

Convergence in nonlinear filtering for stochastic delay systems ^{*}

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

Abstract

We study some approximation schemes for a nonlinear filtering problem when the state process X is the solution of a stochastic delay diffusion equation, and the observation process is a noisy function of X_s for $s \in [t - \tau, t]$, where τ is a constant. The approximating state is given by means of an Euler discretization scheme, and the observation process is a noisy function of the approximating state.

Contents

1	Introduction	2
2	The model and its approximation	3
2.1	The model	3
2.2	Approximation schemes	6
3	Some related results	9
4	Approximation results	13
4.1	General considerations on approximation for filters	13
4.2	The main results	15

^{*}Partially supported by MURST project "Modelli probabilistici e applicazioni"

[†]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 - Università di Roma "La Sapienza", piazzale A. Moro 2, I 00185 Roma, Italy

1 Introduction

Let $(\mathbf{X}, \mathbf{Y}) = (X_t, Y_t)_{t \geq 0}$ be a stochastic system. Assume that the *state process* $\mathbf{X} = (X_t)_{t \geq 0}$ of the system cannot be directly observed, while the other component $\mathbf{Y} = (Y_t)_{t \geq 0}$ is completely observable, and therefore is referred to as the *observation process*. The aim of stochastic nonlinear filtering is to compute the conditional law π_t of the state process at time t , given the observation process up to time t , i.e. the computation of

$$\pi_t(\varphi) = E[\varphi(X_t) / \mathcal{F}_t^Y],$$

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 , $\mathcal{F}_t^Y = \sigma\{Y_s, s \leq t\}$.

A classical model of partially observed system extensively studied in the last past years arises when both the state and the observation processes are diffusion processes.

For this model, under suitable hypotheses on the coefficients, the filtering equations are well known (see for example Pardoux [15] or Kallianpur [6] and the references therein) and different approximation schemes have been studied (see for example Kushner [10] or Le Gland [12] and the references therein).

In this paper we are interested in nonlinear filtering of partially observed delay systems and in computable approximations of the filter.

To our knowledge there are only three papers dealing with nonlinear filtering for delay systems: Kwong and Willsky (1978) [11], Chang (1987) [4], and Kallianpur and Mandal (2002) [7].

In [11] Kwong and Willsky give a characterization of the optimal filter when dealing with nonlinear delay systems with gaussian noises. A Fujisaki-Kallianpur-Kunita equation for the filter is deduced from a representation result which characterizes conditional moment functionals of nonlinear delay systems. However the uniqueness of the solution of this equation is not guaranteed.

In [4] Chang gives a computable approximation for the optimal filter when dealing with one dimensional nonlinear delay filtering systems with gaussian noises. The original model is approximated by a discrete-time model obtained by applying an Euler discretization scheme. An optimal filter for the approximate system is obtained by an explicit procedure and the weak convergence of the approximating process and the approximating filter to the original ones are verified.

In [7] Kallianpur and Mandal study a nonlinear filtering problem where the state process is given as the solution of a stochastic delay differential equation and the observation depends not only on a function of the instantaneous value of the signal but also on the values of the signal from the past. By using some extensions of results obtained by Mohammed [14] for stochastic delay differential equations they prove that the signal process is the unique solution to an appropriate martingale problem. By taking this fact into account the authors prove that the optimal filter corresponding to the nonlinear filtering problem is the unique solution of a Zakai-type equation. A Fujisaki-Kallianpur-Kunita equation for the filter is also deduced from the Zakai-type equation.

The main results of these papers will be presented in a more extensive way in Section 3.

The main result of this paper concerning the convergence of the approximation scheme introduced in Section 2.2 to the system introduced in Section 2.1 is stated and proved in Section 4.2. The proof is based on the convergence results of Bhatt, Kallianpur and Karandikar (1999) [2] and Bhatt and Karandikar (2002) [3], which are recalled in Section 4.1.

2 The model and its approximation

2.1 The model

The state satisfies the following kind of delay differential equation on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, P)$

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

where

(H1) $\tau > 0$, and $\Pi_t X$ is a $C([-\tau, 0], \mathbb{R})$ random valued process defined by

$$\Pi_t X(s) = X(t + s) \quad -\tau \leq s \leq 0,$$

(H2) $\{\tilde{W}(t); t \geq 0\}$ is a standard Brownian motion,

(H3) η is a \mathcal{F}_0 square integrable $C([-\tau, 0], \mathbb{R})$ valued random variable, that is

$$E(\|\Pi_0 X\|^2) = E\left(\sup_{s \in [-\tau, 0]} |\eta(s)|^2\right) < \infty,$$

(H4) a and b are two Borel measurable functionals on $[0, T] \times C([-\tau, 0], \mathbb{R})$ satisfying the Lipschitz condition

$$|a(t, \theta) - a(t, \bar{\theta})|^2 + |b(t, \theta) - b(t, \bar{\theta})|^2 \leq K \|\theta - \bar{\theta}\|^2, \quad (2)$$

and the growth condition

$$|a(t, \theta)|^2 + |b(t, \theta)|^2 \leq K(1 + \|\theta\|^2), \quad (3)$$

for some constant $K > 0$.

The observation process is given by

$$Y(t) = \int_0^t h(u, \Pi_u X) du + W(t), \quad 0 \leq t \leq T, \quad (4)$$

where

(H5) $\{W(t); t \geq 0\}$ is a standard Brownian motion, independent of $\{\tilde{W}(t); t \geq 0\}$,

(H6) $h : [0, T] \times C([-\tau, 0], \mathbb{R}) \rightarrow \mathbb{R}$ is a Borel measurable function such that

$$\int_0^T E[|h(u, \Pi_u X)|^2] du < \infty.$$

In this paper we consider the filtering problem of the delay system $(\Pi_t X, Y(t))_{t \geq 0}$, i.e. we want to compute

$$\pi_t(\phi) = \mathbb{E}[\phi(\Pi_t X) | \mathcal{F}_t^Y],$$

for measurable and bounded functions ϕ mapping $C([-\tau, 0], \mathbb{R})$ into \mathbb{R} .

Conditions **(H1)** to **(H4)** imply the existence and uniqueness of the solution of the stochastic delay differential equation (1) (see Mohammed [14]), though the square integrability condition **(H3)** is not necessary (see Kallianpur and Mandal [7]). However condition **(H3)**, together with the growth condition (3) in **(H4)**, implies that

$$E \left[\sup_{u \in [0, T]} \|\Pi_u X\|^2 \right] < \infty, \quad (5)$$

(see [14], Theorem II.2.1 and Lemma III.1.2).

Note that in the homogeneous case, i.e. $a(t, \theta) = a(\theta)$ and $b(t, \theta) = b(\theta)$, the Lipschitz assumption (2) in **(H4)** implies the growth condition (3). Moreover in the homogeneous case Kallianpur and Mandal prove in [7] that the process $(\Pi_t X)_{t \geq 0}$ is Markov. This result extends a previous result by Mohammed [14] for the case a and b bounded.

Conditions **(H5)** and **(H6)** are classical assumptions in nonlinear filtering theory which guarantee that the filter π_t can be represented via a Kallianpur–Striebel formula

$$\pi_t(\phi) = \frac{\sigma_t(\phi)}{\sigma_t(\mathbf{1})},$$

with

$$\sigma_t(\phi) = E^0 \left[\phi(\Pi_t X) \exp \left\{ \int_0^t h(s, \Pi_s X) dY_s - \frac{1}{2} \int_0^t |h(s, \Pi_s X)|^2 ds \right\} \middle| \mathcal{F}_t^Y \right],$$

where E^0 denotes the expectation w.r.t. the reference measure P^0 , defined by the Radon–Nikodym derivative

$$\frac{dP^0}{dP} = \exp \left\{ - \int_0^T h(s, \Pi_s X) dY_s + \frac{1}{2} \int_0^T |h(s, \Pi_s X)|^2 ds \right\}. \quad (6)$$

This fact will play a fundamental role in the proof of our approximation results.

Furthermore, the following conditions on the observation function h will also be used in our analysis.

(H7) h is sublinear, i.e. $|h(t, \theta)|^2 \leq K(1 + \|\theta\|^2)$,

(H8) h is jointly continuous.

Note that condition **(H7)** together with the integrability condition (5) on the state process implies that **(H6)** is also satisfied.

Condition **(H8)** will be used in the proof of our approximation results, together with the following condition.

(H9) The functions $a(t, \theta)$ and $b(t, \theta)$ are jointly globally Lipschitz, i.e.

$$|a(t, \theta) - a(t', \bar{\theta})|^2 + |b(t, \theta) - b(t', \bar{\theta})|^2 \leq K (\|t - t'\|^2 + \|\theta - \bar{\theta}\|^2).$$

Note that the above joint global Lipschitz condition is stronger than condition **(H4)**, but in the homogeneous case, i.e. when $a(t, \theta) = a(\theta)$ and $b(t, \theta) = b(\theta)$, both conditions are clearly equivalent to (2).

In addition, to prove our approximation results, condition **(H3)** is not strong enough and we will assume that the initial condition of the stochastic delay differential equation (1) is such that

$$\textbf{(H10)} \quad E(\|\Pi_0 X\|^4) = E\left(\sup_{s \in [-\tau, 0]} |\eta(s)|^4\right) < \infty.$$

Note that under this condition it is easy to prove that $\mathbb{E}[\sup_{t \in [0, T]} \|\Pi_t X\|^4]$ is finite.

To conclude this section, note that conditions **(H7)** to **(H9)** are clearly satisfied if, for example, the functions a , b and h are given by

$$a(t, \theta) = \varphi_a(t, \int_{-\tau}^0 \psi_a(u, \theta(u)) du), \quad b(t, \theta) = \varphi_b(t, \int_{-\tau}^0 \psi_b(u, \theta(u)) du),$$

$$h(t, \theta) = \varphi(t, \int_{-\tau}^0 \psi(u, \theta(u)) du),$$

where φ_a , φ_b , φ , ψ_a , ψ_b and ψ are joint Lipschitz functions.

2.2 Approximation schemes

In this paper we will consider some approximation models for both the delay system and the filter, but the main results concern the following approximation scheme:

The approximation X^n of the state process is the linear interpolation of the Euler discretization scheme with step $\delta = \delta_n = T/n$, with $\tau = m\delta$ (as in Chang [4], for the sake of simplicity, we assume that T/τ is rational)¹:

$$\begin{cases} X^n(\ell\delta) = \eta(\ell\delta), & -m \leq \ell \leq 0, \\ X^n((\ell + 1)\delta) = X^n(\ell\delta) + a(\ell\delta, \Pi_{\ell\delta}X^n)\delta \\ \quad + b(\ell\delta, \Pi_{\ell\delta}X^n)[\tilde{W}((\ell + 1)\delta) - \tilde{W}(\ell\delta)], & 0 \leq \ell \leq n. \end{cases} \quad (7)$$

For the approximation of the observation process, we define $Y^n(t)$ by

$$Y^n(t) = \int_0^t h(\lfloor s/\delta \rfloor \cdot \delta, \Pi_{\lfloor s/\delta \rfloor \cdot \delta} X^n) ds + W(t), \quad 0 \leq t \leq T, \quad (8)$$

where $\lfloor x \rfloor$ is the integer part of x .

Remark 2.1. *It is clear that*

$$\left\{ (X^n(\ell\delta), X^n((\ell - 1)\delta), \dots, X^n((\ell - m)\delta)) \right\}_{\ell \geq 0}$$

is an $(m + 1)$ -dimensional Markov chain, and for $t \in [\ell\delta, (\ell + 1)\delta]$, $0 \leq \ell \leq n$

$$\begin{aligned} X^n(t) &= X^n(\ell\delta) + a(\ell\delta, \Pi_{\ell\delta}X^n)(t - \ell\delta) \\ &\quad + b(\ell\delta, \Pi_{\ell\delta}X^n)[\tilde{W}((\ell + 1)\delta) - \tilde{W}(\ell\delta)](t - \ell\delta)/\delta, \end{aligned} \quad (9)$$

with $X^n(0) = \eta(0)$, and

$$Y^n(t) = Y^n(\ell\delta) + h(\ell\delta, \Pi_{\ell\delta}X^n)(t - \ell\delta) + [W(t) - W(\ell\delta)], \quad (10)$$

with $Y^n(0) = 0$.

It is also interesting to note that therefore the approximation $(\bar{Y}^n(t))_{t \geq 0}$ for the observation process proposed by Chang [4] (see also (17) in Section 3), is the linear interpolation of $Y^n(\ell\delta)$.

Furthermore, we would also emphasize the fact that when using (8) we do not make use, as in [4], of a discrete time observation process but still of a continuous time diffusion process, such that $Y^n(l\delta) = \bar{Y}^n(l\delta)$.

¹It is clear that, assuming $T = \frac{p}{q}\tau$, we first fix $m = kq$ a multiple of q and then set $\delta = \tau/m$, so that $T = kp\delta$ and $n = kp$.

Or better, we first fix m , then set $\delta = \tau/m$, and finally take the interval $[-\tau, \lfloor T/\delta \rfloor \delta]$, instead of $[-\tau, T]$, so that $n = n(m)$.

With these approximations for the state process and the observation process we define the approximating filter π_t^n by :

$$\pi_t^n(\phi) = \frac{\sigma_t^n(\phi)}{\sigma_t^n(\mathbf{1})} = \frac{E^{0,n} [\phi(\Pi_t X^n) \mathcal{L}_t^n | \mathcal{F}_t^{Y^n}]}{E^{0,n} [\mathcal{L}_t^n | \mathcal{F}_t^{Y^n}]}, \quad (11)$$

where $E^{0,n}$ denotes the expectation w.r.t. the reference measure $P^{0,n}$, defined by the Radon-Nikodym derivative

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

with

$$\mathcal{L}_t^n = \exp \left\{ \int_0^t h(\lfloor s/\delta \rfloor \cdot \delta, \Pi_{\lfloor s/\delta \rfloor \cdot \delta} X^n) dY_s^n - \frac{1}{2} \int_0^t |h(\lfloor s/\delta \rfloor \cdot \delta, \Pi_{\lfloor s/\delta \rfloor \cdot \delta} X^n)|^2 ds \right\}.$$

Taking into account that $s \mapsto h(\lfloor s/\delta \rfloor \cdot \delta, \Pi_{\lfloor s/\delta \rfloor \cdot \delta} X^n)$ is piecewise constant, we have that

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

where,

$$\log L_{\ell\delta}^n(x(\cdot), y_0, y_\delta, \dots, y_{\ell\delta}) = \sum_{k=0}^{\ell\delta-1} h(k\delta, \Pi_{k\delta} x(\cdot))(y_{k+1} - y_k) - \frac{1}{2} \sum_{k=0}^{\ell\delta-1} |h(k\delta, \Pi_{k\delta} x(\cdot))|^2 \delta,$$

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

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

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

$$\sigma_t^n(\phi) = E \left[\phi(\Pi_t X^n) L_t^n(X^n(\cdot), y_0, y_\delta, \dots, y_{\ell\delta}, y) \right] \Bigg|_{y_0=Y_0^n, y_1=Y_\delta^n, \dots, y_\ell=Y_{\ell\delta}^n, y=Y_t^n}.$$

Note that the filter π_t^n defined above depends on the approximated observation process Y^n , which is not directly observable. To overcome this difficulty we also consider the following approximation $\tilde{\pi}_t^n$ for the filter

$$\tilde{\pi}_t^n(\phi) = \frac{\tilde{\sigma}_t^n(\phi)}{\tilde{\sigma}_t^n(\mathbf{1})}, \quad (12)$$

where

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

Furthermore, as an approximation for the state we will also consider the following diffusion

$$\begin{cases} Z^n(t) := \eta(t) & -\tau \leq t \leq 0, \\ Z^n(t) := \eta(0) + \int_0^t a(\delta \cdot \lfloor s/\delta \rfloor, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} X^n) ds \\ \quad + \int_0^t b(\delta \cdot \lfloor s/\delta \rfloor, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} X^n) d\tilde{W}_s, & 0 \leq t \leq T, \end{cases} \quad (13)$$

which will be used in the proof of our main result, since this diffusion has the property that

$$Z^n(\ell\delta) = X^n(\ell\delta), \text{ for all } \ell \geq -m.$$

Indeed $Z^n(\ell\delta) = X^n(\ell\delta) = \eta(\ell\delta)$, for $-m \leq \ell \leq 0$. The case $\ell \geq 0$ follows by observing that for $t \in [\ell\delta, (\ell+1)\delta]$

$$\begin{aligned} Z^n(t) &= Z^n(\ell\delta) + \int_{\ell\delta}^t a(\ell\delta, \Pi_{\ell\delta} X^n) ds + \int_{\ell\delta}^t b(\ell\delta, \Pi_{\ell\delta} X^n) d\tilde{W}_s \\ &= Z^n(\ell\delta) + a(\ell\delta, \Pi_{\ell\delta} X^n)(t - \ell\delta) + b(\ell\delta, \Pi_{\ell\delta} X^n)[\tilde{W}(t) - \tilde{W}(\ell\delta)], \end{aligned}$$

so that

$$Z^n((\ell+1)\delta) = Z^n(\ell\delta) + a(\ell\delta, \Pi_{\ell\delta} X^n)\delta + b(\ell\delta, \Pi_{\ell\delta} X^n)[\tilde{W}((\ell+1)\delta) - \tilde{W}(\ell\delta)],$$

and finally comparing the above recursive formula with (7).

Other possible choices for the approximation for the observation process are either

$$Y^n(t) = \int_0^t h(s, \Pi_s X^n) ds + W(t), \quad (14)$$

or

$$Y^n(t) = \int_0^t h(s, \Pi_{\lfloor s/\delta \rfloor \cdot \delta} X^n) ds + W(t). \quad (15)$$

These choices are motivated by the fact that we want to use the results of Bhatt, Kallianpur and Karandikar [2] (see also Section 4.1), on the approximation of the filter, when the observation is a diffusion, in both the limit model and the approximation model.

Assuming that conditions **(H1)** to **(H10)** are satisfied, we prove the weak convergence of the filter π_t^n for the approximation model to the filter π_t of the original delay system and the convergence in probability of $\tilde{\pi}_t^n$ to π_t (see Theorem 4.2 in Section 4.2).

To conclude this section, note that in order to compute π_t^n and $\tilde{\pi}_t^n$ we need to compute

(a) the transition probability of the $(m+1)$ -dimensional Markov chain

$$(X^n(\ell\delta), X^n((\ell-1)\delta), \dots, X^n((\ell-m)\delta)),$$

and therefore we need the explicit expression of $a(\ell\delta, \Pi_{\ell\delta} X^n)$ and of $b(\ell\delta, \Pi_{\ell\delta} X^n)$,

- (b) the explicit expression of L_t^n , and therefore we need the explicit expression of $h(\ell\delta, \Pi_{\ell\delta}X^n)$.

With the example given at the end of Section 2.1 we get

$$a(\ell\delta, \Pi_{\ell\delta}X^n) = \varphi_a(t, \int_{-\tau}^0 \psi_a(u, \Pi_{\ell\delta}X^n(u))du),$$

where

$$\begin{aligned} \int_{-\tau}^0 \psi_a(u, \Pi_{\ell\delta}X^n(u))du &= \int_{-\tau}^0 \psi_a(u, X^n(\ell\delta + u))du \\ &= \sum_{k=-m}^{-1} \int_{k\delta}^{(k+1)\delta} \psi_a(u, X^n((\ell+k)\delta) + \frac{u-k\delta}{\delta} [X^n((\ell+k+1)\delta) - X^n((\ell+k)\delta)])du. \end{aligned}$$

Similar expressions hold for $b(\ell\delta, \Pi_{\ell\delta}X^n)$ and $h(\ell\delta, \Pi_{\ell\delta}X^n)$.

3 Some related results

In [7], Kallianpur and Mandal deal with a nonlinear filtering problem described by a state process solution of the stochastic delay differential equation (1), in the homogeneous case, and an observation process given by (4). The authors consider the probability space (Ω, \mathcal{F}, Q) , where

$$\frac{dQ}{dP} = \exp \left\{ - \int_0^T h(u, \Pi_u X) dY_u + \frac{1}{2} \int_0^T |h(u, \Pi_u X)|^2 du \right\}$$

and the product space $(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F}, P \otimes Q)$. Denoting by $(X'(t))_{t \geq 0}$ and $(Y^\dagger(t))_{t \geq 0}$ the stochastic processes defined on $\Omega \times \Omega$ by

$$X'(t, \omega, \omega') = X(t, \omega') \quad Y^\dagger(t, \omega, \omega') = Y(t, \omega),$$

and by $(R(t))_{t \geq 0}$ the stochastic process given by

$$R(t) = \exp \left\{ \int_0^T h(u, \Pi_u X') dY_u^\dagger - \frac{1}{2} \int_0^T |h(u, \Pi_u X')|^2 du \right\},$$

they prove the Kallianpur-Striebel formula: for any ϕ in the space of bounded continuous functions from $C([-\tau, 0], \mathbb{R})$ into \mathbb{R} it holds

$$E_P [\phi(\Pi_t X) / \mathcal{F}_t^Y] = \frac{\sigma_t(\phi, Y)}{\sigma_t(1, Y)},$$

where

$$\sigma_t(\phi, Y) = \int_{\Omega} \phi(\Pi_t X(\omega')) R(t; \omega, \omega') P(d\omega')$$

is the unnormalized conditional expectation of $\phi(\Pi_t X)$ given \mathcal{F}_t^Y .

Then, assuming **(H1)** to **(H6)**, that ϕ is a quasi-tame function as defined in Mohammed [14], and that

$$E_P \left[\left(\int_0^T \{ |a(\Pi_u X)| + |b(\Pi_u X)|^2 \} du \right)^2 \int_0^T |h(u, \Pi_u X)|^2 du \right] < \infty, \quad (16)$$

they prove that the unnormalized conditional expectation of $\phi(\Pi_t X)$ given \mathcal{F}_t^Y solves the Zakai equation

$$\sigma_t(\phi, Y) = E_P[\phi(\eta)] + \int_0^t \sigma_s(A^0 \phi, Y) ds + \int_0^t \sigma_s(\phi h(s, \cdot), Y) dY_s$$

where A^0 is the infinitesimal generator of the process $\{\Pi_t X, t \geq 0\}$, as given in [14].

Before giving some examples of quasi-tame functions ϕ , and of $A^0 \phi$, we note that if furthermore the diffusion coefficient b in the definition of the signal process is bounded, and assumption **(H10)** on the fourth moment of the initial condition is satisfied, then (16) holds.

Example 1 If $\phi(\theta) = \varphi(\theta(0))$, where φ is a C^∞ bounded function, then

$$A^0 \phi(\theta) = \mathcal{L}\varphi(\theta(0)) = a(\theta) \cdot \varphi_x(\theta(0)) + \frac{1}{2} b^2(\theta) \varphi_{xx}(\theta(0)).$$

Example 2 If $\phi(\theta) = \int_{-\tau}^0 \psi(u) g(\theta(u), \theta(0)) du$, where ψ is piecewise C^1 , with $\dot{\psi}$ absolutely integrable, and $g(x_1, x_2)$ a bounded C^∞ function, then

$$\begin{aligned} A^0 \phi(\theta) &= \psi(0) g(\theta(0), \theta(0)) - \psi(-\tau) g(\theta(-\tau), \theta(0)) \\ &\quad - \int_{-\tau}^0 \dot{\psi}(u) g(\theta(u), \theta(0)) du + \int_{-\tau}^0 \psi(u) \mathcal{L}g(y, \cdot)(\theta(0))|_{y=\theta(u)} du \end{aligned}$$

where \mathcal{L} is the operator defined in example 1.

Example 3 If $\phi(\theta) = \varphi\left(\int_{-\tau}^0 \psi(u) g(\theta(u), \theta(0)) du\right)$, where φ is a bounded C^∞ function, with ψ and $g(x_1, x_2)$ as in example 2, then

$$A^0 \phi(\theta) = \varphi_x(\Phi(\theta)) A^0 \Phi(\theta) + \frac{1}{2} \varphi_{xx}(\Phi(\theta)) G(\theta)$$

where $\Phi(\theta) = \int_{-\tau}^0 \psi(u) g(\theta(u), \theta(0)) du$, $A^0 \Phi(\theta)$ is computed as in example 2, and finally

$$G(\theta) = \int_{-\tau}^0 \int_{-\tau}^0 \psi(u) \psi(v) g_{x_2}(\theta(u), \theta(0)) g_{x_2}(\theta(v), \theta(0)) dv du$$

(here g_{x_2} denotes partial differentiation of g with respect to its second argument).

The uniqueness for the solution of the Zakai equation is deduced from the results of Bhatt, Kallianpur and Karandikar [1], and a Fujisaki-Kallianpur-Kunita equation for the filter is deduced from the Zakai equation and the Kallianpur-Striebel formula by a usual method in nonlinear filtering theory.

In [11] Kwong and Willsky consider the delay differential equations for the n -dimensional state process

$$dX(t) = a(t, \Pi_t X) dt + F(t) d\tilde{W}_t, \quad \Pi_0 X = \eta,$$

and for the p -dimensional observation process

$$dY(t) = h(t, \Pi_t X) dt + N(t) dW_t, \quad Y(0) = 0,$$

where $N(t)$ is nonsingular, and a satisfies a Lipschitz condition and a growth condition.

Under additional conditions on the initial trajectory $\eta \in C([-\tau, 0], \mathbb{R}^n)$ and integrability conditions for $\int_0^T |h(t, \Pi_t X)|^2 dt$, $|\phi(\Pi_t X)|^2$ and $\int_0^T |\phi(\Pi_t X) h(t, \Pi_t X)|^2 dt$, the authors derive an evolution equation for functionals of the form $E^t[\phi(\Pi_t X)] \equiv E[\phi(\Pi_t X) | \mathcal{F}_t^Y]$, extending to their framework the result of Kunita [8]. In particular, they prove that the filter solves a Fujisaki-Kallianpur-Kunita equation, namely

$$dE^t[\phi(\Pi_t X)] = E^t[A_t \phi(\Pi_t X)] dt + E^t[\phi(\Pi_t X) (h(t, \Pi_t X) - E^t[h(t, \Pi_t X)])] (N(t) N^*(t))^{-1} d\nu_t$$

where $\nu_t = Y(t) - \int_0^t E^s[h(s, \Pi_s X)] ds$, for functions ϕ in the extended generator of the state process:

$$E[\phi(\Pi_t X) / \Pi_s X] = \phi(\Pi_t X) + \int_0^t E[A_u \phi(\Pi_u X) / \Pi_s X] du.$$

The expression of $A_t \phi(\Pi_t X)$ is similar to that of $A^0 \phi(\Pi_t X)$, but taking into account the fact that the state is multidimensional and given as the solution of a nonhomogeneous stochastic delay equation. For instance in example 1 the operator \mathcal{L} has to be replaced by

$$\mathcal{L}_t \varphi(\theta(0)) = a(t, \theta) \cdot \varphi_x(\theta(0)) + \frac{1}{2} \text{tr}\{F(t) F^*(t) \varphi_{xx}(\theta(0))\}.$$

In [4], Chang considers the optimal filtering problem for a partially observable model, when the state process and the observation process are real valued and satisfy the stochastic nonlinear differential equations

$$dX(t) = a(t, \Pi_t X) dt + d\tilde{W}_t, \quad \Pi_0 X = \eta$$

$$dY(t) = h(t, \Pi_t X) dt + dW_t, \quad Y(0) = 0$$

where \tilde{W} and W are two independent Brownian motions.

Under the assumptions:

(A1) $a : [0, T] \times C([-\tau, 0], \mathbb{R}) \rightarrow \mathbb{R}$; $(t, \vartheta) \mapsto a(t, \vartheta)$ is jointly continuous; the same holds for h ;

(A2) a and h are Lipschitz continuous in the variable ϑ (uniformly in $[0, T]$): there exists a finite measure λ on $[-\tau, 0]$ such that

$$|a(t, \vartheta) - a(t, \vartheta')| + |h(t, \vartheta) - h(t, \vartheta')| \leq \int_{-\tau}^0 |\vartheta(u) - \vartheta'(u)| \lambda(du),$$

for all $t \in [0, T]$, $\vartheta, \vartheta' \in C([-\tau, 0], \mathbb{R})$;

(A3) a and h satisfy the following linear growth condition: there exist a positive constant k and a finite measure γ on $[-\tau, 0]$ such that

$$|a(t, \vartheta)|^2 + |h(t, \vartheta)|^2 \leq k \left(1 + \int_{-\tau}^0 |\vartheta(u)|^2 \gamma(du) \right),$$

for all $t \in [0, T]$ and $\vartheta \in C([-\tau, 0], \mathbb{R})$;

(A4) η is a random function with values in $C([-\tau, 0], \mathbb{R})$ with probability 1, there exists a strictly positive constant k such that

$$E[\exp\{k\|\eta\|^2\}] < \infty,$$

and for any partition $-\tau \leq \tau_0 < \dots < \tau_n = 0$ the $(n+1)$ -dimensional random vector $(\eta(\tau_i); 0 \leq i \leq n)$ has a probability density w.r.t. the Lebesgue measure in \mathbb{R}^{n+1} .

Under the previous hypotheses Chang gives a computable approximation for the optimal filter. The state process of the original model is approximated by the linear interpolation of (7) with $b(t, \theta) = 1$, and the observation process is approximated by the linear interpolation of

$$Y^n((\ell+1)\delta) - Y^n(\ell\delta) = h(\ell\delta, \Pi_{\ell\delta} X^n) \delta + [W((\ell+1)\delta) - W(\ell\delta)], \quad 0 \leq \ell \leq n. \quad (17)$$

Then, the author shows that the optimal filter for the discrete time approximation can be designed by an explicit procedure and he verifies the weak convergence of the approximation process and the approximation filter to the original ones.

It is interesting to note some of the differences among the papers concerning the diffusion coefficient of the state process: namely the diffusion coefficient is a deterministic function $F(t)$ in [11], is just the constant 1 in [4], while it is a function depending only on the state in [7]. Another important observation is that Kwong and Willsky, as well as Chang consider non homogeneous delay systems, while Kallianpur and Mandal consider only homogeneous delay systems. Nevertheless, in the homogeneous case, conditions (A2) and (A3) of Chang imply that the Lipschitz condition

$$|a(\theta) - a(\bar{\theta})|^2 + |b(\theta) - b(\bar{\theta})|^2 \leq K \|\theta - \bar{\theta}\|^2$$

of Kallianpur and Mandal is satisfied with $K = \max\{(\lambda([-\tau, 0]))^2, k(\gamma([-\tau, 0]) \vee 1)\}$. Moreover, condition (A2) of Chang implies also that the observation coefficient h satisfies a Lipschitz condition while condition (A3) implies that the sublinear growth condition (H7) is satisfied.

4 Approximation results

4.1 General considerations on approximation for filters

Suppose that a sequence of partially observed systems $(\mathbf{X}^n, \mathbf{Y}^n)$ converges to a partially observed system (\mathbf{X}, \mathbf{Y}) . Then a natural question is whether the corresponding sequence of filters π^n converges to the filter π of the limit system.

Different kinds of convergence can be considered for both the systems and the filters. When all the processes are defined on the same probability space (Ω, \mathcal{F}, P) , almost sure convergence, convergence in probability or weak convergence (i.e. in distribution) can be studied. Weak convergence is the only notion that can be considered when the systems $(\mathbf{X}^n, \mathbf{Y}^n)$ are defined on different probability spaces $(\Omega^n, \mathcal{F}^n, P^n)$. Furthermore one can distinguish between convergence of π_t^n to π_t as random probability measures, for each t , and convergence of the processes $\boldsymbol{\pi}^n = (\pi_t^n; t \geq 0)$ to $\boldsymbol{\pi} = (\pi_t; t \geq 0)$, as càdlàg probability measure-valued processes. Some now classical result in this direction can be found in Goggin [5].

A different, but connected, approach to the problem of the approximation of the filter takes into account that a realistic approximation depends on the actually observed trajectory. Before describing it we need to recall that under very general conditions (see for instance, Kurtz and Ocone [9]), when denoting by S the state space, by F the observation space, and by $\mathcal{P}(S)$ the space of probability measures on S , there always exist deterministic functionals U^n and U , with paths in $D_{\mathcal{P}(S)}([0, T])$,

$$U^n, U : [0, T] \times D_F([0, T]) \mapsto \mathcal{P}(S)$$

with the properties $U^n(t, \mathbf{y}) = U^n(t, y(\cdot \wedge t))$ and $U(t, \mathbf{y}) = U(t, y(\cdot \wedge t))$, and such that

$$\pi_t^n = U^n(t, \mathbf{Y}^n) \text{ and } \pi_t = U(t, \mathbf{Y}).$$

Note that the functionals U^n and U depend on the joint distribution of $(\mathbf{X}^n, \mathbf{Y}^n)$ and (\mathbf{X}, \mathbf{Y}) and that they are defined in $D_F([0, T])$ almost surely with respect to $P_{\mathbf{Y}^n}$, the law of \mathbf{Y}^n , and with respect to $P_{\mathbf{Y}}$, the law of \mathbf{Y} , respectively.

When dealing with approximation problems the following typical situation can arise: the model is (\mathbf{X}, \mathbf{Y}) , and therefore we observe \mathbf{Y} , while the models $(\mathbf{X}^n, \mathbf{Y}^n)$ are merely *more manageable* approximations, despite the fact that it may be impossible to observe the approximate process \mathbf{Y}^n . In this situation the filter is π_t , and it is quite natural to consider

$$\tilde{\pi}_t^n = U^n(t, \mathbf{Y}) \tag{18}$$

as an approximation depending on the actually observed trajectory. The functional U^n is defined $P_{\mathbf{Y}^n}$ almost surely and therefore to define $\tilde{\pi}_t^n$ it is natural to assume that $P_{\mathbf{Y}}$ is absolutely continuous with respect to $P_{\mathbf{Y}^n}$, otherwise it could be impossible to define $\tilde{\pi}_t^n$. Note that, however, $\tilde{\pi}_t^n$ is not a conditional law. Then one can consider both the almost sure convergence and the convergence in probability of $\tilde{\pi}_t^n$ to π_t . It is important to note

that with this kind of approximation it is even not necessary for the processes \mathbf{Y}^n and \mathbf{Y} to be defined on the same probability space.

Some results on weak convergence, which also contain explicitly the approach discussed above, can be found in Bhatt, Kallianpur and Karandikar (1999) [2] and in Bhatt and Karandikar (2002) [3]. In these papers the observation is a diffusion, in both the limit model and the approximation model. In [2] the state and its approximation may be general processes, while in [3] they are both diffusions and the noises in the signal and the observation may be correlated. In [2] the following limit model $(\mathcal{X}_t, Y_t)_{t \geq 0}$ and approximation model $(\mathcal{X}_t^n, Y_t^n)_{t \geq 0}$ are considered.

The signal process \mathcal{X}_t , with values in a complete separable metric space S , is defined on (Ω, \mathcal{F}, P) and has *càdlàg* paths. The observation process Y_t is given by

$$Y(t) = \int_0^t \mathbf{h}(\mathcal{X}_s) ds + W(t), \quad (19)$$

where $(W(t))_{t \geq 0}$ is a standard Brownian motion, defined on (Ω, \mathcal{F}, P) , independent of \mathcal{X} , and \mathbf{h} is a measurable function on S with values in \mathbb{R}^k , such that

$$P \left(\int_0^T |\mathbf{h}(\mathcal{X}_s)|^2 ds < \infty \right) = 1.$$

The approximation signal processes \mathcal{X}_t^n are defined on $(\Omega^n, \mathcal{F}^n, P^n)$, and take values in S as well. The approximation observation processes Y_t^n are defined by

$$Y^n(t) = \int_0^t \mathbf{h}^n(\mathcal{X}_s^n) ds + W^n(t), \quad (20)$$

where $W^n(t)$ are standard Brownian motions, defined on $(\Omega^n, \mathcal{F}^n, P^n)$, independent of \mathcal{X}^n , and \mathbf{h}^n are measurable functions on S with values in \mathbb{R}^k , such that

$$P^n \left(\int_0^T |\mathbf{h}^n(\mathcal{X}_s^n)|^2 ds < \infty \right) = 1.$$

Assuming that

(B1) \mathbf{h}^n converges to \mathbf{h} uniformly on compact sets,

(B2) \mathbf{h} is continuous,

(B3) \mathcal{X}^n converges weakly to \mathcal{X} in $D_S([0, T])$,

(B4) $\lim_{n \rightarrow \infty} E^n \left(\int_0^T |\mathbf{h}^n(\mathcal{X}_s^n)|^2 ds \right) = E \left(\int_0^T |\mathbf{h}(\mathcal{X}_s)|^2 ds \right)$,

it is proved in [2] that

$$P^n \circ (\boldsymbol{\pi}^n)^{-1} \text{ converge weakly to } P \circ (\boldsymbol{\pi})^{-1}$$

as processes with values in $D_{\mathcal{P}(S)}([0, T])$, where $\mathcal{P}(S)$ is the metric space of probability measures on S , endowed with the Prohorov metric.

Note that the above conditions **(B1)**—**(B4)** are only sufficient conditions, and that in [2] weaker conditions can be found.

Moreover in [2] (see Theorem 3.3-(a)), as a step to prove the above weak convergence result, the authors prove that for any Wiener process $\{B(t); t \geq 0\}$, the $\mathcal{P}(S)$ -valued processes $(U^n(t, \mathbf{B}))_{t \in [0, T]}$ converge in probability to the $\mathcal{P}(S)$ -valued process $(U(t, \mathbf{B}))_{t \in [0, T]}$. This amounts to say that if P^0 is the reference probability measure defined by the Radon-Nikodym derivative

$$\begin{aligned} \frac{dP^0}{dP} &= \exp \left\{ - \int_0^T h(\mathcal{X}_s)(dY_s - h(\mathcal{X}_s)ds) - \frac{1}{2} \int_0^T |h(\mathcal{X}_s)|^2 ds \right\} \\ &= \exp \left\{ - \int_0^T h(\mathcal{X}_s)dY_s + \frac{1}{2} \int_0^T |h(\mathcal{X}_s)|^2 ds \right\}, \end{aligned}$$

i.e. the measure under which the process \mathbf{Y} is a Wiener process, independent of the state process \mathcal{X} , then the $\mathcal{P}(S)$ -valued processes $\tilde{\pi}^n = (\tilde{\pi}_t^n)_{t \in [0, T]} = (U^n(t, \mathbf{Y}))_{t \in [0, T]}$ converge in P^0 -probability to the $\mathcal{P}(S)$ -valued process $\pi = (\pi_t)_{t \in [0, T]} = (U(t, \mathbf{Y}))_{t \in [0, T]}$. In addition, since the measure P is also absolutely continuous w.r.t. P^0 , the authors also prove that

$$\tilde{\pi}^n \text{ converges in } P\text{-probability to } \pi$$

which is the result we had already announced for the second approach described above.

Finally in Remark 7.4 in [3] it is observed that if $(\Omega^n, \mathcal{F}^n, P^n) = (\Omega, \mathcal{F}, P)$, $W^n(t) = W(t)$, in (20), **(B1)** and **(B2)** hold, and moreover

(B3') \mathcal{X}^n converges in P -probability to \mathcal{X} in $D_S([0, T])$,

(B4') $\lim_{n \rightarrow \infty} \int_0^T |\mathbf{h}^n(\mathcal{X}_s^n) - \mathbf{h}(\mathcal{X}_s)|^2 ds = 0$, in P -probability,

then the filters π^n converge to π in P -probability, as measure valued processes.

To conclude this section observe that a sufficient condition for **(B4')** to hold is that

$$\mathbf{(B4'')} \quad \lim_{n \rightarrow \infty} E \left(\int_0^T |\mathbf{h}^n(\mathcal{X}_s^n) - \mathbf{h}(\mathcal{X}_s)|^2 ds \right) = 0.$$

4.2 The main results

In this section, we want to prove by using the convergence results proved in [2] (see also Section 4.1) that the sequence of filters associated with the approximated system (7) and (8) introduced in Section 2.2 converges to the filter associated with the original system (1) and (4). With this aim, in our setting we take $S = [0, T] \times C([- \tau, 0], \mathbb{R})$, with

$$dist_S((t, \theta), (t', \theta')) = \|(t, \theta) - (t', \theta')\|_S = |t - t'| + \|\theta - \theta'\|,$$

$$\mathcal{X}_t = (t, \Pi_t X),$$

and

$$\mathbf{h} : [0, T] \times C([- \tau, 0], \mathbb{R}) \rightarrow \mathbb{R}; \quad (t, \theta) \mapsto h(t, \theta)$$

where h is the observation function in (4).

Note that with these choices the observation processes given by (4) and (19) coincide.

Moreover, if we take

$$\mathcal{X}_t^n = (\delta \cdot \lfloor t/\delta \rfloor, \Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n), \quad (21)$$

as an approximation for the signal process \mathcal{X}_t , with $X^n(t)$ defined in Section 2.2, then the approximation (8) of Section 2.2 for the observation process and the approximation (20) of Section 4.1, with $\mathbf{h}^n = \mathbf{h}$ and $W^n = W$, coincide.

With the approximations (14) or (15) for the observation process it is natural to take

$$\mathcal{X}_t^n = (t, \Pi_t X^n), \quad (22)$$

or

$$\mathcal{X}_t^n = (t, \Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n), \quad (23)$$

as an approximation for the signal process.

Indeed, similarly to the previous case, if we take the approximation (22) (respectively (23)) for the signal process, $\mathbf{h}^n = \mathbf{h}$ and $W^n = W$, then the approximation (14) (respectively (15)) and (20) for the observation process coincide.

Remark 4.1. *Note that we are taking $(\Omega^n, \mathcal{F}^n, P^n) = (\Omega, \mathcal{F}, P)$, and $W^n = W$. Moreover, in order to get the approximation (8) for the observation process in (20), besides taking \mathcal{X}^n defined by (21) and $\mathbf{h}^n = \mathbf{h}$, we can take \mathcal{X}^n as in (23) and*

$$\mathbf{h}^n : [0, T] \times C([- \tau, 0], \mathbb{R}) \rightarrow \mathbb{R}; \quad (t, \theta) \mapsto h(\delta \cdot \lfloor t/\delta \rfloor, \theta). \quad (24)$$

Then, the main result of this paper, concerning the approximated system (7) and (8) is stated in the following theorem. Note that similar results hold for the other approximation schemes introduced in Section 2.2 (see for example Lemma 4.7).

Theorem 4.2. *Assume that conditions **(H1)** to **(H10)** are satisfied. Then, the sequence of filters $\boldsymbol{\pi}^n$ associated with the approximated system (7) and (8) converges weakly and in probability to the original filter $\boldsymbol{\pi}$, in $D_{\mathcal{P}(S)}([0, T])$.*

Moreover the sequence of measure valued processes $\tilde{\boldsymbol{\pi}}^n$ converges in probability to the original filter $\boldsymbol{\pi}$.

In addition, if d denotes the Prohorov metric, then $\max_{k=1, \dots, n} d(\tilde{\pi}_{k\delta}^n, \pi_{k\delta})$ converges in probability to zero.

Since the filter π is continuous in time, the last statement of the Theorem is an immediate consequence of the convergence in probability of $\tilde{\pi}^n$ to π .

In the following, since the measure valued processes π^n and $\tilde{\pi}^n$ in Section 4.1 coincide with the processes given by (11) and (12) respectively, we prove that under the assumptions of Theorem 4.2 the conditions **(B1)** to **(B4)** of the previous section are fulfilled.

In our setting, the convergence condition **(B1)** is obviously satisfied when taking $h^n = h$, and the continuity condition **(B2)** on h is satisfied thanks to the continuity assumption **(H8)** on h .

Furthermore, note that, under assumption **(H8)**, condition **(B1)** is obviously also satisfied when we take h^n defined as in (24).

Moreover, in our framework, condition **(B3)** asserts that $\mathcal{X}^n = ((\delta \cdot \lfloor t/\delta \rfloor, \Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n))_{t \geq 0}$ converges weakly to $\mathcal{X} = ((t, \Pi_t X))_{t \geq 0}$ in $D_S([0, T]) = D_{[0, T] \times C([- \tau, 0], \mathbb{R})}([0, T])$. We will prove the convergence of \mathcal{X}^n to \mathcal{X} , in a sense stronger than weak convergence, in Proposition 4.5 below (see also Remark 4.6).

Assume for the moment that the weak convergence condition **(B3)** is satisfied. Then, with our approximation scheme, condition **(B4)** reads

$$\lim_{n \rightarrow \infty} E \left(\int_0^T |h(\delta \cdot \lfloor s/\delta \rfloor, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} X^n)|^2 ds \right) = E \left(\int_0^T |h(s, \Pi_s X)|^2 ds \right). \quad (25)$$

With the other choices for the approximation models condition **(B4)** reads

$$\lim_{n \rightarrow \infty} E \left(\int_0^T |h(s, \Pi_s X^n)|^2 ds \right) = E \left(\int_0^T |h(s, \Pi_s X)|^2 ds \right), \quad (26)$$

or

$$\lim_{n \rightarrow \infty} E \left(\int_0^T |h(s, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} X^n)|^2 ds \right) = E \left(\int_0^T |h(s, \Pi_s X)|^2 ds \right), \quad (27)$$

respectively.

Under the sublinear growth condition **(H7)**, and the continuity condition **(H8)** for the observation function h , if the weak convergence condition **(B3)** holds, we can assume w.l.o.g. that in our case the integrals inside the expectation in the left hand side of (25) converge to the integral inside the expectation in the right hand side of (25). Therefore, to prove that condition **(B4)** holds, we only need a uniform integrability condition. To this end observe that

$$\begin{aligned} & \sup_n E \left[\left(\int_0^T |h(\delta \cdot \lfloor s/\delta \rfloor, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} X^n)|^2 ds \right)^2 \right] \leq \sup_n E \left[T \int_0^T |h(\delta \cdot \lfloor s/\delta \rfloor, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} X^n)|^4 ds \right] \\ & \leq \sup_n T \int_0^T E \left[C' \left(1 + \sup_{t \in [-\tau, T]} |X^n(t)|^4 \right) \right] ds < \infty \end{aligned}$$

and hence, condition **(B4)** will hold thanks to the following result, which implies also the stronger condition **(B4'')**, and therefore the convergence in probability of the filters (see Section 4.1).

Lemma 4.3. *Assume that conditions **(H1)** to **(H4)** are satisfied, then*

$$\sup_n \mathbb{E} \left[\sup_{t \in [0, T]} \|\Pi_t X^n\|^2 \right] = \sup_n \mathbb{E} \left[\sup_{u \in [-\tau, T]} |X^n(u)|^2 \right] < \infty.$$

Moreover, if condition **(H10)** also holds, then

$$\sup_n \mathbb{E} \left[\sup_{t \in [0, T]} \|\Pi_t X^n\|^4 \right] = \sup_n \mathbb{E} \left[\sup_{u \in [-\tau, T]} |X^n(u)|^4 \right] < \infty.$$

Proof. First observe that

$$\begin{aligned} \sup_{t \in [-\tau, T]} |X^n(t)| &= \max(|X^n(k\delta)|; k\delta \leq T) \leq \max(\|\eta\|, \max(|X^n(k\delta)|; 0 \leq k\delta \leq T)) \\ &= \max(\|\eta\|, \max(|Z^n(k\delta)|; 0 \leq k\delta \leq T)) \leq \max \left(\|\eta\|, \sup_{t \in [0, T]} |Z^n(t)| \right), \end{aligned}$$

where $Z^n(t)$ is given by (13), and set

$$M_t^n := \int_0^t b_n(u) d\tilde{W}_u \quad \text{where } b_n(u) := b(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} X^n). \quad (28)$$

For any $\ell \in \{1, 2\}$ take $\alpha = 2\ell$ and $\beta = 2\ell/(2\ell - 1)$, so that $(1/\alpha) + (1/\beta) = 1$. Then for any stopping time σ , there exists a suitable constant C_ℓ such that

$$\begin{aligned} &\sup_{u \in [-\tau, t \wedge \sigma]} |X^n(u)|^{2\ell} \\ &\leq C_\ell \left\{ \|\eta\|^{2\ell} + \left(\int_0^t \sup_{u \in [0, s \wedge \sigma]} |a(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} X^n)| ds \right)^{2\ell} + \sup_{s \in [0, t \wedge \sigma]} |M_s^n|^{2\ell} \right\} \\ &\leq C_\ell \left\{ \|\eta\|^{2\ell} + t^{2\ell/\beta} \left(\int_0^t \sup_{u \in [0, s \wedge \sigma]} |a(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} X^n)|^\alpha ds \right)^{2\ell/\alpha} + \sup_{s \in [0, t \wedge \sigma]} |M_s^n|^{2\ell} \right\} \\ &\leq C_\ell \left\{ \|\eta\|^{2\ell} + t^{2\ell-1} \int_0^t \ell K^\ell \left(1 + \sup_{u \in [-\tau, s \wedge \sigma]} |X^n(u)|^{2\ell} \right) ds + \sup_{s \in [0, t]} |M_{s \wedge \sigma}^n|^{2\ell} \right\}. \end{aligned}$$

If $M_{t \wedge \sigma}^n$ is a martingale, setting

$$\phi_{\sigma, \ell}^n(t) := \mathbb{E} \left[\sup_{u \in [-\tau, t \wedge \sigma]} |X^n(u)|^{2\ell} \right],$$

and applying Doob's inequality for $p = 2\ell$ to $M_{t \wedge \sigma}^n$ yields

$$\phi_{\sigma, \ell}^n(t) \leq C'_\ell \left\{ 1 + \|\eta\|^{2\ell} + \int_0^t \phi_{\sigma, \ell}^n(s) ds + \mathbb{E} \left[|M_{t \wedge \sigma}^n|^{2\ell} \right] \right\}, \quad (29)$$

for all $t \in [0, T]$, for a suitable constant $C'_\ell = C'_\ell(T)$.

Taking $\sigma = \sigma_N^n := \inf\{s > 0; \sup_{u \in [-\tau, s]} |X^n(u)| \geq N\}$, we have

$$\int_0^T \mathbb{E}[\mathbf{1}_{s \leq \sigma_N^n} b_n^{2\ell}(s)] ds < \infty, \quad (30)$$

and $M_{t \wedge \sigma}^n = M_{t \wedge \sigma_N^n}^n$ is a martingale. Indeed, by the sublinearity condition (3) on b we have

$$\int_0^t \mathbb{E}[\mathbf{1}_{s \leq \sigma_N^n} b_n^{2\ell}(s)] ds \leq \int_0^t \mathbb{E}[\ell K^\ell (1 + \sup_{u \in [-\tau, s \wedge \sigma_N^n]} |X^n(u)|^{2\ell})] ds, \quad (31)$$

and the r.h.s. of the previous inequality is finite since $\sup_{u \in [-\tau, s \wedge \sigma_N^n]} |X^n(u)| \leq N$. Then

$$\mathbb{E}[|M_{t \wedge \sigma_N^n}^n|^2] = \mathbb{E} \left[\left(\int_0^t \mathbf{1}_{s \leq \sigma_N^n} b_n(s) d\tilde{W}_s \right)^2 \right] = \int_0^t \mathbb{E}[\mathbf{1}_{s \leq \sigma_N^n} b_n^2(s)] ds,$$

and (see e.g. Lemma 4.12, page 125, in Liptser and Shiryaev [13])²

$$\mathbb{E}[|M_{t \wedge \sigma}^n|^4] = \mathbb{E} \left[\left(\int_0^t \mathbf{1}_{s \leq \sigma_N^n} b_n(s) d\tilde{W}_s \right)^4 \right] \leq 6^2 t \int_0^t \mathbb{E}[\mathbf{1}_{s \leq \sigma_N^n} b_n^4(s)] ds.$$

Then, taking into account (29) and (31), and invoking Gronwall's inequality we get a bound for $\phi_{\sigma, \ell}^n(T) = \phi_{\sigma_N^n, \ell}^n(T)$, uniform in n and N . Therefore, applying Fatou's Lemma and making use of the fact that $\sigma_N^n \rightarrow \infty$ as $N \rightarrow \infty$, we get the theses. \square

Remark 4.4. *Note that with the same technique one could prove that*

$$\sup_n \mathbb{E} \left[\sup_{t \in [0, T]} \|\Pi_t Z^n\|^{2\ell} \right] = \sup_n \mathbb{E} \left[\sup_{u \in [-\tau, T]} |Z^n(u)|^{2\ell} \right] < \infty.$$

for $\ell = 1, 2$, whenever $\mathbb{E}[\|\eta\|^{2\ell}] < \infty$, i.e. when assumption **(H4)** or **(H10)** holds.

²Apply Lemma 4.12, page 125, with

$$f(t, \omega) := \mathbf{1}_{t \leq \sigma_N^n} b_n(t),$$

which assures that, if

$$\int_0^T \mathbb{E}[f(s, \omega)^{2m}] ds < \infty$$

then

$$\mathbb{E} \left[\left(\int_0^t f(s, \omega) dW_s \right)^{2m} \right] \leq [m(2m-1)]^m t^{m-1} \int_0^t \mathbb{E}[f(s, \omega)^{2m}] ds$$

In order to check the weak convergence condition **(B3)** we use the following result.

Proposition 4.5. *Assume that conditions **(H1)** to **(H4)** and **(H9)** are satisfied. Then*

$$\lim_{n \rightarrow \infty} E \left[\sup_{t \in [0, T]} \left\| (\delta \cdot \lfloor t/\delta \rfloor, \Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n) - (t, \Pi_t X) \right\|_S^2 \right] = 0. \quad (32)$$

Remark 4.6. *Note that condition (32) implies the convergence in probability in $D_S([0, T])$, with the uniform metric, which in fact is stronger than **(B3)**.*

As an intermediate step we will need the following result concerning the weak convergence of the approximation $Z^n(t)$.

Lemma 4.7. *Under the hypotheses of Proposition 4.5*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \in [0, T]} \|\Pi_t Z^n - \Pi_t X\|^2 \right] = 0.$$

Before giving the proofs of Proposition 4.5 and Lemma 4.7 we introduce the operator P^δ which gives the linear interpolation of a function $(x(s))_{s \in [-\tau, T]}$, with step δ , so that $(P^\delta x(s))_{s \in [-\tau, T]}$ is the linear interpolation of $(\ell\delta, x(\ell\delta))$, for $\ell = -m, \dots, n$. Moreover, denoting by

$$\omega_x(\delta; [-\tau, T]) := \sup_{\substack{s, t \in [-\tau, T] \\ |s-t| \leq \delta}} |x(s) - x(t)|$$

the modulus of continuity of the function $(x(s))_{s \in [-\tau, T]}$, we observe that

$$\sup_{t \in [0, T]} \|\Pi_{\delta \cdot \lfloor t/\delta \rfloor} P^\delta x - \Pi_t x\| \leq 2\omega_x(\delta; [-\tau, T]), \quad (33)$$

$$\omega_x(\delta; [-\tau, T]) \leq 2 \sup_{t \in [-\tau, T]} |x(t)| \quad (34)$$

The second inequality is obvious, while the first one follows by observing that $\Pi_{\delta \cdot \lfloor t/\delta \rfloor} P^\delta x = P^\delta \Pi_{\delta \cdot \lfloor t/\delta \rfloor} x$, and therefore

$$\begin{aligned} \|\Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta x - \Pi_u x\| &\leq \|\Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta x - \Pi_{\delta \cdot \lfloor u/\delta \rfloor} x\| + \|\Pi_{\delta \cdot \lfloor u/\delta \rfloor} x - \Pi_u x\| \\ &= \|P^\delta \Pi_{\delta \cdot \lfloor u/\delta \rfloor} x - \Pi_{\delta \cdot \lfloor u/\delta \rfloor} x\| + \|\Pi_{\delta \cdot \lfloor u/\delta \rfloor} x - \Pi_u x\| \\ &\leq 2 \sup_{\substack{s, t \in [0, T] \\ |s-t| \leq \delta}} \|\Pi_s x - \Pi_t x\| = 2 \sup_{\substack{s, t \in [-\tau, T] \\ |s-t| \leq \delta}} |x(s) - x(t)|. \end{aligned}$$

Finally, we can apply inequality (33) to the state process. Since the paths of X are

continuous, the modulus of continuity $\omega_X(\delta, [-\tau, T])$ converge to zero as $\delta = \delta_n$ converge to zero, and therefore, taking into account (34) and (5), by the dominated convergence theorem we get that

$$\lim_{n \rightarrow \infty} \mathbb{E} [\omega_X^2(\delta, [-\tau, T])] = 0, \quad (35)$$

and that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \in [0, T]} \|\Pi_{\delta \cdot \lfloor t/\delta \rfloor} P^\delta X - \Pi_t X\|^2 \right] = 0. \quad (36)$$

Proof. of Proposition 4.5

First of all we note that

$$\|(\delta \cdot \lfloor t/\delta \rfloor, \Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n) - (t, \Pi_t X)\|_S^2 \leq 2\delta^2 + 2 \|\Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n - \Pi_t X\|^2,$$

Then adding and subtracting $\Pi_{\delta \cdot \lfloor t/\delta \rfloor} P^\delta X$ in the second term on the right hand side of the above expression it yields

$$\|\Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n - \Pi_t X\|^2 \leq 2 \|\Pi_{\delta \cdot \lfloor t/\delta \rfloor} X^n - \Pi_{\delta \cdot \lfloor t/\delta \rfloor} P^\delta X\|^2 + 2 \|\Pi_{\delta \cdot \lfloor t/\delta \rfloor} P^\delta X - \Pi_t X\|^2.$$

Then, taking into account (36), and that

$$\begin{aligned} \sup_{t \in [0, T]} \|\Pi_t X^n - \Pi_t P^\delta X\| &= \sup_{k: k\delta \in [-\tau, T]} |X^n(k\delta) - X(k\delta)| \\ &= \sup_{k: k\delta \in [-\tau, T]} |Z^n(k\delta) - X(k\delta)| \leq \sup_{t \in [-\tau, T]} |Z^n(t) - X(t)| \end{aligned}$$

the result follows by Lemma 4.7. □

Proof. of Lemma 4.7

Noticing that $P^\delta Z^n(s) = X^n(s)$, for $s \in [-\tau, 0] \cup [0, T]$, then we can rewrite (13) as

$$Z^n(t) = \eta(0) + \int_0^t a(\delta \cdot \lfloor s/\delta \rfloor, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} P^\delta Z^n) ds + \int_0^t b(\delta \cdot \lfloor s/\delta \rfloor, \Pi_{\delta \cdot \lfloor s/\delta \rfloor} P^\delta Z^n) d\tilde{W}_s.$$

Therefore, taking into account that $Z^n(t) = X(t) = \eta(t)$ for $t \in [-\tau, 0]$,

$$\begin{aligned} &\sup_{s \in [-\tau, t]} |Z^n(s) - X(s)|^2 \\ &\leq 2 \left(\int_0^t |a(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n) - a(u, \Pi_u X)| du \right)^2 \\ &\quad + 2 \sup_{s \in [0, t]} \left(\int_0^s [b(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n) - b(u, \Pi_u X)] d\tilde{W}_u \right)^2 \\ &\leq 2t \int_0^t |a(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n) - a(u, \Pi_u X)|^2 du \\ &\quad + 2 \sup_{s \in [0, t]} \left(\int_0^s [b(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n) - b(u, \Pi_u X)] d\tilde{W}_u \right)^2 \end{aligned}$$

By Lemma 4.3, Remark 4.4 and the sublinearity of b ,

$$\int_0^s [b(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n) - b(u, \Pi_u X)] d\tilde{W}_u$$

is a martingale.

Then taking the expectations we can apply Doob's inequality, and we get for $t \in [0, T]$

$$\begin{aligned} & \mathbb{E} \left[\sup_{u \in [0, t]} \|\Pi_u Z^n - \Pi_u X\|^2 \right] = \mathbb{E} \left[\sup_{s \in [-\tau, t]} |Z^n(s) - X(s)|^2 \right] \\ & \leq 2t \int_0^t \mathbb{E} [|a(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n) - a(u, \Pi_u X)|^2] du \\ & + 8 \int_0^t \mathbb{E} \left[|b(\delta \cdot \lfloor u/\delta \rfloor, \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n) - b(u, \Pi_u X)|^2 \right] du \\ & \leq \max(2T, 8) \int_0^t K \mathbb{E} \left[\sup_{u \in [0, s]} \|\Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n - \Pi_u X\|^2 + \delta^2 \right] ds \\ & \leq C(T) \int_0^t \mathbb{E} \left[\sup_{u \in [0, s]} (\|\Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta Z^n - \Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta X\|^2 + \|\Pi_{\delta \cdot \lfloor u/\delta \rfloor} P^\delta X - \Pi_u X\|^2) + \delta^2 \right] ds \\ & \leq C(T) \int_0^t \mathbb{E} \left[\sup_{u \in [0, s]} \|\Pi_u Z^n - \Pi_u X\|^2 + 4\omega_X^2(\delta; [-\tau, T]) + \delta^2 \right] ds, \end{aligned}$$

where we have used (33). Then Gronwall's inequality gives the upper bound

$$\mathbb{E} \left[\sup_{u \in [0, T]} \|\Pi_u Z^n - \Pi_u X\|^2 \right] \leq C'(T) (\mathbb{E} [\omega_X^2(\delta; [-\tau, T])] + \delta^2)$$

and the proof is accomplished, since $\delta = \delta_n$ and (according to (35)) $\mathbb{E} [\omega_X^2(\delta; [-\tau, T])]$ go to zero as n goes to infinity. \square

References

- [1] BHATT, A. G., KALLIANPUR, G., AND KARANDIKAR, R. L. Uniqueness and robustness of solution of measure-valued equations of nonlinear filtering. *Ann. Probab.* 23, 4 (1995), 1895–1938.
- [2] BHATT, A. G., KALLIANPUR, G., AND KARANDIKAR, R. L. Robustness of the nonlinear filter. *Stochastic Process. Appl.* 81, 2 (1999), 247–254.
- [3] BHATT, A. G., AND KARANDIKAR, R. L. Robustness of the nonlinear filter: the correlated case. *Stochastic Process. Appl.* 97, 1 (2002), 41–58.

- [4] CHANG, M. H. Discrete approximation of nonlinear filtering for stochastic delay equations. *Stochastic Anal. Appl.* 5, 3 (1987), 267–298.
- [5] GOGGIN, E. M. Convergence in distribution of conditional expectations. *Ann. Probab.* 22, 2 (1994), 1097–1114.
- [6] KALLIANPUR, G. *Stochastic filtering theory*, vol. 13 of *Applications of Mathematics*. Springer-Verlag, New York, 1980.
- [7] KALLIANPUR, G. MANDAL, P. K. Nonlinear filtering with stochastic delay equations. In *Advances on Theoretical and Methodological Aspects of Probability and Statistics; IISA 1998; Volume 2; Editor: Balakrishnan, N.* Gordon and Breach Science Publishers, 2002, pp. 3–36.
- [8] KUNITA, H. Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal.* 1 (1971), 365–393.
- [9] KURTZ, T. G., AND OCONE, D. L. Unique characterization of conditional distributions in nonlinear filtering. *Ann. Probab.* 16, 1 (1988), 80–107.
- [10] KUSHNER, H. J. *Weak convergence methods and singularly perturbed stochastic control and filtering problems*, vol. 3 of *Systems & Control: Foundations & Applications*. Birkhäuser Boston Inc., Boston, MA, 1990.
- [11] KWONG, R. H., AND WILLSKY, A. S. Estimation and filter stability of stochastic delay systems. *SIAM J. Control Optim.* 16, 4 (1978), 660–681.
- [12] LE GLAND, F. Time discretization of nonlinear filtering equations. In *Proceedings of the 28th IEEE Conference on Decision and Control, Vol. 1–3 (Tampa, FL, 1989)* (New York, 1989), IEEE, pp. 2601–2606.
- [13] LIPTSER, R. S., AND SHIRYAEV, A. N. *Statistics of random processes. I*, expanded ed., vol. 5 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1977. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.
- [14] MOHAMMED, S. E. A. *Stochastic functional differential equations*, vol. 99 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1984.
- [15] PARDOUX, É. Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In *École d’Été de Probabilités de Saint-Flour XIX—1989*, vol. 1464 of *Lecture Notes in Math*. Springer, Berlin, 1991, pp. 67–163.