ELSEVIER

Stochastic Processes and their Applications 97 (2002) 41–58

stochastic processes and their applications

www.elsevier.com/locate/spa

Robustness of the nonlinear filter: the correlated case

Abhay G. Bhatt *, Rajeeva L. Karandikar

Indian Statistical Institute, Delhi Centre, 7, S.J.S. Sansanwal Marg, New Delhi 110016, India

Received 11 October 2000; received in revised form 19 June 2001; accepted 27 June 2001

Abstract

We consider the question of robustness of the optimal nonlinear filter when the signal process X and the observation noise are possibly correlated. The signal X and observations Y are given by a SDE where the coefficients can depend on the entire past. Using results on pathwise solutions of stochastic differential equations we express X as a functional of two independent Brownian motions under the reference probability measure P_0. This allows us to write the filter π as a ratio of two expectations. This is the main step in proving robustness.

In this framework we show that when (X^n, Y^n) converge to (X, Y) in law, then the corresponding filters also converge in law. Moreover, when the signal and observation processes converge in probability, so do the filters.

We also prove that the paths of the filter are continuous in this framework.

MSC: Primary 60G35; Secondary 60H10; 60G17; 60G44

Keywords: Nonlinear filtering; Correlated signal and noise; Robustness; Pathwise formulae for SDE

1. Introduction

We consider a general nonlinear filtering model with $\mathbb{R}^d$ valued signal process X and $\mathbb{R}^k$ valued observation process Y where the observation noise is Gaussian. We consider the case where the observation noise could possibly be correlated with the signal X. Let π denote the optimal nonlinear filter defined by

$$\pi_t(f) = E[f(X_t)|\mathscr{F}_t^Y], \quad f \in C_b(\mathbb{R}^d).$$

Here and in the sequel, for any process η, we will denote by $\mathscr{F}_t^\eta$ the σ-field generated by $\{\eta_s\colon 0 \leqslant s \leqslant t\}$. $C_b(S)$ denotes the space of bounded continuous functions on a metric space S.

We consider approximating processes (X^n, Y^n) converging to (X, Y). Let π^n denote the corresponding nonlinear filter defined by

$$\pi_t^n(f) = E[f(X_t^n)|\mathscr{F}_t^{Y^n}], \quad f \in C_b(\mathbb{R}^d).$$

* Corresponding author. Tel.: +91-11-651-6200; fax: +91-11-685-6779.

E-mail addresses: abhay@isid.ac.in (A.G. Bhatt), rlk@isid.ac.in (R.L. Karandikar).

0304-4149/02/$ - see front matter
PII: S0304-4149(01)00121-1

In this article we address the question of robustness of the filter π: Does π^n converge to π (as $C([0,T],\mathscr{M}_+(\mathbb{R}^d))$ valued processes)? Here, $\mathscr{M}_+(\mathbb{R}^d)$ denotes the space of positive finite measures equipped with the topology of weak convergence. $C([0,T],S)$ is the space of continuous functions from $[0,T]$ to S equipped with the topology of uniform convergence and the corresponding Borel σ-field.

It is well known that in general convergence in probability of (U^n,V^n) to (U,V) does not even guarantee the weak convergence of the conditional expectations $E[U^n|V^n]$ to $E[U|V]$. Goggin (1994) obtained sufficient conditions for convergence of conditional expectations and applied it to deduce weak convergence of π^n to π assuming, among other things, independence of signal and observation noise. In Bhatt et al. (1995) weak convergence of π^n to π was shown in the signal–noise independent case using uniqueness of solution of the (measure valued) Zakai equation. This required the assumption that the signal is Markov. Both these results required stringent integrability conditions to be satisfied.

Similar questions arise when one tries to prove convergence of approximate filter (computed via time discretisation or otherwise) to the optimal filter. See Goggin (1992), Elliott and Glowinski (1989), Florschinger and Le Gland (1991), Budhiraja and Kallianpur (1996).

In Bhatt et al. (1999) robustness of the filter (again in the signal–noise independent case) was deduced directly from the Kallianpur Striebel Bayes' formula under minimal integrability conditions. The technique used was to express the filters π,π^n as

$$\pi(\omega)=H(Y(\omega)),\quad \pi^n(\omega)=H^n(Y^n(\omega))$$

for suitable Wiener functionals H and H^n and then showing that H^n converges to H in probability.

Here, we extend these robustness results to the correlated case (Theorem 6.1). The main hurdle in using this approach in the correlated case is that the usual analogue of the Bayes' formula expresses the filter as a ratio of two quantities each of which is a conditional expectation as opposed to expectation in the independent case. Here we consider the model

$$\mathrm{d}X_t=a(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}W_t^1+b(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}W_t^2+c(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}t,$$

$$\mathrm{d}Y_t=h(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}t+\mathrm{d}W_t^2,$$

$$\mathscr{X}_t(s)=X_{t\wedge s},$$

$$\mathscr{Y}_t(s)=Y_{t\wedge s},$$

where W^1 and W^2 are independent Brownian motions.

The main technique that we employ here is the following. We express X as a solution of the following SDE.

$$\mathrm{d}X_t=a(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}W_t^1+b(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}Y_t$$
$$+(c(t,\mathscr{X}_t,\mathscr{Y}_t)-b(t,\mathscr{X}_t,\mathscr{Y}_t)h(t,\mathscr{X}_t,\mathscr{Y}_t))\,\mathrm{d}t.$$

We assume that the reference probability measure P_0 exists. Now using pathwise solutions of stochastic differential equations, we express X as a functional of the two independent Brownian motions W^1, Y (under P_0). This now facilitates writing π as a ratio of two expectations (instead of conditional expectations) leading to a new Bayes' formula for the filter in this framework. This in turn gives a way of proving robustness of the filter in the correlated case, i.e. when the approximating processes (X^n, Y^n) satisfy equations similar to the ones for (X, Y) given above, then weak convergence of (X^n, Y^n) to (X, Y) implies that π^n converges to π weakly.

We extend the robustness results in another direction. All the robustness results referred to above show weak convergence of the filter. Here, we show that $\pi^n \to \pi$ in probability when $(X^n, Y^n) \to (X, Y)$ in probability (Theorem 7.3). We use the weak convergence of $\pi^n \to \pi$ along with a technical lemma (Lemma 7.2) to conclude convergence in probability of $\pi^n \to \pi$. This technique allows us to avoid the exponential integrability conditions on the observation function h. Here again, the pathwise formula for the stochastic integral plays an important role—it allows us to substitute the path of the *integrator* in a stochastic integral with another process.

We also prove that the paths of the filter are continuous in this framework.

Throughout the article $\Rightarrow$ will denote convergence in law.

2. The filtering model

We will consider the filtering model where the signal process X and the observation process Y are given by the following system of differential equations:

$$\mathrm{d}X_t = a(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}W_t^1 + b(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}W_t^2 + c(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}t, \tag{2.1a}$$

$$\mathrm{d}Y_t = h(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}t + \mathrm{d}W_t^2, \tag{2.1b}$$

$$\mathscr{X}_t(s) = X_{t\wedge s}, \tag{2.1c}$$

$$\mathscr{Y}_t(s) = Y_{t\wedge s} \tag{2.1d}$$

for $0 \leqslant t \leqslant T$, where X, W^1 are the $\mathbb{R}^d$ valued processes, Y, W^2 are the $\mathbb{R}^k$ valued processes and W^1, W^2 are the independent Wiener processes. Further, it is assumed that X_0 is independent of W^1, W^2 and $Y_0 = 0$. All these processes are defined on a complete probability space $(\Omega, \mathscr{F}, P)$. Here $(\mathscr{X}_t)$ and $(\mathscr{Y}_t)$ are, respectively, $C([0,T], \mathbb{R}^d)$ and $C([0,T], \mathbb{R}^k)$ valued *path* processes. Let E be equal to $[0,T] \times C([0,T], \mathbb{R}^d) \times C([0,T], \mathbb{R}^k)$. We will denote by $\mathbb{M}^{d\times k}$, the space of matrices of order $d \times k$. Here, the functions a, b, c and h,

$$a : E \to \mathbb{M}^{d\times d}, \quad b : E \to \mathbb{M}^{d\times k}, \quad c : E \to \mathbb{R}^d, \quad h : E \to \mathbb{R}^k$$

are assumed to be continuous and each satisfying the following condition:

$$|f(t,\xi,\eta) - f(t,\xi',\eta')| \leqslant K\left(\sup_{0\leqslant u\leqslant t} |\xi_u - \xi'_u| + \sup_{0\leqslant u\leqslant t} |\eta_u - \eta'_u|\right)$$
$$\forall \xi, \xi' \in C([0,T], \mathbb{R}^d), \;\; \eta, \eta' \in C([0,T], \mathbb{R}^k). \tag{2.2}$$

Under these conditions, (2.1) admits a unique solution (see Kallianpur, 1980). It is easy to see that the solution (X,Y) also satisfies

$$\begin{aligned} \mathrm{d}X_t = {} & a(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}W_t^1 + b(t,\mathscr{X}_t,\mathscr{Y}_t)\,\mathrm{d}Y_t \\ & + (c(t,\mathscr{X}_t,\mathscr{Y}_t) - b(t,\mathscr{X}_t,\mathscr{Y}_t)h(t,\mathscr{X}_t,\mathscr{Y}_t))\,\mathrm{d}t. \end{aligned} \tag{2.3}$$

Under the additional assumption that bh also satisfies condition (2.2), it follows that (2.3) along with (2.1b)–(2.1d) admit a unique solution.

Let q_t be defined by

$$q_t = \exp\left\{ \int_0^t \sum_{i=1}^k h^i(u,\mathscr{X}_u,\mathscr{Y}_u)\,\mathrm{d}Y_u^i - \frac{1}{2}\int_0^t |h(u,\mathscr{X}_u,\mathscr{Y}_u)|^2\,\mathrm{d}u \right\}. \tag{2.4}$$

We will assume that

$$\frac{\mathrm{d}P_0}{\mathrm{d}P} = q_T^{-1} = \exp\left\{ -\int_0^T \sum_{i=1}^k h^i(u,\mathscr{X}_u,\mathscr{Y}_u)\,\mathrm{d}W_u^{2,i} - \frac{1}{2}\int_0^T |h(u,\mathscr{X}_u,\mathscr{Y}_u)|^2\,\mathrm{d}u \right\} \tag{2.5}$$

defines a probability measure P_0.

Remark 2.1. P_0 defined above is always a probability measure if h is bounded. More generally, when

$$E\exp\left\{ \frac{1}{2}\int_0^T |h(u,\mathscr{X}_u,\mathscr{Y}_u)|^2\,\mathrm{d}u \right\} < \infty$$

P_0 is a probability measure. See Novikov (1972) or Kallianpur (1980).

This probability measure P_0 is called the *reference probability measure*. Under P_0, Y and W^1 are independent Brownian motions, also independent of X_0. Further, as a consequence of Girsanov's theorem, we get

$$P \circ X_0^{-1} = P_0 \circ X_0^{-1} = \pi_0. \tag{2.6}$$

Moreover, the optimal filter π_t admits a representation

$$\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)} \quad \forall f \in C_b(\mathbb{R}^d), \tag{2.7}$$

where

$$\sigma_t(f) = E_{P_0}[f(X_t)q_t|\mathscr{F}_t^Y]. \tag{2.8}$$

This observation is well known when the signal is Markovian (see Elliott (1982, Theorem 18.21)) and the same proof carries over to this case. This can also be verified using Lemma 11.3.3 of Kallianpur (1980).

3. The Bayes' formula

We will express σ_t as a functional on the Wiener space. For this purpose let $\Omega_0 = \mathbb{R}^d$, $\Omega_1 = C([0,T],\mathbb{R}^d)$, $\Omega_2 = C([0,T],\mathbb{R}^k)$ and let $\tilde{\Omega} = \Omega_0 \times \Omega_1 \times \Omega_2$. It was shown in

Theorem 4.3 of Karandikar (1989) that the solution X of the SDE (2.3) can be expressed as a functional $\mathfrak{e}$ of the initial condition X_0 and the underlying driving processes W^1 and Y, i.e. there exists a mapping $\mathfrak{e}:\tilde{\Omega}\to\Omega_1$ such that

$$X_t(\omega)=\mathfrak{e}(X_0(\omega),\mathscr{W}^1(\omega),\mathscr{Y}(\omega))(t)\quad\forall t,\ \text{a.s. } P_0. \tag{3.1}$$

(See also the discussion on pathwise formulae in stochastic calculus in Karandikar, 1995.) Here $\mathscr{W}^1$ is the path process of W^1. It should be noted that the mapping $\mathfrak{e}$ depends on the coefficients a, b, c and h appearing in (2.3) and does not depend on the underlying measure.

Let $\mathfrak{I}: C([0,T],\mathbb{R})\times C([0,T],\mathbb{R})\to C([0,T],\mathbb{R})$ denote the *pathwise integral map* so that for continuous semimartingales U,V

$$\mathfrak{I}_t(U,V)=\int_0^t U_s\,\mathrm{d}V_s\quad\forall t \text{ a.s.}$$

See Karandikar (1995). Define $\tilde{q}$ on $\tilde{\Omega}$ by

$$\tilde{q}_t(\tilde{\omega})=\exp\left\{\sum_{i=1}^k\mathfrak{I}_t\left(h^i(\cdot,\mathfrak{e}(\omega_0,\omega_1,\omega_2),\omega_2),\omega_2^i\right)\right.$$
$$\left.-\frac{1}{2}\int_0^t|h(u,\mathfrak{e}(\omega_0;\omega_1,\omega_2),\omega_2)|^2\,\mathrm{d}u\right\}, \tag{3.2}$$

where $\tilde{\omega}=(\omega_0,\omega_1,\omega_2)$. Then,

$$\tilde{q}_t(X_0(\omega),\mathscr{W}^1(\omega),\mathscr{Y}(\omega))=q_t(\omega)\quad\text{a.s. }[P_0]. \tag{3.3}$$

Let Q_1,Q_2 be the Wiener measures on Ω_1 and Ω_2, respectively, and let $\tilde{Q}$ on $\tilde{\Omega}$ be defined by

$$\tilde{Q}=\pi_0\times Q_1\times Q_2.$$

Define

$$\tilde{\sigma}_t(f,\omega_2)=\iint f(\mathfrak{e}(\omega_0,\omega_1,\omega_2)(t))\tilde{q}_t(\omega_0,\omega_1,\omega_2)\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1). \tag{3.4}$$

Note that $\tilde{\sigma}(f,\cdot)=E_{\tilde{Q}}[f(\mathfrak{e}(t))\tilde{q}_t|\mathscr{H}_t]$, where $\mathscr{H}_t$ is the σ-field generated by $\{\omega_2(u): 0\leqslant u\leqslant t\}$. Now using the fact that $P_0\circ(X_0,W^1,Y)^{-1}=\tilde{Q}$, and Eqs. (2.6), (2.8), (3.1) and (3.3), we get

$$\tilde{\sigma}_t(f,Y(\omega))=\sigma_t(f)(\omega)\quad\text{a.s. }P_0. \tag{3.5}$$

Now defining $\tilde{\pi}$ by

$$\tilde{\pi}_t(f,\omega_2)=\frac{\tilde{\sigma}_t(f,\omega_2)}{\tilde{\sigma}_t(\mathbf{1},\omega_2)}, \tag{3.6}$$

it follows that

$$\tilde{\pi}_t(f,Y(\omega))=\pi_t(f)(\omega)\quad\text{a.s. }P. \tag{3.7}$$

Eqs. (3.5) and (3.7) express the unnormalized filter and the optimal filter, respectively, as Wiener functionals of the observation path Y. These are analogues of similar representations obtained in the signal–noise independent case in Bhatt et al. (1995, 1999).

4. Continuity of the filter

In this section we will show that the filter $\tilde{\pi}_t$ is a continuous process. The argument is similar to that in signal–noise independent case given in Bhatt and Karandikar (1999b). Note that by its definition, $\tilde{q}_t$ is continuous in t for every $\tilde{\omega} \in \tilde{\Omega}$ and further is a mean one $\tilde{Q}$-martingale. Let

$$\tilde{\rho}_t(\omega_2) = \int\int \tilde{q}_t(\omega_0, \omega_1, \omega_2)\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1).$$

It follows that $(\tilde{\rho}_t, \mathscr{G}_t)$ is a Q_2-martingale, where $(\mathscr{G}_t)$ is the canonical filtration on $(\Omega_2, \mathscr{B}(\Omega_2), Q_2)$ (satisfying usual hypotheses). Since Q_2 is the Wiener measure on Ω_2, it follows that $\tilde{\rho}_t$ admits a continuous modification denoted by ρ^*_t.

Let

$$N = \{\omega_2 \in \Omega_2\colon\ \tilde{\rho}_r(\omega_2) \neq \rho^*_r(\omega_2) \text{ for some rational } r\}.$$

Then it follows that $Q_2(N) = 0$. Note that

$$\tilde{\rho}^n_t = \int\int (n \wedge \tilde{q}_t(\omega_0, \omega_1, \omega_2))\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1), \quad 0 \leqslant t \leqslant T$$

is a continuous process and hence $(\mathscr{G}_t)$-predictable. Further, $\tilde{\rho}_t$ is the pointwise limit of $\tilde{\rho}^n_t$ as n tends to ∞ and hence $\tilde{\rho}$ is also $(\mathscr{G}_t)$-predictable.

Fix a $(\mathscr{G}_t)$-stopping time τ. Let $\tau^n(\omega_2) = 2^{-n}[2^n\tau(\omega_2) + 1]$ (here, $[x]$ denotes the integer part of x). Note that $\tau^n(\omega_2)$ is rational and hence for $\omega_2 \notin N$,

$$\tilde{\rho}_{\tau^n(\omega_2)}(\omega_2) = \rho^*_{\tau^n(\omega_2)}(\omega_2). \tag{4.1}$$

Fix $\omega_2 \notin N$. Using (4.1) and Fatou's lemma we conclude that

$$\begin{aligned}
\rho^*_{\tau(\omega_2)}(\omega_2) &= \lim_n \rho^*_{\tau^n(\omega_2)}(\omega_2)\\
&= \lim_n \tilde{\rho}_{\tau^n(\omega_2)}(\omega_2)\\
&= \liminf_n \int\int \tilde{q}_{\tau^n(\omega_2)}(\omega_0, \omega_1, \omega_2)\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1)\\
&\geqslant \int\int \liminf_n \tilde{q}_{\tau^n(\omega_2)}(\omega_0, \omega_1, \omega_2)\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1)\\
&= \int\int \tilde{q}_{\tau(\omega_2)}(\omega_0, \omega_1, \omega_2)\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1)\\
&= \tilde{\rho}_{\tau(\omega_2)}(\omega_2).
\end{aligned}$$

Thus, $\rho^*_\tau \geqslant \tilde{\rho}_\tau$ a.s. $[Q_2]$. By Fubini and the definition of $\tilde{\rho}$ it follows that $E_{Q_2}[\tilde{\rho}_\tau] = 1$. Also, (ρ^*_t) is a mean one continuous martingale and hence $E_{Q_2}[\rho^*_\tau] = 1$. These observations give us

$$Q_2(\tilde{\rho}_\tau = \rho^*_\tau) = 1 \quad \text{for all stopping times } \tau. \tag{4.2}$$

Since $\tilde{\rho}$ and ρ^* are predictable processes, (4.2) implies that

$$Q_2(\tilde{\rho}_t = \rho^*_t \text{ for all } t) = 1 \tag{4.3}$$

(see Métivier, 1982, Theorem 14.2). Thus $\tilde{\rho}_t$ is continuous a.s. Q_2. Noting that $Q_2(\tilde{\rho}_T > 0) = 1$, $\tilde{\rho}_t$ is a Q_2 martingale, and that $\tilde{\rho}_t(\omega_2) = \tilde{\sigma}_t(\mathbf{1}, \omega_2)$ we get

$$Q_2\left(\omega_2\colon \inf_{0\leqslant t\leqslant T} \tilde{\sigma}_t(\mathbf{1},\omega_2) > 0\right) = 1. \tag{4.4}$$

(See Ethier and Kurtz, 1986, Proposition II.2.15.) It follows that

$$\frac{1}{\tilde{\sigma}_t(\mathbf{1},\cdot)} \quad \text{is continuous in } t \text{ a.s. } Q_2. \tag{4.5}$$

This discussion leads us to the following result.

Theorem 4.1. *The paths of the processes* $(\tilde{\sigma}_t)$ *and* $(\tilde{\pi}_t)$ *are* Q_2*-a.s. continuous.*

Proof. Let $N_1 = \{\omega_2\colon \tilde{\rho}_t(\omega_2)$ is not continuous in $t\}$. As seen earlier, $Q_2(N_1) = 0$. Fix $t_n \to t$.

Also, for $\omega_2 \notin N_1$,

$$\tilde{q}_{t_n}(\omega_0,\omega_1,\omega_2) \to \tilde{q}_t(\omega_0,\omega_1,\omega_2) \quad \forall \omega_0,\omega_1$$

and

$$\int\int \tilde{q}_{t_n}(\omega_0,\omega_1,\omega_2)\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1) \to \int\int \tilde{q}_t(\omega_0,\omega_1,\omega_2)\,\mathrm{d}\pi_0(\omega_0)\,\mathrm{d}Q_1(\omega_1).$$

Thus, for $\omega_2 \notin N_1$, $\{q_{t_n}(\cdot,\cdot,\omega_2)\colon n \geqslant 1\}$ is $\pi_0 \otimes Q_1$-uniformly integrable. Since $f(\mathfrak{e}(\omega_0,\omega_1,\omega_2)(t))$ is bounded and is continuous in t for all $(\omega_0,\omega_1,\omega_2)$, it now follows from (3.4) that $\tilde{\sigma}_{t_n}(f,\omega_2) \to \tilde{\sigma}_t(f,\omega_2)$ for all $\omega_2 \notin N_1$. Almost sure continuity of $\tilde{\sigma}$ is now immediate.

Continuity of $\tilde{\pi}$ follows from this and Eqs. (3.6), (4.5). □

Remark 4.1. It should be noted that we have not explicitly used the continuity of the signal process in the above arguments. Similar arguments would yield continuity of the filter even when (2.1a) has a jump component.

5. Approximating the filtering model

Let a,b,c,h be as in the previous section. We will assume that these satisfy (2.2) and (X,Y) defined on $(\Omega,\mathscr{F},P)$ satisfy (2.1). We will now consider processes (X^n,Y^n) which approximate (X,Y).

Let $K < \infty$ be fixed. Let a^n, b^n, c^n and h^n be continuous functions

$$a^n : E \to \mathbb{M}^{d\times d}, \quad b^n : E \to \mathbb{M}^{d\times k}, \quad c^n : E \to \mathbb{R}^d, \quad h^n : E \to \mathbb{R}^k$$

each satisfying condition (2.2) (with the same fixed K). As before we also assume that the product function $b^n h^n$ also satisfies condition (2.2).

Let X^n, Y^n be solutions to the system of equations

$$\mathrm{d}X_t^n = a^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n)\,\mathrm{d}W_t^{n,1} + b^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n)\,\mathrm{d}W_t^{n,2} + c^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n)\,\mathrm{d}t, \tag{5.1a}$$

$$\mathrm{d}Y_t^n = h^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n)\,\mathrm{d}t + \mathrm{d}W_t^{n,2}, \tag{5.1b}$$

$$\mathscr{X}_t^n(s) = X_{t\wedge s}^n, \tag{5.1c}$$

$$\mathscr{Y}_t^n(s) = Y_{t\wedge s}^n \tag{5.1d}$$

for $0 \leqslant t \leqslant T$, where the processes $X^n, Y^n, W^{n,1}, W^{n,2}$ are defined on some complete probability space $(\Omega^n, \mathscr{F}^n, P^n)$, $X^n, W^{n,1}$ are $\mathbb{R}^d$ valued processes, $Y^n, W^{n,2}$ are $\mathbb{R}^k$ valued processes and $W^{n,1}, W^{n,2}$ are independent Wiener processes. The processes $\mathscr{X}^n$ and $\mathscr{Y}^n$ are $C([0,T],\mathbb{R}^d)$ and $C([0,T],\mathbb{R}^k)$ valued, respectively. Further, it is assumed that X_0^n is independent of $W^{n,1}, W^{n,2}$ with $P^n \circ (X_0^n)^{-1} = \pi_0^n$ and $Y_0^n = 0$.

We will assume that

$$\pi_0^n \Rightarrow \pi_0 \tag{5.2}$$

and for $\xi \in C([0,T],\mathbb{R}^d)$ and $\eta \in C([0,T],\mathbb{R}^k)$

$$\begin{aligned}
&\sup_{0\leqslant t\leqslant T} |a^n(t,\xi,\eta) - a(t,\xi,\eta)| \to 0,\\
&\sup_{0\leqslant t\leqslant T} |b^n(t,\xi,\eta) - b(t,\xi,\eta)| \to 0,\\
&\sup_{0\leqslant t\leqslant T} |c^n(t,\xi,\eta) - c(t,\xi,\eta)| \to 0,\\
&\sup_{0\leqslant t\leqslant T} |h^n(t,\xi,\eta) - h(t,\xi,\eta)| \to 0.
\end{aligned} \tag{5.3}$$

Even in the case when the processes are Markovian the above condition is weaker than uniform convergence of coefficients on compacts. Here we require that the coefficients converge uniformly in t for every fixed ξ, η.

Let q^n be defined as in (2.4) with $h, \mathscr{X}, \mathscr{Y}$ replaced by $h^n, \mathscr{X}^n, \mathscr{Y}^n$. Define P_0^n by

$$\frac{\mathrm{d}P_0^n}{\mathrm{d}P^n} = (q_T^n)^{-1}. \tag{5.4}$$

We assume that P_0^n is a probability measure. Again, $P_0^n \circ (X_0^n)^{-1} = \pi_0^n$ and under $P_0^n, Y^n, W^{n,1}$ are Brownian motions and $X_0^n, Y^n, W^{n,1}$ are independent. Moreover, X^n, Y^n also satisfy

$$\begin{aligned}
\mathrm{d}X_t^n &= a^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n)\,\mathrm{d}W_t^{n,1} + b^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n)\,\mathrm{d}Y_t^n\\
&\quad + (c^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n) - b^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n)h^n(t,\mathscr{X}_t^n,\mathscr{Y}_t^n))\,\mathrm{d}t.
\end{aligned} \tag{5.5}$$

Here again the SDE, along with (5.1b)–(5.1d) admits a unique strong (and weak) solution. Let $\mathfrak{e}^n$ be the pathwise solution map of this SDE so that

$$X^n = \mathfrak{e}^n(X_0^n, W^{n,1}, Y^n) \quad \text{a.s. } P_0^n. \tag{5.6}$$

Analogous to (3.2) define $\tilde{q}^n$ on $\tilde{\Omega}$ by

$$\begin{aligned}
\tilde{q}_t^n(\tilde{\omega}) = \exp\Bigg\{&\sum_{i=1}^k \mathfrak{I}_t(h^{n,i}(\cdot, \mathfrak{e}^n(\omega_0,\omega_1,\omega_2),\omega_2),\omega_2^i)\\
&- \frac{1}{2}\int_0^t |h^n(u,\mathfrak{e}^n(\omega_0,\omega_1,\omega_2),\omega_2)|^2\,\mathrm{d}u\Bigg\}.
\end{aligned} \tag{5.7}$$

Then,

$$q_t^n(\omega) = \tilde{q}_t^n(X_0^n(\omega), \mathscr{W}^{n,1}(\omega), \mathscr{Y}^n(\omega)) \quad \text{a.s. } [P_0^n]. \tag{5.8}$$

Theorem 5.1. (a) $P_0^n \circ (X^n, Y^n)^{-1} \Rightarrow P^0 \circ (X, Y)^{-1}$.
(b) $P^n \circ (X^n, Y^n)^{-1} \Rightarrow P \circ (X, Y)^{-1}$.

Proof. If all the processes were defined on the same space and if the convergence of X_0^n to X_0 was in probability, then the Theorem would follow from the results on stability of solutions of SDE given in Emery (1979) or Karandikar (1989, Theorem 3.3). The situation at hand is handled via Skorokhod representation.

For part (a), using Skorokhod representation theorem we get a sequence of random variables $\tilde{X}_0^n$ and $\tilde{X}_0$ on a probability space $(\Gamma, \mathscr{G}, R)$ such that $\tilde{X}_0^n \to \tilde{X}_0$ a.s. $[R]$, and $R \circ (\tilde{X}_0^n)^{-1} = \pi_0^n$ and $R \circ (\tilde{X}_0)^{-1} = \pi_0$.

Recall the definitions of Ω_1, Ω_2 and Q_1, Q_2 from the previous section. Let β, Z be the coordinate processes on Ω_1, Ω_2, respectively. Let $\tilde{X}, \tilde{X}^n$ be defined on $(\Gamma, \mathscr{G}, R) \otimes (\Omega_1, \mathscr{F}^\beta, Q_1) \otimes (\Omega_2, \mathscr{F}^Z, Q_2)$ by

$$\tilde{X}_t(\gamma, \omega_1, \omega_2) = \mathfrak{e}(\tilde{X}_0(\gamma), \omega_1, \omega_2)(t), \tag{5.9}$$

$$\tilde{X}_t^n(\gamma, \omega_1, \omega_2) = \mathfrak{e}^n(\tilde{X}_0^n(\gamma), \omega_1, \omega_2)(t). \tag{5.10}$$

Then $\tilde{X}, \tilde{X}^n$ are solutions of the Eqs. (2.3) and (5.5), respectively, with the driving processes β, Z. The stability results referred above now imply

$$\sup_{0 \leqslant t \leqslant T} |\tilde{X}_t^n - \tilde{X}_t| \to 0 \quad \text{in } R \otimes Q_1 \otimes Q_2 \text{ probability}. \tag{5.11}$$

The result (a) follows from this as the law of $(\tilde{X}^n, Z)$ under $R \otimes Q_1 \otimes Q_2$ is the same as the law of (X^n, Y^n) under P_0^n and law of $(\tilde{X}, Z)$ under $R \otimes Q_1 \otimes Q_2$ is the same as the law of (X, Y) under P_0.

For part (b) instead of considering the pathwise solutions of Eqs. (2.3) and (5.5), we look at pathwise solutions to (2.1) and (5.1), respectively. The rest of the argument is same as in part (a). □

6. Robustness of the filter

We continue to use the notations introduced in the previous sections. We start by noting that the functional $\tilde{\sigma}$ defined by (3.4) can also be expressed as

$$\tilde{\sigma}_t(f, \omega_2) = \iint f(\tilde{X}_t(\gamma, \omega_1, \omega_2)) p_t(\gamma, \omega_1, \omega_2) \, \mathrm{d}R(\gamma) \, \mathrm{d}Q_1(\omega_1), \tag{6.1}$$

where the probability measure R is as in the proof of Theorem 5.1 and

$$p_t(\gamma, \omega_1, \omega_2) = \tilde{q}_t(\tilde{X}_0(\gamma), \omega_1, \omega_2). \tag{6.2}$$

Note that

$$p_t = \exp\left\{ \int_0^t \sum_{i=1}^k h^i(u, \tilde{\mathscr{X}}, \mathscr{Z}) \, \mathrm{d}Z_u^i - \frac{1}{2} \int_0^t |h(u, \tilde{\mathscr{X}}, \mathscr{Z})|^2 \, \mathrm{d}u \right\}. \tag{6.3}$$

Similarly, we define p^n and $\tilde{\sigma}^n$ by

$$p_t^n(\gamma, \omega_1, \omega_2) = \tilde{q}_t^n(\tilde{X}_0^n(\gamma), \omega_1, \omega_2) \tag{6.4}$$

and

$$\tilde{\sigma}_t^n(f,\omega_2) = \iint f(\tilde{X}_t^n(\gamma,\omega_1,\omega_2))\, p_t^n(\gamma,\omega_1,\omega_2)\, \mathrm{d}R(\gamma)\, \mathrm{d}Q_1(\omega_1). \tag{6.5}$$

Once again note that

$$p_t^n = \exp\left\{ \int_0^t \sum_{i=1}^k h^{n,i}(u,\tilde{\mathscr{X}}^n,\mathscr{Z})\, \mathrm{d}Z_u^i - \frac{1}{2}\int_0^t |h^n(u,\tilde{\mathscr{X}}^n,\mathscr{Z})|^2\, \mathrm{d}u \right\}. \tag{6.6}$$

Let $\tilde{\pi}^n$ be defined by

$$\tilde{\pi}_t^n(f,\omega_2) = \frac{\tilde{\sigma}_t^n(f,\omega_2)}{\tilde{\sigma}_t^n(\mathbf{1},\omega_2)}. \tag{6.7}$$

Define σ^n and π^n by

$$\sigma_t^n(f)(\omega^n) = \tilde{\sigma}_t^n(f, Y^n(\omega^n)), \tag{6.8}$$

$$\pi_t^n(f)(\omega^n) = \tilde{\pi}_t^n(f, Y^n(\omega^n)). \tag{6.9}$$

As in (3.7) we get

$$\pi_t^n(f) = E_{P^n}[f(X_t^n)|\mathscr{F}_t^{Y^n}]. \tag{6.10}$$

Now we state the main theorem of this article. We restate all the assumptions explicitly.

Theorem 6.1. *Let* (X,Y), (X^n,Y^n) *be solutions of SDE's* (2.1) *and* (5.1), *respectively. Let the coefficients* $a,b,c,h,bh,a^n,b^n,c^n,h^n,b^nh^n$ *satisfy condition* (2.2) *for a fixed constant* K. *Assume that* P_0 *and* P_0^n *defined by* (2.5) *and* (5.4), *respectively, are probability measures. Further let* (5.2) *and* (5.3) *be satisfied. Then,*
(a) $\sup_{0\leqslant t\leqslant T} |\tilde{\sigma}_t^n(f,\omega_2) - \tilde{\sigma}_t(f,\omega_2)| \to 0$ *in* Q_2 *probability.*
(b) $\sup_{0\leqslant t\leqslant T} |\tilde{\pi}_t^n(f,\omega_2) - \tilde{\pi}_t(f,\omega_2)| \to 0$ *in* Q_2 *probability.*
(c) $P^n \circ (\sigma^n)^{-1} \Rightarrow P\circ\sigma^{-1}$.
(d) $P^n \circ (\pi^n)^{-1} \Rightarrow P\circ\pi^{-1}$.

Proof. (a) We will first show that for $t_n \to t$, $p_{t_n}^n \to p_t$ in $L^1(R\otimes Q_1\otimes Q_2)$, where p and p^n are defined by (6.3) and (6.6), respectively. Note that

$$\begin{aligned} &\sup_{0\leqslant t\leqslant T} |h^n(t,\tilde{\mathscr{X}}^n,\mathscr{Z}) - h(t,\tilde{\mathscr{X}},\mathscr{Z})| \\ &\quad\leqslant \sup_{0\leqslant t\leqslant T} |h^n(t,\tilde{\mathscr{X}}^n,\mathscr{Z}) - h^n(t,\tilde{\mathscr{X}},\mathscr{Z})| \\ &\qquad + \sup_{0\leqslant t\leqslant T} |h^n(t,\tilde{\mathscr{X}},\mathscr{Z}) - h(t,\tilde{\mathscr{X}},\mathscr{Z})| \\ &\quad\to 0 \quad \text{as } n\to\infty \text{ in } R\otimes Q_1\otimes Q_2 \text{ probability.} \end{aligned} \tag{6.11}$$

We have used (5.9)–(5.11) to get that the first term on the RHS of the above inequality tends to zero, since for every n, the function h^n satisfies the Lipschitz condition (2.2) with the same fixed K. The second term tends to zero by (5.3).

It now follows that

$$\sup_{0\leqslant t\leqslant T}\left| \int_0^t h^{n,i}(u,\tilde{\mathscr{X}}^n,\mathscr{Z})\, \mathrm{d}Z_u^i - \int_0^t h^i(u,\tilde{\mathscr{X}},\mathscr{Z})\, \mathrm{d}Z_u^i \right| \to 0 \quad \text{as } n\to\infty$$

in $R\otimes Q_1\otimes Q_2$ probability for $1\leqslant i\leqslant k$ and

$$\int_0^T |h^n(u,\tilde{\mathscr{X}}^n,\mathscr{Z})-h(u,\tilde{\mathscr{X}},\mathscr{Z})|^2\,\mathrm{d}u\to 0 \quad \text{as} \quad n\to\infty \tag{6.12}$$

in $R\otimes Q_1\otimes Q_2$ probability. As a consequence we get that

$$p^n_{t_n}\to p_t \quad \text{in } R\otimes Q_1\otimes Q_2 \text{ probability.}$$

Since

$$\iiint p^n_{t_n}\,\mathrm{d}R\,\mathrm{d}Q_1\,\mathrm{d}Q_2=\iiint p_t\,\mathrm{d}R\,\mathrm{d}Q_1\,\mathrm{d}Q_2=1$$

for all n, we get

$$p^n_{t_n}\to p_t \quad \text{in } L^1(R\otimes Q_1\otimes Q_2). \tag{6.13}$$

Since $f(\tilde{X}^n_{t_n})$ is bounded and converges to $f(\tilde{X}_t)$ in $R\otimes Q_1\otimes Q_2$ probability it follows that

$$\iiint |f(\tilde{X}^n_{t_n})p^n_{t_n}-f(\tilde{X}_t)p_t|\,\mathrm{d}R\,\mathrm{d}Q_1\,\mathrm{d}Q_2\to 0.$$

Invoking Fubini's theorem this gives

$$\iint |f(\tilde{X}^n_{t_n})p^n_{t_n}-f(\tilde{X}_t)p_t|\,\mathrm{d}R\,\mathrm{d}Q_1\to 0 \quad \text{in } Q_2 \text{ probability.}$$

Thus, we get

$$|\tilde{\sigma}^n_{t_n}(f,\omega_2)-\tilde{\sigma}_t(f,\omega_2)|\to 0 \quad \text{in } Q_2 \text{ probability.} \tag{6.14}$$

This now implies (a).

(b) Note that as in (4.4) we have $\inf_{0\leqslant t\leqslant T}\tilde{\sigma}^n_t(\mathbf{1},\omega_2)>0$ a.s. $[Q_2]$ for all n. Part (b) now follows from part (a), (3.6), (6.7) and (4.4).

(c) Note that for $G\in C_b(C[0,T],\mathscr{M}_+(\mathbb{R}^d))$,

$$\begin{aligned} E_{P^n}[G(\sigma^n)] &= E_{P^n_0}[G(\sigma^n)q^n_T] \\ &= E_{R\otimes Q_1\otimes Q_2}[G(\tilde{\sigma}^n)p^n_T]. \end{aligned}$$

Similarly,

$$\begin{aligned} E_P[G(\sigma)] &= E_{P_0}[G(\sigma)q_T] \\ &= E_{R\otimes Q_1\otimes Q_2}[G(\tilde{\sigma})p_T]. \end{aligned}$$

The result now follows from (a) and (6.13).

Part (d) follows similarly using (b). $\square$

Remark 6.1. Instead of assuming (5.3) we can get the same conclusions as in Theorem 6.1 if we assume the weaker condition

$$P^n\circ\left(X^n,\int_0^\cdot h^n(u,\mathscr{X}^n,\mathscr{Y}^n)\,\mathrm{d}u,\int_0^T |h^n(u,\mathscr{X}^n,\mathscr{Y}^n)|^2\,\mathrm{d}u\right)^{-1}$$
$$\Rightarrow P\circ\left(X,\int_0^\cdot h(u,\mathscr{X},\mathscr{Y})\,\mathrm{d}u,\int_0^T |h(u,\mathscr{X},\mathscr{Y})|^2\,\mathrm{d}u\right)^{-1}.$$

This can be seen as follows. In the proof of Theorem 6.1 the convergence of the coefficients was used to prove that the expression in (6.12) converges to 0 in probability, where the processes $\tilde{X}^n$ and $\tilde{X}$ are versions of X^n and X, respectively, defined on some appropriate representation space. Here, by choosing another appropriate representation space and proceeding as in Lemma 3.1 of Bhatt et al. (1999) we can show that the expression corresponding to (6.12) goes to zero in probability. It may be noted that the Lemma referred to above proved the same result in the context of signal–noise independent case. The remainder of the proof is similar to that of Theorem 6.1.

Remark 6.2. Note that we have deduced robustness of the filter without any reference to the FKK (or Zakai) equation though it is known that π satisfies the FKK equation. Indeed, at this level of generality, uniqueness of solution to the FKK equation may not hold.

When the dependence of the coefficients a, b, c, h on X is Markovian, i.e. $a(t, \xi, \eta) = \hat{a}(t, \xi_t, \eta)$ for a suitable $\hat{a}$ (and similar conditions on b, c, h), uniqueness of solution to the Zakai and FKK equations was proved in Bhatt and Karandikar, 1999a. No continuity assumptions on the coefficients are required—only requirement being that system of Eqs. (2.1) admits a unique weak solution.

Remark 6.3. Here we have stated the results for finite-dimensional signal and noise. However, the methods used can be easily carried over to infinite-dimensional setting.

7. Convergence in probability of the filter

In the previous section we considered the question of robustness of the filter under a fairly general framework. In this section we will further assume that the approximating processes are all defined on the same space and will show that in this setup the filters will converge in probability.

We need to use the Emery topology on the space of semimartingales which is given by the following metric d. (See Emery, 1979). For a semimartingale Z, define

$$r(Z) = \sum_{n=1}^{\infty} 2^{-n} E\left\{1 \wedge \left(\sup_{0 \leqslant t \leqslant n} |Z_t|\right)\right\}$$

and for semimartingales Z_1, Z_2,

$$d(Z_1, Z_2) = \sup\left\{r\left(\int f d(Z_1 - Z_2)\right) : f \text{ predictable and unformly bounded by } 1\right\}.$$

Let X and Y satisfy the stochastic differential equations

$$\mathrm{d}X_t = a(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}W_t^1 + b(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}W_t^2 + c(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}t,$$

$$\mathrm{d}Y_t = h(t, \mathscr{X}_t, \mathscr{Y}_t)\,\mathrm{d}t + \mathrm{d}W_t^2,$$

$$\mathscr{X}_t(s) = X_{t\wedge s},$$

$$\mathscr{Y}_t(s) = Y_{t\wedge s} \tag{7.1}$$

for $0 \leqslant t \leqslant T$, where X, W^1 are the $\mathbb{R}^d$ valued processes, Y, W^2 are the $\mathbb{R}^k$ valued processes and W^1, W^2 are the independent Wiener processes. Let X_0 be independent of W^1, W^2 and $Y_0 = 0$. All these processes are defined on a complete probability space $(\Omega, \mathscr{F}, P)$. As before the functions a, b, c, h and bh will be assumed to satisfy condition (2.2). Further, let X^n, Y^n satisfy the equations

$$\mathrm{d}X_t^n = a^n(t, \mathscr{X}_t^n, \mathscr{Y}_t^n)\,\mathrm{d}W_t^1 + b^n(t, \mathscr{X}_t^n, \mathscr{Y}_t^n)\,\mathrm{d}W_t^2 + c^n(t, \mathscr{X}_t^n, \mathscr{Y}_t^n)\,\mathrm{d}t,$$

$$\mathrm{d}Y_t^n = h^n(t, \mathscr{X}_t^n, \mathscr{Y}_t^n)\,\mathrm{d}t + \mathrm{d}W_t^2,$$

$$\mathscr{X}_t^n(s) = X_{t\wedge s}^n,$$

$$\mathscr{Y}_t^n(s) = Y_{t\wedge s}^n \tag{7.2}$$

for $0 \leqslant t \leqslant T$, where the W^1, W^2 are as in (7.1). We will assume that for all n, X_0^n is independent of W^1, W^2 with $P \circ (X_0^n)^{-1} = \pi_0^n$ and $Y_0^n = 0$. We will continue to assume that the coefficient functions a^n, b^n, c^n, h^n and $b^n h^n$ satisfy (2.2) and (5.3). We will now assume

$$X_0^n \to X_0 \quad \text{in } P\text{-probability}. \tag{7.3}$$

Under these conditions we have the following result. (See Emery (1979) or Karandikar (1989, Theorem 3.3).)

Theorem 7.1. $(X^n, Y^n) \to (X, Y)$ *in the Emery topology on the space of semimartingales.*

We will need to use the following consequences of convergence in Emery topology. Whenever f^n and f are predictable, locally bounded processes with

$$P\left(\sup_{0\leqslant t\leqslant T} |f_t^n - f_t| > \varepsilon\right) \to 0 \quad \forall \varepsilon > 0, \tag{7.4}$$

then we have

$$P\left(\sup_{0\leqslant t\leqslant T} \left|\int_0^t f^n\,\mathrm{d}Y^n - \int_0^t f\,\mathrm{d}Y\right| > \varepsilon\right) \to 0 \quad \forall \varepsilon > 0. \tag{7.5}$$

(Similar statement holds for X^n and X.) In particular, we have for all $\varepsilon > 0$,

$$P\left(\sup_{0\leqslant t\leqslant T} |Y_t^n - Y_t| > \varepsilon\right) \to 0, \tag{7.6}$$

$$P\left(\sup_{0\leqslant t\leqslant T} |X_t^n - X_t| > \varepsilon\right) \to 0. \tag{7.7}$$

We will once again assume that P_0 and P_0^n defined, respectively, by (2.5) and (5.4) are probability measures on $(\Omega, \mathscr{F})$. Let π and π^n be the optimal nonlinear filters (defined by (2.7) and (6.9), respectively). Similarly, let σ and σ^n be the unnormalized filters (defined by (2.8) and (6.8), respectively).

Recall that $\sigma_t(f)(\omega)=\tilde{\sigma}_t(f,Y(\omega))$ a.s. $[P]$ and $\sigma_t^n(f)(\omega)=\tilde{\sigma}_t^n(f,Y^n(\omega))$, where $\tilde{\sigma}_t$ and $\tilde{\sigma}_t^n$ are defined by (3.4) and (6.5), respectively. A similar statement holds for the normalized filters. In Theorem 6.1 we showed that $\tilde{\pi}^n$ converges in probability to $\tilde{\pi}$ on the Wiener space and deduced the weak convergence of π^n to π. Here, we will show that the convergence of π^n to π is in fact in P-probability. The following lemma is a crucial step towards this. It seems to be a simple measure of theoretical result. However, we are unable to find a reference for the same and hence include the proof here.

Lemma 7.2. *Let U be a random variable and $\{U_n\}$ be a sequence of random variables on a probability space $(\Omega^*,\mathscr{F}^*,P^*)$ such that*
(a) $P^*\circ(U_n)^{-1}\Rightarrow P^*\circ U^{-1}$,
(b) $\liminf U_n \geqslant U$ *a.s.* P^*.
Then $U_n\to U$ in P^*-probability.

Proof. Let $V_n=\tan^{-1}(U_n)$, $V=\tan^{-1}(U)$. Then V_n and V are bounded, $P^*\circ(V_n)^{-1}\Rightarrow P^*\circ(V)^{-1}$ and

$$\liminf V_n \geqslant V \quad \text{a.s. } [P^*]. \tag{7.8}$$

Since $\{V_n\}$ are bounded, we get $E(V_n)\to E(V)$.

On the other hand, using boundedness of $\{V_n\}$, we get by an application of Fatou's lemma

$$E\left(\liminf_{n\to\infty} V_n\right)\leqslant \liminf_{n\to\infty} E(V_n)=E(V). \tag{7.9}$$

Now (7.8) and (7.9) imply

$$\liminf_{n\to\infty} V_n = V \quad \text{a.s. } [P^*].$$

Let $\tilde{V}_m=\inf_{n\geqslant m}V_n$. Then $\tilde{V}_m\to\liminf V_n=V$ a.s.

We thus have $\tilde{V}_n\leqslant V_n$, $\tilde{V}_n\to V$ a.s. and $V_n\Rightarrow V$. Since $\{\tilde{V}_n\}$ and $\{V_n\}$ are converging in law, the sequence $\{(\tilde{V}_n,V_n)\}$ is tight as $\mathbb{R}^2$-valued random variables. If $(\tilde{V}_{n_k},V_{n_k})$ is a convergent subsequence, with $(\tilde{V}_0,V_0)$ as a weak limit, then $\tilde{V}_{n_k}\leqslant V_{n_k}$ implies that $\tilde{V}_0\leqslant V_0$ a.s. On the other hand, $\tilde{V}_0,V_0$ both have same law as V. Hence, $\tilde{V}_0=V_0$ a.s. We then conclude,

$$(\tilde{V}_n,V_n)\Rightarrow(V,V).$$

It then follows that $P(|\tilde{V}_n-V_n|\geqslant\varepsilon)\to P(|V-V|\geqslant\varepsilon)=0$ for any $\varepsilon>0$. Since $\tilde{V}_n\to V$ a.s., it follows that $V_n\to V$ in probability. $\square$

Recall that $\mathscr{M}_+(\mathbb{R}^d)$ denotes the space of positive finite measures on $\mathbb{R}^d$ with Prohorov metric.

Theorem 7.3. *Let $(X,Y),(X^n,Y^n)$ be solutions of SDE's* (7.1) *and* (7.2), *respectively. Let the coefficients $a,b,c,h,bh,a^n,b^n,c^n,h^n,b^nh^n$ satisfy condition* (2.2) *for a fixed constant K. Assume that P_0 and P_0^n defined by* (2.5) *and* (5.4), *respectively, are*

probability measures. Further let (5.3) *and* (7.3) *be satisfied. Then,*
(a) $\sigma^n \to \sigma$ *in* P*-probability as* $C([0,T],\mathscr{M}_+(\mathbb{R}^d))$ *valued processes.*
(b) $\pi^n \to \pi$ *in* P*-probability as* $C([0,T],\mathscr{M}_+(\mathbb{R}^d))$ *valued processes.*

Proof. (a) Let $(\Gamma,\mathscr{G},R)$ be the Skorokhod representation space as in the proof of Theorem 5.1. Let $\hat{X}_0^n$, $\hat{X}_0$ be random variables defined on this space such that $\hat{X}^n \to \hat{X}$ a.s. $[R]$, and $R\circ(\hat{X}_0^n)^{-1}=\pi_0^n$, $R\circ(\hat{X})^{-1}=\pi_0$.

Consider the product space

$$(\hat{\Omega},\hat{\mathscr{F}},\hat{P})=(\Gamma,\mathscr{G},R)\otimes(\Omega_1,\mathscr{F}^\beta,Q_1)\otimes(\Omega,\mathscr{F},P).$$

(Recall that $\Omega_1=C([0,T],\mathbb{R}^d)$ and that Q_1 is the Wiener measure.) Define $\hat{X}^n$,$\hat{X}$ and $\hat{p}^n$, $\hat{p}$ on $\hat{\Omega}$ by

$$\hat{X}_t^n(\gamma,\omega_1,\omega)=\mathfrak{e}^n(\hat{X}_0^n(\gamma),\omega_1,Y^n(\omega)),\quad \hat{X}_t(\gamma,\omega_1,\omega)=\mathfrak{e}(\hat{X}_0(\gamma),\omega_1,Y(\omega)),\quad (7.10)$$

$$\hat{p}_t^n(\gamma,\omega_1,\omega)=p_t^n(\gamma,\omega_1,Y^n(\omega)),\quad \hat{p}_t(\gamma,\omega_1,\omega)=p_t(\gamma,\omega_1,Y(\omega)),\qquad (7.11)$$

where $\mathfrak{e}^n,\mathfrak{e}$, p^n, p are as in (5.6), (3.1), (6.6) and (6.3), respectively. Note that $(\hat{X}^n,Y^n)$ is a solution of the SDE (5.5) with $W^{n,1}$ replaced by β and hence the law of $\hat{X}^n$ under $\hat{P}$ is same as the law of X^n under P. Similarly, the law of $\hat{X}$ under $\hat{P}$ is same as the law of X under P. It follows that for $f\in C_b(\mathbb{R}^d)$,

$$\sigma_t^n(f,\omega)=\iint f(\hat{X}_t^n(\gamma,\omega_1,\omega))\,\hat{p}_t^n(\gamma,\omega_1,\omega)\,\mathrm{d}R(\gamma)\,\mathrm{d}Q_1(\omega_1),$$

$$\sigma_t(f,\omega)=\iint f(\hat{X}_t(\gamma,\omega_1,\omega))\,\hat{p}_t(\gamma,\omega_1,\omega)\,\mathrm{d}R(\gamma)\,\mathrm{d}Q_1(\omega_1).\qquad (7.12)$$

Note that $\hat{p}^n$ and $\hat{p}$ can also be represented by

$$\hat{p}_t^n=\exp\left\{\int_0^t\sum_{i=1}^k h^{n,i}(u,\hat{\mathscr{X}}^n,\mathscr{Y}^n)\,\mathrm{d}Y_u^{n,i}-\frac{1}{2}\int_0^t|h^n(u,\hat{\mathscr{X}}^n,\mathscr{Y}^n)|^2\,\mathrm{d}u\right\},\qquad (7.13)$$

$$\hat{p}_t=\exp\left\{\int_0^t\sum_{i=1}^k h^i(u,\hat{\mathscr{X}},\mathscr{Y})\,\mathrm{d}Y_u^i-\frac{1}{2}\int_0^t|h(u,\hat{\mathscr{X}},\mathscr{Y})|^2\,\mathrm{d}u\right\}.\qquad (7.14)$$

As in (6.11) we get

$$\sup_{0\leqslant t\leqslant T}|h^n(t,\hat{\mathscr{X}}^n,\hat{\mathscr{Y}}^n)-h(t,\hat{\mathscr{X}},\hat{\mathscr{Y}})|\to 0\quad\text{in }\hat{P}\text{ probability.}\qquad (7.15)$$

Using (7.4) and (7.5) we conclude that

$$\sup_{0\leqslant t\leqslant T}\left|\int_0^t h^{n,i}(u,\hat{\mathscr{X}}^n,\hat{\mathscr{Y}}^n)\,\mathrm{d}Y_u^{n,i}-\int_0^t h^i(u,\hat{\mathscr{X}},\hat{\mathscr{Y}})\,\mathrm{d}Y_u^i\right|\to 0\quad\text{in }\hat{P}\text{ probability}\qquad (7.16)$$

and

$$\int_0^T|h^n(u,\hat{\mathscr{X}}^n,\hat{\mathscr{Y}}^n)-h(u,\hat{\mathscr{X}},\hat{\mathscr{Y}})|^2\,\mathrm{d}u\to 0\quad\text{in }\hat{P}\text{ probability.}\qquad (7.17)$$

Let $t_n \to t$ and let $f \in C_b(\mathbb{R}^d)$. As a consequence of (7.16) and (7.17) we get

$$\hat{p}^n_{t_n} \to \hat{p}_t \quad \text{in } \hat{P} \text{ probability} \tag{7.18}$$

and hence

$$f(\hat{X}^n_{t_n})\hat{p}^n_{t_n} \to f(\hat{X}_t)\hat{p}_t \quad \text{in } \hat{P} \text{ probability.} \tag{7.19}$$

Let $\theta_n = f(\hat{X}^n_{t_n})\hat{p}^n_{t_n}$ and let $\theta = f(\hat{X}_t)\hat{p}_t$. Consider any subsequence (θ_{n_k}) of θ_n. Then $\theta_{n_k} \to \theta$ in $R \otimes Q_1 \otimes P$ probability. Thus there exists a further subsequence, say n_{k_j} such that

$$\theta_{n_{k_j}} \to \theta \quad \text{a.s. } [R \otimes Q_1 \otimes P] \text{ as } j \to \infty.$$

Now applying Fatou's lemma and using (7.12) we get

$$\liminf_{j\to\infty} \iint \theta_{n_{k_j}} \, \mathrm{d}R \, \mathrm{d}Q_1 \geqslant \iint \liminf_{j\to\infty} \theta_{n_{k_j}} \, \mathrm{d}R \, \mathrm{d}Q_1 = \iint \theta \, \mathrm{d}R \, \mathrm{d}Q_1. \tag{7.20}$$

On the other hand, using (7.12) and Theorem 6.1 we have

$$\liminf_{j\to\infty} \iint \theta_{n_{k_j}} \, \mathrm{d}R \, \mathrm{d}Q_1 = \sigma^{n_{k_j}}_{t_{n_{k_j}}}(f,\cdot) \Rightarrow \sigma_t(f,\cdot) = \iint \theta \, \mathrm{d}R \, \mathrm{d}Q_1. \tag{7.21}$$

Thus, using (7.20), (7.21) and Lemma 7.2 we get that

$$\sigma^{n_{k_j}}_{t_{n_{k_j}}}(f,\cdot) \to \sigma_t(f,\cdot)$$

in P-probability. Since the subsequence (n_k) was arbitrary, we have shown that any subsequence of $\sigma^n_{t_n}$ has a further subsequence that converges in P-probability to σ_t. This implies that

$$\sigma^n_{t_n}(f,\cdot) \to \sigma_t(f,\cdot) \quad \text{in } P\text{-probability.} \tag{7.22}$$

Since t_n is an arbitrary sequence converging to t, we get

$$\sup_{0\leqslant t\leqslant T} |\sigma^n_t(f,\cdot) - \sigma_t(f,\cdot)| \to 0 \quad \text{in } P\text{-probability.} \tag{7.23}$$

This holds for all $f \in C_b(\mathbb{R}^d)$. Hence we get (a).

(b) As in (4.4) we have

$$P\left(\omega : \inf_{0\leqslant t\leqslant T} \sigma^n_t(\mathbf{1},\omega) > 0\right) = 1 \quad \forall n,$$

$$P\left(\omega : \inf_{0\leqslant t\leqslant T} \sigma_t(\mathbf{1},\omega) > 0\right) = 1.$$

The proof of part (b) follows from this fact along with the definitions of π^n and π and part (a). $\square$

Remark 7.1. In literature, we find that robustness of the filter is studied vis-a-vis convergence in law of π^n to π. This may be useful when we want to simulate the true filter $\pi = \tilde{\pi}(Y)$, but are only able to simulate an approximate filter $\pi^n = \tilde{\pi}^n(Y^n)$. Here we have shown that this convergence holds in probability. In the context of filtering theory, this is of more practical relevance, as seen in the following two Remarks.

Remark 7.2. Suppose that the true (signal, observation) pair is given by (7.1) but is approximated by the (signal, observation) pair modelled by (7.2) (for a large parameter n). Then the true filter π is given by $\tilde{\pi}(Y)$ where as the filter computed based on the model (7.2) will be $\tilde{\pi}^n(Y)$. Since the law of Y is absolutely continuous with respect to the Wiener measure, Theorem 6.1 would imply that $\tilde{\pi}^n(Y)$ converges to $\tilde{\pi}(Y)$ in probability. Thus, no serious error is committed by using an approximate model.

Remark 7.3. Now suppose that the true (signal, observation) pair is given by (7.2) (for a large parameter n) but is approximated by the (signal, observation) pair modelled by (7.1). Then the true filter π is given by $\tilde{\pi}^n(Y^n)$ (for the true observations are Y^n) where as the filter computed based on the model (7.1) will be $\tilde{\pi}(Y^n)$. Using arguments similar to those used in the proof of Theorem 7.3, it can be shown that $\tilde{\pi}(Y^n)$ converges to $\tilde{\pi}(Y)$ in probability. Since $\tilde{\pi}^n(Y^n)$ also converges in probability to $\tilde{\pi}(Y)$, it follows that $\tilde{\pi}^n(Y^n) - \tilde{\pi}(Y^n)$ converges in probability to 0 again justifying the approximation.

Remark 7.4. In the usual signal–noise independent case, where the observation Y is given by

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t$$

with W being a Wiener process independent of signal X, if we approximate X by X^n in probability and consider

$$Y_t^n = \int_0^t h^n(X_s^n)\,\mathrm{d}s + W_t,$$

we can conclude that the filters π^n converges to π in probability if

$$\int_0^T |h^n(X_s^n) - h(X_s)|^2\,\mathrm{d}s \to 0 \quad \text{in probability.}$$

This supplements the results in Bhatt et al. (1999) where we had shown convergence in law.

References

Bhatt, A.G., Kallianpur, G., Karandikar, R.L., 1995. Uniqueness and robustness of solution of measure-valued equations of nonlinear filtering. Ann. Probab. 23 (4), 1895–1938.

Bhatt, A.G., Kallianpur, G., Karandikar, R.L., 1999. Robustness of the optimal filter. Stochastic Process. Appl. 81, 247–254.

Bhatt, A.G., Karandikar, R.L., 1999a. Characterization of the optimal filter: the nonMarkov case. Stochastics Stochastic Rep. 66, 177–204.

Bhatt, A.G., Karandikar, R.L., 1999b. Path continuity of the nonlinear filter. Statist. Probab. Lett. 54, 75–78.

Budhiraja, A., Kallianpur, G., 1996. Approximation to the solution of the Zakai equation using Multiple Wiener and Stratonovich integral expansions. Stochastics Stochastic Rep. 56, 271–315.

Elliott, R.J., 1982. Stochastic Calculus and Applications. Springer, New York, Heidelberg, Berlin.

Elliott, R.J., Glowinski, R., 1989. Approximations to solutions of the Zakai filtering equation. Stochastic Anal. Appl. 7 (2), 145–168.

Emery, M., 1979. Equations differentielles Lipschitziennes: etude de la stability. In: Dellacherie, C., Meyer, P.A., Weil, M. (Eds.), Séminaire de Probabilités XIII, Lecture Notes in Mathematics, Vol. 721, Springer, Berlin, pp. 280–293.

Ethier, S.N., Kurtz, T.G., 1986. Markov Processes: Characterization and Convergence. Wiley, New York.

Florschinger, P., Le Gland, F., 1991. Time discretization of Zakai equation for diffusion processes observed in correlated noise. Stochastics Stochastic Rep. 35, 233–256.
Goggin, E.M., 1992. Convergence of filters with applications to the Kalman–Bucy case. IEEE Trans. Inform. Theory 38, 1091–1100.
Goggin, E.M., 1994. Convergence in distribution of conditional expectations. Ann. Probab. 22, 1097–1114.
Kallianpur, G., 1980. Stochastic Filtering theory. Springer, New York, Heidelberg, Berlin.
Karandikar, R.L., 1989. On Métivier–Pellaumail inequality, Emery topology and Pathwise formulae in stochastic calculus. Sankhya Ser. A 51, 121–143.
Karandikar, R.L., 1995. On pathwise stochastic integration. Stochastics Process. Appl. 57, 11–18.
Métivier, M., 1982. Semimartingales. Walter de Gruyter, Berlin, New York.
Novikov, A.A., 1972. On an identity for stochastic integrals. Theory Probab. Appl. 17, 717–720.