Robustness of the nonlinear filter

Abhay G. Bhatt[a,*,1], G. Kallianpur[b], Rajeeva L. Karandikar[a,1]

[a] *Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi 110016, India*

[b] *Center for Stochastic Processes, University of North Carolina at Chapel Hill, NC, USA*

Received 8 June 1998; received in revised form 17 November 1998; accepted 16 December 1998

Abstract

In the nonlinear filtering model with signal and observation noise independent, we show that the filter depends continuously on the law of the signal. We do not assume that the signal process is Markov and prove the result under minimal integrability conditions. The analysis is based on expressing the nonlinear filter as a Wiener functional via the Kallianpur–Striebel Bayes formula.

MSC classification: primary 60G35; 62M20 secondary 60G44; 60H15; 60J35

Keywords: Nonlinear filtering; Robustness

1. Introduction

In a recent paper (Bhatt et al., 1995), we had proved that the filter depends continuously on the law of the signal process (in the signal noise-independent case). The approach in this paper was via the characterization of the filter as the unique solution to the Zakai (and FKK) equation and thus was applicable to the case of Markov signals. Moreover, the proof required an exponential integrability condition to be satisfied by the signal (see (8.5) in the paper cited above).

Here we will again restrict attention to the case when the signal and the observation noise are independent. Using only the Kallianpur–Striebel Bayes formula we will show continuous dependence of the filter on the signal. This allows us to consider signal processes which may not be Markov. Also, we are able to do away with the exponential integrability condition.

The Bayes formula allows us to view the filter as a functional on the Wiener space evaluated at the observation path (see (2.8) below). Careful analysis of this functional is the crucial step in deducing the robustness of the filter.

Recently, Goggin (1992,1993,1994) has looked at the robustness question from the point of view of convergence of conditional expectations. This result requires assumptions such as equicontinuity of the Radon–Nikodym derivatives of the respective reference probability measures. A result due to (di Masi and Runggaldier, 1982) on robustness of filter for the case when observation noise includes Poisson noise is deduced by her. In the case of Wiener observation noise, she has also obtained a result on robustness when the approximating sequence of signal process arises via a specific approximation scheme and approximation of the Wiener noise is done via a Gaussian random walk. Stettner (1989) and Kunita (1991) also show that the filter is a Feller continuous Strong Markov process (under some suitable conditions) and thus continuous dependence of the filter on the initial condition follows. In all the papers cited above, the convergence of the filter is shown for each fixed t.

It should be noted that here we are considering a fairly general case. In particular, the result is applicable when the state space is a complete separable metric space and when the observation function h may be unbounded. The signal process is allowed to be fairly general (only r.c.l.l. paths are assumed) and the only condition required is (4.7). Also, we consider convergence of the filter in $D([0,T],\mathscr{P}(E))$.

2. The filtering model

Consider the nonlinear filtering model

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t, \quad 0 \leqslant t \leqslant T, \tag{2.1}$$

where X is the signal process, assumed to take values in a complete separable metric space E and having r.c.l.l. paths, the observation noise W is assumed to be an $\mathbb{R}^k$ valued Brownian motion, h is a measurable function and Y is the observation process. The optimal filter π_t is given by

$$\langle \pi_t, f\rangle = E[f(X_t)|\mathscr{F}_t^Y], \quad \forall f \in C_{\mathrm{b}}(E). \tag{2.2}$$

Here $C_{\mathrm{b}}(E)$ is the class of bounded continuous functions on E, the processes X and W are defined on a probability space $(\Omega,\mathscr{F},P)$ and

$$\mathscr{F}_t^Y = \sigma\{Y_s\colon 0 \leqslant s \leqslant t\}$$

is the observation σ-field.

The function h is assumed to satisfy

$$\xi = \int_0^T |h(X_s)|^2\,\mathrm{d}s < \infty \quad \text{a.s. } [P]. \tag{2.3}$$

In this paper, we will restrict our attention to the case when the signal process X is independent of the observation noise W. In this case, there is an explicit expression for the filter π_t given by the Kallianpur–Striebel Bayes formula which we describe below.

It is straight forward to verify that the measure P_0 defined by

$$\frac{\mathrm{d}P_0}{\mathrm{d}P} = \exp\left\{-\int_0^T \sum_{i=1}^k h^i(X_s)\,\mathrm{d}W_s^i - \frac{1}{2}\sum_{i=1}^k \int_0^T |h^i(X_s)|^2\,\mathrm{d}s\right\} \tag{2.4}$$

is a probability measure on $(\Omega,\mathscr{F})$. Further, under P_0, Y is a Brownian motion independent of X and the law of X under P_0 is same as the law of X under P.

Let $\Omega^0 = C([0,T],\mathbb{R}^k)$, $\mathscr{F}^0$ be the Borel σ-field on Ω^0 and Q be the Wiener measure on $(\Omega^0,\mathscr{F}^0)$. Let $\tilde{Y}$ be the coordinate process on Ω^0. Let $\tilde{X}$ be a process defined on some probability space $(\hat{\Omega},\hat{\mathscr{F}},\hat{P})$ where the law of $\tilde{X}$ is same as the law of X. Let $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{P}) = (\hat{\Omega},\hat{\mathscr{F}},\hat{P}) \otimes (\Omega^0,\mathscr{F}^0,Q)$.

Note that the law of (X,Y) on $(\Omega,\mathscr{F},P_0)$ is the same as the law of $(\tilde{X},\tilde{Y})$ on $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{P})$.

Define F by

$$\langle F_t(\omega^0), f\rangle = \int f(\tilde{X}_t(\hat{\omega}))q_t(\hat{\omega},\omega^0)\,\mathrm{d}\hat{P}(\hat{\omega}), \quad \forall f \in C_b(E), \tag{2.5}$$

where

$$q_t(\hat{\omega},\omega^0) = \exp\left\{\sum_{i=1}^{k}\int_0^t h^i(\tilde{X}_s(\hat{\omega}))\,\mathrm{d}\tilde{Y}_s^i(\omega^0) - \frac{1}{2}\sum_{i=1}^{k}\int_0^t (h^i(\tilde{X}_s(\hat{\omega})))^2\,\mathrm{d}s\right\}. \tag{2.6}$$

Also let H be defined by

$$\langle H_t(\omega^0), f\rangle = \frac{F_t(\omega^0)}{\langle F_t(\omega^0), 1\rangle}. \tag{2.7}$$

Then

$$E[f(X_t)|\mathscr{F}_t^Y] = \frac{\langle F_t(Y), f\rangle}{\langle F_t(Y), 1\rangle} = H_t(Y) \quad \text{a.s. } P. \tag{2.8}$$

This is the Kallianpur–Striebel formula. See (Kallianpur and Karandikar, 1988, appendix). In view of (2.8) we define the conditional distribution π_t of X_t given $\mathscr{F}_t^Y$ under P by

$$\langle \pi_t, f\rangle(\omega) = \langle H_t(Y(\omega)), f\rangle. \tag{2.9}$$

Here is a simple result needed in the sequel.

Lemma 2.1. *Let*

$$\rho_t(\omega^0) = \int q_t(\hat{\omega},\omega^0)\,\mathrm{d}\hat{P}(\hat{\omega}), \quad 0 \leqslant t \leqslant T.$$

Then ρ_t admits a continuous modification $\tilde{P}$ (under Q) and further

$$Q\left(\omega^0 : \inf_{0\leqslant t\leqslant T} \tilde{P}_t(\omega^0) > 0\right) = 1. \tag{2.10}$$

Proof. Note that $(q_t, \mathscr{G}_t)$ is a martingale on $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{P})$ where

$$\mathscr{G}_t = \sigma\{(\tilde{X}_s,\tilde{Y}_s)\colon 0\leqslant s\leqslant t\}.$$

This follows from the independence of $\tilde{X}$ and $\tilde{Y}$. See Kallianpur (1980).

As a consequence $(\rho_t,\mathscr{F}_t)$ is a martingale on $(\Omega^0,\mathscr{F}^0,Q)$ where $\mathscr{F}_t = \sigma\{\tilde{Y}_s\colon 0\leqslant s\leqslant t\}$. Since $\tilde{Y}$ is a Brownian motion, this implies that ρ_t admits a continuous modification. The last part follows from the fact that $Q(\rho_T > 0) = 1$. □

This result along with a classical result due to Yor on path properties of the nonlinear filter give us the following result.

Theorem 2.2. *(F_t) admits a r.c.l.l. modification (under Q).*

Proof. The classical result due to Yor (1977) implies that $\pi_t(\omega)$ has r.c.l.l. modification under P. Since the law of Y under P is equivalent to the Wiener measure Q, it follows from (2.9) that the process H_t has r.c.l.l. modification under Q. Noting that $\langle F_t, 1\rangle = \rho_t$ it now follows from the equation

$$H_t(\omega^0) = \frac{F_t(\omega^0)}{\tilde{P}_t(\omega^0)}$$

and the observation that $\tilde{P}_t$ has continuous paths a.s. Q, we conclude that F_t has a r.c.l.l. modification under Q. □

3. Robustness of the filter

Let X^n, X be $D[(0,T],E)$-valued processes, defined respectively on $(\Omega^n, \mathscr{F}^n, P^n)$ and $(\Omega, \mathscr{F}, P)$ such that $X^n \Rightarrow X$. Here and in the sequel, $\Rightarrow$ denotes convergence in distribution of random variables as well as weak convergence of probability measures.

We take the observation models to be

$$Y_t^n = \int_0^t h^n(X_s^n)\,\mathrm{d}s + W_t^n, \quad 0 \leqslant t \leqslant T$$

and

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t, \quad 0 \leqslant t \leqslant T,$$

where for every n, X^n and W^n are independent processes defined on $(\Omega^n, \mathscr{F}^n, P^n)$. Also X and W defined on $(\Omega, \mathscr{F}, P)$ are independent. W^n and W are $\mathbb{R}^k$-valued Brownian motions and h^n, h are measurable functions from E into $\mathbb{R}^k$. Also, as in (2.3), we assume that

$$\xi^n = \int_0^T |h^n(X_s^n)|^2\,\mathrm{d}s < \infty \quad \text{a.s. } [P^n]. \tag{3.1}$$

Let us define

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

$$Z_t^n = \int_0^t h^n(X_s^n)\,\mathrm{d}s.$$

We will assume that

$$(X^n, Z^n, \xi^n) \Rightarrow (X, Z, \xi) \tag{3.2}$$

in the sense of convergence in distribution as $D([0,T],E) \times C([0,T],\mathbb{R}^k) \times \mathbb{R}$-valued random variables.

Under this assumption, we will prove that the conditional distribution of X_t^n given $\sigma\{Y_s^n:\ 0 \leqslant s \leqslant t\}$ converges weakly to the conditional distribution of X_t given $\sigma\{Y_s: 0 \leqslant s \leqslant t\}$ as measure-valued processes.

Before proceeding, we will give a set of sufficient conditions that imply (3.2).

$$h^n \text{ converges to } h \text{ uniformly on compacts}, \tag{3.3}$$

$$h \text{ is a continuous function}, \tag{3.4}$$

$$X^n \Rightarrow X, \tag{3.5}$$

$$\lim_{n\to\infty} \mathbb{E}^{P^n}\left[\int_0^T |h^n(X_s^n)|^2\,\mathrm{d}s\right] = \mathbb{E}^P\left[\int_0^T |h(X_s)|^2\,\mathrm{d}s\right]. \tag{3.6}$$

It can be seen that conditions (3.3)–(3.6) imply (3.2). Also (3.6) can be replaced by the weaker condition

$$\lim_{K\to\infty} \sup_n P^n\left(\int_0^T |h^n(X_s^n)|^2 1_{\{|h^n(X_s^n)|>K\}}\,\mathrm{d}s \geqslant \varepsilon\right) = 0, \quad \forall \varepsilon > 0. \tag{3.7}$$

Let P_0 be defined by (2.4). Define P_0^n similarly with X^n in place of X, W^n in place of W and h^n in place of h.

It should be noted that the law of Y under the reference probability measure P_0 is same as the law of the approximating sequence Y^n under the corresponding reference probability measures P_0^n (both the laws are Wiener measure). This fact is crucially used in the sequel. When this is not the case, the method given below needs to be modified and would require additional assumptions. Work on this aspect is under progress.

Using Skorokhod's representation theorem, get $\tilde{X}^n, \tilde{Z}^n, \tilde{\xi}^n, \tilde{X}, \tilde{Z}, \tilde{\xi}$ on $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{P})$ such that

$$(\tilde{X}^n, \tilde{Z}^n, \tilde{\xi}^n) \to (\tilde{X}, \tilde{Z}, \tilde{\xi}) \quad \text{a.s. } [\hat{P}] \tag{3.8}$$

and

$$\mathscr{L}(\tilde{X}^n, \tilde{Z}^n, \tilde{\xi}^n) = \mathscr{L}(X^n, Z^n, \xi^n), \ \ \mathscr{L}(\tilde{X}, \tilde{Z}, \tilde{\xi}) = \mathscr{L}(X, Z, \xi). \tag{3.9}$$

Lemma 3.1. $\int_0^T |h^n(\tilde{X}_s^n) - h(\tilde{X}_s)|^2\,\mathrm{d}s \to 0$ *a.s.* $[\hat{P}]$.

Proof. It follows from the definitions of Z^n, ξ^n, Z, ξ and (3.9) that

$$\tilde{Z}_t^n = \int_0^t h^n(\tilde{X}_s^n)\,\mathrm{d}s \quad \text{a.s. } [\hat{P}],$$

$$\tilde{\xi}^n = \int_0^T |h^n(\tilde{X}_s^n)|^2\,\mathrm{d}s \quad \text{a.s. } [\hat{P}],$$

$$\tilde{Z}_t = \int_0^t h(\tilde{X}_s)\,\mathrm{d}s \quad \text{a.s. } [\hat{P}],$$

$$\tilde{\xi} = \int_0^T |h(\tilde{X}_s)|^2\,\mathrm{d}s \quad \text{a.s. } [\hat{P}].$$

Thus, (3.8) and the observation that for $g_n, g \in L^2[0,T]$,

$$\int_0^t g_n(s)\,\mathrm{d}s \to \int_0^t g(s)\,\mathrm{d}s, \quad \forall t \in [0,T]$$

and

$$\int_0^T |g_n(s)|^2\,\mathrm{d}s \to \int_0^T |g(s)|^2\,\mathrm{d}s$$

implies

$$\int_0^T |g_n(s) - g(s)|^2\,\mathrm{d}s \to 0$$

gives the required result. □

Recall from Section 2 the definitions of Ω^0, $\hat{\Omega}$ and $\hat{Y}$. We will consider $\tilde{X}^n, \tilde{X}$ as processes on $\tilde{\Omega}$.

Again let F, q_t be defined by (2.5), (2.6) and similarly, F^n, q_t^n by

$$\langle F_t^n(\omega^0), f\rangle = \int f(\tilde{X}_t^n(\hat{\omega}))q_t^n(\hat{\omega},\omega^0)\,\mathrm{d}\hat{P}(\hat{\omega}), \quad \forall f \in C_b(E) \tag{3.10}$$

and

$$q_t^n(\hat{\omega},\omega^0) = \exp\left\{\sum_{i=1}^k \int_0^t h^{n,i}(\tilde{X}_s^n(\hat{\omega}))\,\mathrm{d}\tilde{Y}_s^i(\omega^0) - \frac{1}{2}\sum_{i=1}^k \int_0^t (h^{n,i}(\tilde{X}_s^n(\hat{\omega})))^2\,\mathrm{d}s\right\}. \tag{3.11}$$

Then we have that like F_t, F_t^n also admits a r.c.l.l. modification. We will continue to denote these r.c.l.l. modifications by F_t, F_t^n.

Let us note that if $\tilde{P}^n$ is defined by $\mathrm{d}\tilde{P}^n = q_T^n\,\mathrm{d}\tilde{P}$ then the law of $\tilde{X}^n$ under $\tilde{P}^n$ is the same as the law of X^n under P^n. Let us note that the law of $(\tilde{X}^n, \tilde{Y})$ under $\tilde{P}^n$ equals the law of (X^n, Y^n) under the reference probability measure P_0^n. The fact that we can achieve this with the second component (in $(\tilde{X}^n, \tilde{Y})$) not depending upon n simplifies lots of arguments that follow.

Let d be the Prohorov metric on $\mathscr{M}_+(E)$ – the set of positive finite measures on E. We will also denote by $\mathscr{P}(E)$ – the set of probability measures on E.

The following is our main result on robustness.

Theorem 3.2. *Assume that X is continuous in probability and that* (3.2) *holds. Then for $\varepsilon > 0$,*

$$\lim_{n\to\infty} Q\left(\sup_{t\in[0,T]} d(F_t^n, F_t) > \varepsilon\right) = 0. \tag{3.12}$$

In particular, $F^n \to F$ in Q-probability as $D([0,T], \mathscr{M}_+(E))$-valued random variables.

Proof. It suffices to prove that for $t_n \to t$, $F_{t_n}^n$ converges to F_t in Q-probability, i.e. $\forall \varepsilon > 0$

$$\lim_{n\to\infty} Q(d(F_{t_n}^n, F_t) > \varepsilon) = 0.$$

This in turn is implied by

$$\lim_{n\to\infty} Q(|\langle F_{t_n}^n, g\rangle - \langle F_t, g\rangle| > \varepsilon) = 0 \tag{3.13}$$

for all bounded continuous functions g on E and $\forall \varepsilon > 0$.

From Lemma 3.1 it follows that $q^n_{t_n} \to q_t$ in $\hat{P} \otimes Q$ probability. Further note that

$$\int q^n_{t_n}\,\mathrm{d}(\hat{P} \otimes Q) = 1$$

and

$$\int q_t\,\mathrm{d}(\hat{P} \otimes Q) = 1.$$

Hence (by Scheffe's Lemma)

$$q^n_{t_n} \to q_t \quad \text{in } L^1(\hat{P} \otimes Q). \tag{3.14}$$

Since X (and hence $\tilde{X}$) is continuous in probability we get (using Ethier and Kurtz (1986, Proposition 3.6.5)) that for $g \in C_b(E)$

$$g(\tilde{X}^n_{t_n}) \to g(\tilde{X}_t) \quad \text{a.s. } [\hat{P}]. \tag{3.15}$$

Thus

$$g(\tilde{X}^n_{t_n})q^n_{t_n} \to g(\tilde{X}_t)q_t \quad \text{in } L^1(\hat{P} \otimes Q). \tag{3.16}$$

Eq. (3.13) now follows from Eqs. (3.16), (2.5), (3.10) and Fubini's theorem. □

Let

$$H^n_t = \frac{F^n_t}{\langle F^n_t, 1\rangle}, \qquad H_t = \frac{F_t}{\langle F_t, 1\rangle}.$$

As seen in (2.9)

$$E[f(X_t)|\mathscr{F}^Y_t] = \langle H_t(Y), f\rangle$$

and similarly

$$E[f(X^n_t)|\mathscr{F}^{Y^n}_t] = \langle H^n_t(Y^n), f\rangle.$$

As seen in Lemma 2.1, $\langle F_t(Y), 1\rangle$ is a martingale and

$$\inf_t \langle F_t(Y), 1\rangle > 0 \quad \text{a.s.-}P. \tag{3.17}$$

Similarly, we have

$$\inf_t \langle F^n_t(Y^n), 1\rangle > 0 \quad \text{a.s.-}P^n. \tag{3.18}$$

An immediate consequence of the above theorem is the following.

Theorem 3.3. (a) $H^n \to H$ *in Q-probability as $D([0,T],\mathscr{P}(E))$-valued random variables.*
(b) $P^n \circ (\pi^n)^{-1} \Rightarrow P \circ (\pi)^{-1}$.

Proof. The first part follows immediately from Theorem 3.2 and (3.17), (3.18). For (b) note that for any $G \in C_b(D([0,T],\mathscr{P}(E)))$:

$$\begin{aligned} \mathbb{E}^{P^n}[G(\pi^n)] &= \mathbb{E}^{P^n}[G(H^n(Y^n))] \\ &= \mathbb{E}^{\hat{P}}[G(H^n)q^n_T] \\ &\to \mathbb{E}^{\hat{P}}[G(H)q_T] \\ &= \mathbb{E}^{P}[G(H(Y))] \\ &= \mathbb{E}^{P}[G(\pi)]. \qquad \square \end{aligned}$$

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, 1895–1938.

di Masi, G.B., Runggaldier, W.J., 1982. On approximation methods for nonlinear filtering. Nonlinear Filtering and Stochastic Control. Lecture Notes in Mathematics, vol. 972, Springer, Berlin, pp. 249–259.

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

Goggin, E., 1992. Convergence of filters with applications to the Kalman-Bucy case. IEEE Trans. Inform. Theory 38, 1091–1100.

Goggin, E., 1993. Limits of conditional expectations. IEEE Trans. Inform. Theory 39, 1967–1972.

Goggin, E., 1994. Convergence in distributions of conditional expectations. Ann. Probab. 22, 1097–1114.

Kallianpur, G., 1980. Stochastic Filtering Theory. Springer, New York.

Kallianpur, G., Karandikar, R.L., 1988. White noise theory of Prediction, Filtering and Smoothing. Gordon and Breach, New York.

Kunita, H., 1991. Ergodic Properties of Nonlinear Filtering Processes. In: Alexander, K.S., Watkins, J.C. (Eds.), Spatial Stochastic Processes. Birkhauser, Boston.

Stettner, L., 1989. On invariant measures of filtering processes. In: Stochastic Differential Systems. Lecture Notes in Control and Information Sciences, vol. 126, Springer, Berlin.

Yor, M., 1977. Sur les théories du filtrage et de la prédiction. Séminaire de Probabilités XI. Lecture Notes in Mathematics, vol. 581, Springer, Berlin, pp. 257–297.