q] \right)^{1/q} \leq c^{\overline{M}} n^{-1/2} \quad (36)$$

for every $q \geq 1$, and

$$\sup_{t \in [0, T]} \left(E[|X(t) - X_n^M(t)|^2] \right)^{1/2} \leq c^M n^{-1} \quad (37)$$

where c^E , $c^{\overline{M}}$ and c^M denote constants that only depend on T , a , b , and η .

Similarly to the procedure of Section 3, we can define $\tilde{\pi}_n^E = (\tilde{\pi}_n^E(t))_{t \in [0, T]}$ by the following analogous of equality (7)

$$\tilde{\pi}_n^E(t) = U_n^E(t, \mathbf{Y})$$

where U_n^E is the functional associated to the filter of the partially observed system $(\mathbf{X}_n^E, \mathbf{Y}^n)$, where \mathbf{Y}^n is given by the analogous of (8) with $\hat{\mathbf{X}}^n$ replaced by \mathbf{X}_n^E . It is easy to see that, when we replace $\hat{\mathbf{X}}^n$ with \mathbf{X}_n^E in (16), the estimate still holds.

Then, taking into account that the initial path η has not been approximated, and making use of Remark 1.5 and estimate (35), we can easily derive the following upper bound

$$\sup_{t \in [0, T]} E \left[d_{BL}(\pi(t), \tilde{\pi}_n^E(t)) \right] \leq k n^{-\frac{1}{2}}$$

for a suitable constant k . In the same way we can construct $\tilde{\pi}_n^{\overline{M}}$ and $\tilde{\pi}_n^M$ as approximation schemes for the filter using $\mathbf{X}_n^{\overline{M}}$ and \mathbf{X}_n^M as approximations for the state. Moreover, taking into account Remark 1.5 together with estimates (36) and (37), we can derive explicit upper bounds for $\sup_{t \in [0, T]} E \left[d_{BL}(\pi(t), \tilde{\pi}_n^{\overline{M}}(t)) \right]$ and $\sup_{t \in [0, T]} E \left[d_{BL}(\pi(t), \tilde{\pi}_n^M(t)) \right]$, of order $n^{-1/2}$ and n^{-1} respectively.

Note that when the initial path is not approximated, the Milstein scheme as an approximation for the state process gives the fastest rate of convergence for the filter approximations. This property is lost when approximating the initial path by the piecewise linear interpolation of $\eta(\ell\delta)$, for $\ell = -m, \dots, 0$. Indeed, in the latter case, independently of the approximation scheme for the state process, the rate of convergence for the filter approximations depends on the square root of

$$\sup_{s, t \in [-\tau, 0], |s-t| \leq \delta} E \left[|\eta(t) - \eta(s)|^2 \right],$$

and so, under condition (A3), the rate is of order $n^{-1/2}$.

As already observed in Remark 1.4, when using (18) as an approximation for the observation process, the rate of convergence for the filter approximations is always of order $n^{-1/2}$, independently of the chosen approximation scheme for the state process. It is interesting to note that when we choose the Milstein scheme as described above, then, by means of inequality (20) in the proof of Theorem 1.3 we get that

$$\limsup_{n \rightarrow \infty} \sqrt{n} \sup_{t \in [0, T]} E \left[d_{BL}(\pi(t), \tilde{\pi}_n^M(t)) \right] \leq 4 L_h (CT)^{1/2} + 1, \quad (38)$$

where C is defined in (23). Though the above theoretical result is better than the one obtained in Theorem 3.4, it is important to note that usually, in order to compute $\tilde{\pi}_n^M(t)$ by means of the analogous of (22) with $\hat{\mathbf{X}}^n$ replaced by \mathbf{X}_n^M , one has to implement a Montecarlo method. With this aim, one has first of all to restrict the time interval to a finite discretization grid. Then, if we consider the grid $t_\ell = \ell\delta$ and take n equal to a power of 2, so that we have just to simulate $X_n^M(\ell\delta)$, besides the simulation error, one has to take into account that the stochastic integral $J_n(\ell\delta)$ is approximated by a sum involving $\tilde{W}(s_j)$ for a finite numbers of times s_j . Therefore the approximation scheme used in this simulation is not anymore the Milstein scheme, but is a scheme involving \tilde{W} evaluated in a finite numbers of times. Then the asymptotical (and theoretical) error is not given by (38), and it appears that when computed, it is even worse than the asymptotic error (34) obtained for the modified Milstein scheme.

References

- [1] M. ARRIOJAS, Y. HU, S.-E. A. MOHAMMED, G. PAP, *A delayed Black and Scholes formula*, Stoch. Anal. Appl., 25 (2007), no. 2, pp. 471–492.
- [2] A. G. BHATT, G. KALLIANPUR, AND R. L. KARANDIKAR, *Robustness of the nonlinear filter*, Stochastic Process. Appl., 81 (1999), pp. 247–254.
- [3] A. CALZOLARI, P. FLORCHINGER, AND G. NAPPO, *Approximation of nonlinear filters for Markov systems with delayed observations*, SIAM J. Control Optim., 45 (2006), pp. 599–633.
- [4] A. CALZOLARI, P. FLORCHINGER, AND G. NAPPO, *Convergence in nonlinear filtering for stochastic delay systems*, SIAM J. Control Optim., 46 (2007), pp. 1615–1636.
- [5] M. H. CHANG, *Discrete approximation of nonlinear filtering for stochastic delay equations*, Stochastic Anal. Appl., 5 (1987), pp. 267–298.
- [6] D. CRISAN, *Numerical methods for solving the stochastic filtering problem*, Numerical methods and stochastics (Toronto, ON, 1999), Fields Inst. Commun., Amer. Math. Soc., Providence, RI, 34 (2002), pp. 1–20.
- [7] P. DEL MORAL, Feynman-Kac formulae. Genealogical and interacting particle systems with applications. *Probability and its Applications (New York)*. Springer-Verlag, New York, 2004.
- [8] N. S. GOEL, S. C. MAITRA, AND E. W. MONTROLL, *On the Volterra and other nonlinear models of interacting populations*, Rev. Modern Phys., 43 (1971), pp. 231276.
- [9] N. HOFMANN AND T. MUELLER-GRONBACH, *A modified Milstein scheme for approximation of stochastic delay differential equations with constant time lag*, J. Computation Appl. Math., 197 (2006), pp. 89–121.
- [10] Y. HU, S.-E. A. MOHAMMED, AND F. YAN, *Numerical solution of stochastic differential systems with memory*, Departments of Mathematics, Southern Illinois University, Carbondale (2001), preprint.
- [11] Y. HU, S.-E. A. MOHAMMED, AND F. YAN, *Discrete-time approximations of stochastic delay equations: The Milstein scheme*, Ann. Probab., 32 (2004), pp. 265–314.
- [12] KURTZ, T. G., AND OCONE, D. L., *Unique characterization of conditional distributions in nonlinear filtering*, Ann. Probab., 16, (1988), 1, pp. 80–107.
- [13] H. J. KUSHNER, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Systems Control Found. Appl. 3, Birkhäuser Boston, Boston, MA, 1990.
- [14] H. J. KUSHNER, *Numerical approximations for nonlinear stochastic systems with delays*, Stochastics, 77 (2005), pp. 211–240.
- [15] F. LE GLAND, *Time discretization of nonlinear filtering equations*, in Proceedings of the 28th IEEE Conference on Decision and Control (Tampa, FL, 1989), Vols. 1–3, IEEE, New York, 1989, pp. 2601–2606.