Electronic Journal of Probability

Vol. 9 (2004), Paper no. 18, pages 560-574.

Journal URL
http://www.math.washington.edu/~ejpecp/

NONLINEAR FILTERING FOR REFLECTING DIFFUSIONS IN RANDOM ENVIRONMENTS VIA NONPARAMETRIC ESTIMATION

Michael A. Kouritzin
Department of Mathematical and Statistical Sciences, University of Alberta
Edmonton, Alberta, Canada T6G 2G1
mkouritz@math.ualberta.ca

Wei Sun
Department of Mathematical and Statistical Sciences, University of Alberta
Edmonton, Alberta, Canada T6G 2G1
and
Department of Mathematics and Statistics, Concordia University
Montreal, Quebec, Canada H4B 1R6
wsun@alcor.concordia.ca

Jie Xiong
Department of Mathematical and Statistical Sciences, University of Alberta
Edmonton, Alberta, Canada T6G 2G1
and
Department of Mathematics, University of Tennessee
Knoxville, Tennessee, USA TN 37996-1300
jxiong@math.utk.edu

Abstract: We study a nonlinear filtering problem in which the signal to be estimated is a reflecting diffusion in a random environment. Under the assumption that the observation noise is independent of the signal, we develop a nonparametric functional estimation method for finding workable approximate solutions to the conditional distributions of the signal state. Furthermore, we show that the pathwise average distance, per unit time, of the approximate filter from the optimal filter is asymptotically small in time. Also, we use simulations based upon a particle filter algorithm to show the efficiency of the method.

Keywords and phrases: Nonlinear filtering, reflecting diffusion, random environment, nonparametric estimation, pathwise average distance, particle filter.

AMS subject classification (2000): Primary, 60G35; Secondary, 62M20.

Submitted to EJP on March 6, 2003 Final version accepted on July 12, 2004.

1 Introduction

Herein, we consider a nonlinear filtering problem in which the signal to be estimated is a reflecting diffusion process in a random environment. The motivation comes from the tracking problem of a dinghy lost on a lake or ocean based upon infrared camera readings. Whereas in calm waters the dinghy moves as a diffusion process reflecting at the shores, lake swells provide a random environment potentially altering its motion greatly.

For concreteness, we let the lake be $D = (0,1)^d$, denote by ∂D the boundary of D, and define $\overline{D} = D \cup \partial D$. Formally, the $\overline{D}$-valued signal process X_t for the dinghy can be described by the following stochastic differential equation (SDE)

$$dX_t = \left(-\frac{1}{2}a\nabla W + c\right)(X_t)dt + \sigma_1(X_t)dB_t + \chi_{\partial D}(X_t)(aU)(X_t)d\xi_t, \tag{1.1}$$

where W is a $C(\overline{D})$-valued random variable, B_t is a standard $\mathbb{R}^{d_1}$-valued Brownian motion independent of W, $\sigma_1 : \mathbb{R}^d \to \mathbb{R}^{d\times d_1}$ is a measurable matrix-valued function, $a = \sigma_1\sigma_1^T$, $c = (c_1, \ldots, c_d)^T$ is a measurable vector field, χ is the indicator function, U is the unit inward normal, and ξ_t is the local time of X_t. Note that Equation (1.1) is purely heuristic because almost all paths of the random environment W may be only continuous not differentiable. To use a symmetric Dirichlet form to construct a diffusion process formally associated with Equation (1.1), we assume that $c_i = \frac{1}{2}\sum_{j=1}^d \frac{\partial a_{ij}}{\partial x_j}$ for $1 \le i \le d$. Here, the derivatives are taken in the Schwartz distribution sense.

The $\mathbb{R}^d$-valued observation sequence $\{Y_j\}$ is defined by

$$Y_j = h(X_{t_j}) + V_j, \tag{1.2}$$

where $t_j = j\varepsilon$ for some fixed constant $\varepsilon > 0$ and $j \in \mathbb{N}$, $h : \overline{D} \to \overline{D}$ is a one-to-one continuous differentiable vector field so the Jacobian $J(h) \neq 0$ on $\overline{D}$, and $\{V_j\}$ is a sequence of i.i.d. $N(0, \sigma_2)$ random vectors independent of X_t for some fixed positive definite matrix σ_2.

Recently, there have been many developments in the study of diffusion processes in random environments (see Tanaka (1995) for a survey). In Kouritzin, Long and Sun (2003), we began the study of the nonlinear filtering problem in which the signal to be estimated is a diffusion in a random medium and the signal may be correlated with the observation noise. Using Dirichlet form theory, we introduced a precise nonlinear filtering model for the signal process and established a multiple Wiener integrals representation for the unnormalized pathspace filtering process. Among other things, this representation generalizes Kunita (1982)'s Wiener chaos expansion [cf. also Ocone (1983)] to the singular coefficients case. Combining this representation with the idea of Lototsky, Mikulevicius and Rozovskii (1997)'s spectral approach we thus provided the capability to develop a numerical scheme for nonlinear filtering of diffusions in Brownian environments.

In this work, we consider the same filtering problem from a different viewpoint. Our practical simulation experience suggests that particle and space discretization methods of implementing filters often work better than chaos methods. Therefore, it is important to come up with particle system approximations for the conditional distributions of the signal based upon the observations. The difficulty that arises is that we do not want to introduce an extremely large number of particles to account for the random environment.

Instead, we try to "learn" the environment. Under the assumption that the observation noise is independent of the signal, we develop a nonparametric estimation method for finding workable approximate solutions to the conditional distributions of the signal state given the back observations. In this connection, we also refer the interested reader to Chow, Khasminskii and Liptser (1997, 2001), Elliott (2001), Kouritzin, Rémillard and Chan (2001) for some other recent works on nonparametric and parametric estimations for filtering problems. Since the functional estimations in our approximate filters are based upon long time observations, it is important to study the convergence of the approximate filters over an infinite interval. Following Budhiraja and Kushner (1999, 2000a, 2000b)'s occupation measure arguments, we show that the pathwise average distance, per unit time, of the approximate filter from the optimal filter is asymptotically small in time.

The remainder of this article is organized as follows. In Section 2, we perform nonparametric functional estimations for an unknown path of the random environment. In Section 3, we construct approximate filters using the obtained estimators and study their approximations to optimal filters over very long time intervals. Moreover, we introduce a particle filter algorithm for combined state and nonparametric estimation. In Section 4, we use simulation results to show the efficiency of the nonparametric estimation method. Finally, in the Appendix, we give the proof of the limit result in Section 3.

2 Nonparametric functional estimations

We start by introducing a precise signal and observation model. With the most natural example of a random environment, a Brownian sheet, in mind, we assume that the random environment W is a $C(\overline{D})$-valued random variable. That is, we assume that each randomly given outcome of the environment is a continuous function on $\overline{D}$. In this section, we will use the observations to construct estimators for the fixed unknown path of the random environment. Hereafter, to simplify the notation, we let W denote the random environment outcome as well as the $C(\overline{D})$-valued random variable as the context will clarify the exact meaning.

Suppose that $a = \{a_{ij}\}_{i,j=1}^d$ is a symmetric matrix-valued function satisfying the uniformly elliptic condition

$$\frac{1}{\lambda}\sum_{i=1}^d \eta_i^2 \le \sum_{i,j=1}^d a_{ij}(x)\eta_i\eta_j \le \lambda \sum_{i=1}^d \eta_i^2, \quad \forall \eta \in \mathbb{R}^d, \ \forall x \in D$$

for some constant $\lambda > 1$. Let W be the fixed unknown path of the random environment. We use $H^{1,2}(D)$ to denote the $(1,2)$-Sobolev space on D and define

$$d\mu := e^{-W(x)}dx.$$

Then, we consider on $L^2(\overline{D};\mu)$ the symmetric bilinear form

$$\begin{cases} \mathcal{E}(u,v) = \frac{1}{2}\int_D \sum_{i,j=1}^d a_{ij}(x)\frac{\partial u}{\partial x_i}(x)\frac{\partial v}{\partial x_j}(x)\mu(dx), \quad u,v \in \mathcal{D}(\mathcal{E}), \\ \mathcal{D}(\mathcal{E}) = H^{1,2}(D). \end{cases} \tag{2.1}$$

One can check that $\mathcal{E}$ is a regular Dirichlet form satisfying the local property so it is associated with a strong Markov diffusion process $(C_{\overline{D}}[0,\infty),\ \mathcal{B}(C_{\overline{D}}[0,\infty)), (X_t)_{t\geq 0}, (P_x)_{x\in\overline{D}})$ [cf. Fukushima, Oshima and Takeda (1994), Theorem 7.2.1 and Theorem 7.2.2]. Hereafter, $C_{\overline{D}}[0,\infty)$ is the path space on $\overline{D}$ and $\mathcal{B}(C_{\overline{D}}[0,\infty))$ is its Borel σ-algebra. For any $t\geq 0$ and $\omega\in C_{\overline{D}}[0,\infty)$, the coordinate $X_t(\omega)$ refers to the state of the process at time t on the path ω, and for any $x\in\overline{D}$, the probability measure P_x describes the probabilistic behavior of the paths when they are started at the point x at time 0. The association means that if we denote by $(p_t)_{t>0}$ and $(T_t)_{t>0}$ the semigroups associated with X_t and $\mathcal{E}$, respectively, then $p_t f = T_t f\ dx - a.e.$ for any $f\in L^\infty(\overline{D};\mu)$ and any $t>0$. Comparing their (formal) generators, one can consider X_t to be a solution to the formal Skorohod SDE (1.1) [cf. Freidlin (1985), Section 1.6].

If $W\equiv 0$, we use $\mathcal{E}^0$ to denote the Dirichlet form (2.1). Moreover, we denote its associated Markov process and transition semigroup by X_t^0 and $(p_t^0)_{t>0}$, respectively. From, e.g. Theorem 2.2. of Fukushima and Tomisaki (1996), we know that X_t^0 is in fact a strong Feller diffusion. The relationship between $(p_t)_{t>0}$ and $(p_t^0)_{t>0}$ can be characterized by

$$p_t f = e^{\frac{W}{2}} p_t^0(e^{-\frac{W}{2}} f),\quad \forall f\in B_b(\overline{D}),\ \forall t>0, \tag{2.2}$$

where $B_b(\overline{D})$ denotes the set of all bounded measurable functions on $\overline{D}$. Therefore, X_t is also a strong Feller diffusion.

If $W\equiv 0$ and $a\equiv I_d$, the d-dimensional unit matrix, then the diffusion associated with (2.1) is the reflecting Brownian motion on $\overline{D}$. By the ergodicity of the reflecting Brownian motion and the result on comparison of irreducibility [Fukushima, Oshima and Takeda (1994), Corollary 4.6.4], one can see that $\mathcal{E}$ is also irreducible and thus X_t is an ergodic diffusion with the stationary distribution $\mu/\mu(D)$.

Let σ_2 be a fixed $d\times d$-positive definite matrix and $\{V_j\}$ a sequence of i.i.d. $N(0,\sigma_2)$ random vectors on some probability measure space $(\Omega_V,\mathcal{F}_V,P_V)$. We define the observation sequence $\{Y_j\}$ by Equation (1.2) and denote $\mathcal{Y}_j=\sigma\{Y_i, 1\leq i\leq j\}$ for $j\in\mathbb{N}$. Furthermore, we let $\Omega^1 = C_{\overline{D}}[0,\infty)\times\Omega_V$, $\mathcal{F}^1=\mathcal{B}(C_{\overline{D}}[0,\infty))\times\mathcal{F}_V$, and

$$P^1 = \frac{1}{\mu(D)}\int_D (P_x\times P_V)\mu(dx).$$

In the sequel, we use $||f||_\infty$ and $||f||_2$ to denote respectively the supremum norm and $L^2(\overline{D};\mu)$-norm of any measurable function f on $\overline{D}$.

Theorem 1. *Suppose that the fixed unknown path W of the random environment and the Jacobian $J(h)$ are 1-periodic in each argument of the vector x. Then, there exists a sequence $\{\Theta_n\}$ of 1-periodic $C^\infty(\overline{D})$-valued random variables adapted to the filtration $\{\mathcal{Y}_n\}$ such that*

$$\lim_{n\to\infty} ||\Theta_n - (W+\ln\mu(D))||_\infty = 0$$

in probability P^1.

Proof. We denote $O_0=\ln\mu(D)$ and define

$$C_k := \frac{1}{\mu(D)}\int_D e^{-2\pi i k^T h(x)}\mu(dx)$$

$$= \int_D e^{-2\pi i k^T x} e^{-(W\circ h^{-1}(x)+O_0)} |J(h^{-1})(x)| dx, \quad \forall k \in \mathbb{Z}^d.$$

Then, by the ergodicity of X_t and the independence of $\{X_t\}$ and $\{V_j\}$, we obtain from the Birkhoff ergodic theorem [cf. Rosenblatt (1971), Theorem IV.3.1] that for $k \in \mathbb{Z}^d$

$$\begin{aligned}
(2.3) \qquad & \lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^m e^{-2\pi i k^T Y_j} \\
&= \lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^m e^{-2\pi i k^T (h(X_{t_j})+V_j)} \\
&= \frac{1}{\mu(D)} \int_D \int_{\mathbb{R}^d} e^{-2\pi i k^T (h(x)+y)} \frac{1}{(2\pi)^{\frac{d}{2}} |\sigma_2|^{\frac{1}{2}}} e^{-\frac{y^T \sigma_2^{-1} y}{2}} dy \mu(dx) \\
&= e^{-2\pi^2 k^T \sigma_2 k} C_k, \qquad P_x \times P_V - a.e., \quad \forall x \in \overline{D}.
\end{aligned}$$

From (2.3) we see that the Fourier coefficients of $e^{-(W\circ h^{-1}+O_0)}|J(h^{-1})|$ can be estimated using the observations so we may apply trigonometric Fourier series to conduct functional estimation for $W + O_0$. To do this, we first give an estimation of the convergence rate for (2.3). Following, e.g. Lemma 1.2 of Ledoux (2001), one can see that $\mathcal{E}$ satisfies the Poincaré inequality and thus $(T_t)_{t>0}$ has the exponential L^2-convergence rate. Namely, there exists a constant $\alpha > 0$ such that

$$\left|\left| p_t f - \frac{1}{\mu(D)} \int_D f d\mu \right|\right|_2 \le e^{-\alpha t} ||f||_2, \quad \forall f \in L^2(\overline{D}; \mu), \ \forall t > 0.$$

In particular, for any $k \in \mathbb{Z}^d$

$$(2.4) \qquad \left|\left| p_t \left(e^{-2\pi i k^T h} - C_k \right) \right|\right|_2 \le 2e^{-\alpha t}, \quad \forall t > 0.$$

Therefore, we obtain from the independence of $\{X_t\}$ and $\{V_j\}$, the independence of $\{V_j\}$, the Markovian property of X_t and (2.4) that

$$\begin{aligned}
(2.5) \qquad & \sup_{k\in\mathbb{Z}^d} \int \left| \frac{1}{m} \sum_{j=1}^m e^{-2\pi i k^T Y_j} - e^{-2\pi^2 k^T \sigma_2 k} C_k \right|^2 dP^1 \\
&\le 2 \sup_{k\in\mathbb{Z}^d} \left\{ \int \left| \frac{1}{m} \sum_{j=1}^m e^{-2\pi i k^T h(X_{t_j})} \left(e^{-2\pi i k^T V_j} - e^{-2\pi^2 k^T \sigma_2 k} \right) \right|^2 dP^1 \right. \\
&\quad \left. + e^{-4\pi^2 k^T \sigma_2 k} \int \left| \frac{1}{m} \sum_{j=1}^m e^{-2\pi i k^T h(X_{t_j})} - C_k \right|^2 dP^1 \right\} \\
&= 2 \sup_{k\in\mathbb{Z}^d} \left\{ \int \frac{1}{m^2} \left[\sum_{j=1}^m \left| e^{-2\pi i k^T V_j} - e^{-2\pi^2 k^T \sigma_2 k} \right|^2 \right. \right. \\
&\quad + \sum_{j\neq j'}^m e^{-2\pi i k^T (h(X_{t_j}) - h(X_{t_{j'}}))} \left(e^{-2\pi i k^T V_j} - e^{-2\pi^2 k^T \sigma_2 k} \right)
\end{aligned}$$

$$
\cdot \left(e^{2\pi i k^T V_{j'}} - e^{-2\pi^2 k^T \sigma_2 k}\right)\Big]\, dP^1
$$

$$
+\frac{1}{e^{4\pi^2 k^T \sigma_2 k} m^2}\left[\sum_{j=1}^{m}\int_D p_{t_j}|e^{-2\pi i k^T h} - C_k|^2 \mu(dx)\right.
$$

$$
\left.+\sum_{j<j'}^{m}\int_D 2\,\mathrm{Re}\left\{(e^{-2\pi i k^T h} - C_k)p_{t_{j'}-t_j}(e^{2\pi i k^T h} - \overline{C}_k)\right\}\mu(dx)\right]\Bigg\}
$$

$$
\leq 2\sup_{k\in\mathbb{Z}^d}\left\{\frac{1}{m^2}\int\sum_{j=1}^{m}\left|e^{-2\pi i k^T V_j} - e^{-2\pi^2 k^T \sigma_2 k}\right|^2 dP^1\right.
$$

$$
\left.+\frac{1}{m}||e^{-2\pi i k^T h} - C_k||_2^2 + \frac{2}{m}\sum_{j=1}^{m-1}||p_{t_j/2}(e^{-2\pi i k^T h} - C_k)||_2^2\right\}
$$

$$
\leq \frac{16}{m}\left(1+\sum_{j=1}^{m-1}e^{-\alpha j\varepsilon}\right)
$$

$$
\leq \frac{16}{m}\left(1+\frac{1}{e^{\alpha\varepsilon}-1}\right).
$$

Let $\{m_l\}$ be a sequence of positive integers satisfying

$$
\lim_{l\to\infty}\frac{(2l+1)^{3d}}{m_l}\bigvee_{k_1,\dots,k_d=-l}^{l} e^{4\pi^2 k^T \sigma_2 k} = 0.
$$

By (2.5), we find that

$$
\lim_{l\to\infty}\sum_{k_1,\dots,k_d=-l}^{l} e^{2\pi^2 k^T \sigma_2 k}\left|\frac{1}{m_l}\sum_{j=1}^{m_l} e^{-2\pi i k^T Y_j} - e^{-2\pi^2 k^T \sigma_2 k}C_k\right| = 0 \tag{2.6}
$$

in probability P^1. For a $q \in \mathbb{Z}_+^d$ satisfying $q_u \leq l$ for any $1 \leq u \leq d$, we define

$$
S_{l,q}(x) = \sum_{|k_u|\leq q_u, 1\leq u\leq d}\left(\frac{e^{2\pi^2 k^T \sigma_2 k}}{m_l}\sum_{j=1}^{m_l} e^{-2\pi i k^T Y_j}\right)e^{2\pi i k^T x},\quad \forall x \in \overline{D}.
$$

Then, by (2.6),

$$
\lim_{l\to\infty}\bigvee_{q_1,\dots,q_d=0}^{l}\max_{x\in\overline{D}}\left|S_{l,q}(x) - \sum_{|k_u|\leq q_u, 1\leq u\leq d} C_k e^{2\pi i k^T x}\right| = 0 \tag{2.7}
$$

in probability P^1. We denote the Cesàro means by

$$
C^l(x) = \frac{1}{(l+1)^d}\sum_{q_1,\dots,q_d=0}^{l}\sum_{|k_u|\leq q_u, 1\leq u\leq d} C_k e^{2\pi i k^T x},\quad \forall x \in \overline{D}.
$$

Since $W, J(h^{-1}) \in C(\overline{D})$ and both of them are 1-periodic in each argument of the vector x,

$$\lim_{l\to\infty} \left|\left|C^l - e^{-(W\circ h^{-1}+O_0)}|J(h^{-1})|\right|\right|_\infty = 0 \tag{2.8}$$

[cf. Zygmund (1959), Chapter XVII, Theorem 1.20]. Now, we consider the approximate Cesàro means

$$\Lambda_l = \frac{1}{(l+1)^d} \sum_{q_1,\ldots,q_d=0}^{l} S_{l,q}.$$

By (2.7) and (2.8), we know that

$$\lim_{l\to\infty} \left|\left|\Lambda_l - e^{-(W\circ h^{-1}+O_0)}|J(h^{-1})|\right|\right|_\infty = 0$$

in probability P^1. Thus

$$\lim_{l\to\infty} || \ln\{(\Lambda_l \circ h)|J(h)|\} + (W + O_0)||_\infty = 0$$

in probability P^1. Let $\{g_l\}$ be a sequence of smooth functions on $\overline{D}$ satisfying

$$\lim_{l\to\infty} ||g_l - |J(h)|||_\infty = 0.$$

We define

$$\Phi_l = -\ln\{(\Lambda_l \circ h) g_l\}.$$

Then

$$\lim_{l\to\infty} ||\Phi_l - (W + O_0)||_\infty = 0$$

in probability P^1. Therefore, we obtain the desired adapted sequence $\{\Theta_n\}$ by letting

$$\Theta_n = \begin{cases} 0, & n < m_1, \\ \Phi_n, & m_n \le n < m_{n+1}. \end{cases}$$

Remark 1. Many random environments in practical applications only take effect in some subdomains of D. So it is not a strong restriction to assume in Theorem 1 that W satisfies the periodic boundary condition. Also, a lot of practical nonlinear sensor functions (via some linear transformations if necessary) satisfy the periodic boundary condition in Theorem 1. In the proof of Theorem 1, we made no attempt to optimize the positive integers m_l. In practical filtering, we usually let m_l be some suitably chosen large numbers (cf. the simulation in Section 4 for an example of choosing m_l).

3 Approximate filters and approximations over long time intervals

We let $(\Omega^2, \mathcal{F}^2, (B_t^2)_{t\ge 0}, P^2)$ be a standard $\mathbb{R}^{d_1}$-valued Brownian motion, define $\Omega = \Omega^1 \times \Omega^2$, $\mathcal{F} = \mathcal{F}^1 \times \mathcal{F}^2$, $P = P^1 \times P^2$, and take $\{\Theta_n\}$ to be a sequence of $C^\infty(\overline{D})$-valued random variables adapted to the filtration $\{\mathcal{Y}_n\}$ such that

$$\lim_{n\to\infty} ||\Theta_n - (W + O)||_\infty = 0$$

in probability P^1, where W is the fixed unknown path of the random environment and O is some constant (cf. Theorem 1). Then, we define on $(\Omega, \mathcal{F}, P)$ a sequence of reflecting diffusion processes $\{X_t^n\}$ by

$$dX_t^n = \left(-\frac{1}{2}a\nabla\Theta_n + c\right)(X_t^n)dt + \sigma_1(X_t^n)dB_t^2 + \chi_{\partial D}(X_t^n)(aU)(X_t^n)d\xi_t^n, \tag{3.1}$$

where ξ_t^n is the local time of X_t^n for $n \in \mathbb{N}$. Compared with the signal process X_t, X_t^n is a strong Feller process defined using the approximation Θ_n to W that is constructed using only the observations rather than the unknown path W.

Let E denote expectation with respect to P. For $\phi \in C(\overline{D})$, the filtering problem is to evaluate

$$\langle \Pi_j, \phi\rangle = E[\phi(X_{t_j})|\mathcal{Y}_j], \tag{3.2}$$

which is the least square estimate of $\phi(X_{t_j})$ given all the observations up to time t_j. We define for $j \in \mathbb{N}$,

$$\varsigma_j(x) = \exp\left\{h^T(x)\sigma_2^{-1}Y_j - \frac{h^T(x)\sigma_2^{-1}h(x)}{2}\right\}, \quad \forall x \in \overline{D}$$

and set $\eta_j = \prod_{i=1}^j \varsigma_i(X_{t_i})$. Then, we define a probability measure P^0 on $(\Omega, \mathcal{F})$ by

$$\left.\frac{dP^0}{dP}\right|_{\mathcal{F}_t} = \eta_{\lfloor t/\varepsilon\rfloor}^{-1}, \quad \forall t \geq 0,$$

where $\lfloor t/\varepsilon\rfloor$ denotes the greatest integer not more than t/ε and $\mathcal{F}_t := \sigma\{X_s, 0 \leq s \leq t\} \vee \mathcal{Y}_{\lfloor t/\varepsilon\rfloor} \vee \sigma\{B_s^2, 0 \leq s \leq t\}$. Let E^0 denote expectation with respect to P^0. Under P^0 the distribution of $\{X_{t_j}\}$ is the same as under P and $\{Y_j\}$ is a sequence of i.i.d. $N(0, \sigma_2)$ random vectors independent of $\{X_{t_j}\}$.

Owing to the Markov property of X_t, the optimal filter defined by Equation (3.2) satisfies the semigroup relation

$$\langle \Pi_j, \phi\rangle = \frac{E^0_{\{\Pi_{j-1}, Y_j\}}[\phi(X_{t_1})\varsigma_j(X_{t_1})]}{E^0_{\{\Pi_{j-1}, Y_j\}}[\varsigma_j(X_{t_1})]}, \quad \forall j \in \mathbb{N}. \tag{3.3}$$

Hereafter, we use the notation $E^0_{\{\Pi_{j-1}, Y_j\}}[F(X_{t_1}, Y_j)]$ for the conditional expectation (under P^0) of a function F of X_{t_1}, Y_j given the data Y_j, where the initial distribution of X is Π_{j-1}.

For $n \in \mathbb{N}$, we define the approximate filter $\Pi^n_\cdot$ using the recursive representation

$$\langle \Pi_j^n, \phi\rangle = \frac{E^2_{\{\Pi_{j-1}^n\}}[\phi(X_{t_1}^n)\varsigma_j(X_{t_1}^n)]}{E^2_{\{\Pi_{j-1}^n\}}[\varsigma_j(X_{t_1}^n)]}, \quad \forall j \in \mathbb{N}. \tag{3.4}$$

Hereafter, we use the notation $E^2_{\{\Pi_{j-1}^n\}}[F(X_{t_1}^n)]$ for the expectation (under P^2) of a function F of $X_{t_1}^n$, where the initial distribution of X^n is Π_{j-1}^n.

Now, we can state the main theoretical result of this section which shows that the pathwise average distance, per unit time, of the approximate filter from the optimal filter is asymptotically small in time. The proof of this limit result will be given in the Appendix.

Theorem 2. *Let the filtering model be as in Section 2. Define the approximate filter $\Pi^n_\cdot$ via Equation (3.4), where X^n_t satisfies Equation (3.1). Let $\{N_n\}$ be a sequence of positive integers satisfying $\lim_{n\to\infty} n/N_n = 0$. Then, for any $\phi \in C(\overline{D})$,*

$$\lim_{n\to\infty} \frac{1}{N_n} \sum_{j=1}^{N_n} \left[\langle \Pi^n_j, \phi \rangle - \langle \Pi_j, \phi \rangle \right]^2 = 0$$

in probability P.

The approximate filters defined by Equation (3.4) usually require excessive computations. So it is attractive to construct numerically feasible approximations via particle filters. In the remaining part of this section, we consider the simplest one which is based upon pure random sampling of the approximating process X^l_t. Here $l \in \mathbb{N}$, the order of the approximate Cesàro mean, is a suitably chosen large number. The basic algorithm for combined state and nonparametric estimation is as follows.

Initialization: Let $\{X^{i,l}_{t_j}\}_{i=1}^{M_l}$ be a set of particles, where M_l is the number of particles. In the initialization stage, each particle's state is independently initialized according to the uniform distribution on $\overline{D}$.

Evolution: In the evolution stage, each of the particles is independently evolved according to the approximate SDE of the signal as described in Equation (3.1). Here Θ_l is defined through the observation sequence:
(a) Let $m_l \in \mathbb{N}$ be a sufficiently large number. For a $q \in \mathbb{Z}^d_+$ satisfying $q_u \le l$ for any $1 \le u \le d$, we define

$$S_{l,q}(x) = \sum_{|k_u| \le q_u, 1 \le u \le d} \left(\frac{e^{2\pi^2 k^T \sigma_2 k}}{m_l} \sum_{j=1}^{m_l} e^{-2\pi i k^T Y_j} \right) e^{2\pi i k^T x}, \quad \forall x \in \overline{D}$$

and

$$\Lambda_l = \frac{1}{(l+1)^d} \sum_{|k_u| \le q_u, 1 \le u \le d} S_{l,q}.$$

(b) Let $g_l \in C^\infty(\overline{D})$ be an approximate function of h. We define

$$\Theta_l = -\ln\{(\Lambda_l \circ h) g_l\}.$$

Selection: In the selection stage, particles are weighted based upon their likelihood given the current observation. The approximate filter Π^l_j is defined by the sample average:

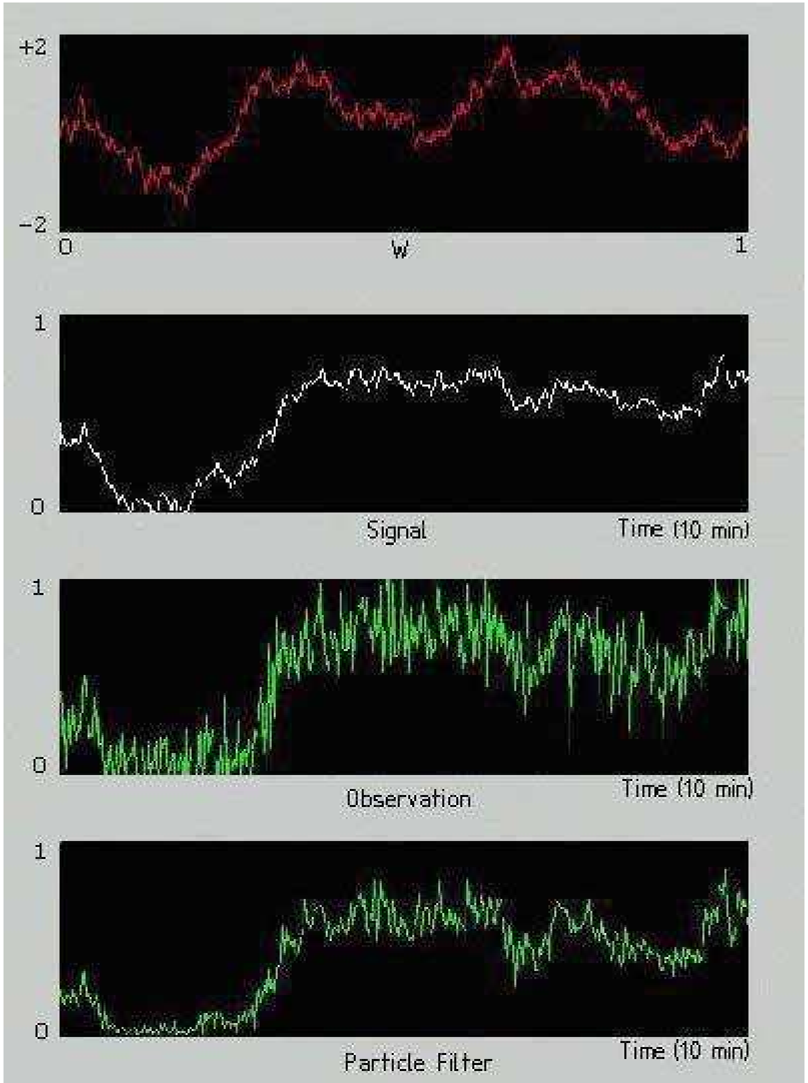
$$\langle \Pi^l_j, \phi \rangle = \frac{\sum_{i=1}^{M_l} \phi(X^{i,l}_{t_j}) \varsigma_j(X^{i,l}_{t_j})/M_l}{\sum_{i=1}^{M_l} \varsigma_j(X^{i,l}_{t_j})/M_l}.$$

The evolution and selection steps are repeated at each observation time.

Similarly as in the proof of Theorem 3.1 of Budhiraja and Kushner (2000b), one can show that the conclusion of Theorem 2 holds (cf. Appendix). Also, this scheme can be generalized in many ways. Common variance reduction methods such as antithetic variables and stratified sampling can be used. We refer the interested reader to Budhiraja and Kushner (2000b) for more details. Furthermore, one expects to improve the performance of the particle filter algorithm using many recently developed powerful interacting and branching particle filters.

4 Simulation Results

Simulation results based upon the above particle filter algorithm are presented below for one dimensional case. We set $W(x) = Z_x - xZ_1$, where Z is the standard $\mathbb{R}^1$-valued Brownian motion, and use the nonlinear sensor function $h(x) = \frac{1}{3}(1 + x - \cos \pi x)$. We specify the signal noise via $\sigma_1 = 0.1$, the observation noise via $\sqrt{\sigma_2} = 0.04$, and the time period between observations to be $\varepsilon = 1$ time unit. We let $l = 12$, $m_l = 1200$, and use $M_l = 2000$ particles for the approximate filter.



The curve in the first figure represents an unknown path of the random environment. The curves in the second and third figures represent the signal state and the observation, respectively. When the values of the observations are out of the interval $[0,1]$, we truncate them. The fourth figure gives the estimations of the signal based upon the weighted particle method. Here, the filter does not have access to the true random environment, but rather must contend with noisy estimates. The filter performs better as time goes on and we get improved functional estimation in environment. The results show that our nonparametric estimation method provides an effective solution to the nonlinear filtering problem for reflecting diffusions in random environments.

Appendix: Proof of Theorem 2

Many arguments in the proof are similar to those used in Theorem 2.1 of Budhiraja and Kushner (2000b); however, the proof is nontrivial considering the dependence of X^n and Y and the construction of our approximate filters. Also, we have to be very careful when dealing with asymptotic properties of nonlinear filters because of the gap recently found in Kunita's classic paper (1971) [cf. Budhiraja (2003)]. In the following, we will refer to the proof in Budhiraja and Kushner (2000b), and to concentrate on the differences.

Proof. The basic idea of the proof is to apply the method of occupation measures. For each $n, j \in \mathbb{N}$, we define the process

$$\Psi^n(j,\cdot) = (X_{j+\cdot}, \Pi^n_{j+\cdot}, Y_{j+\cdot} - Y_j, B_{j+\cdot} - B_j),$$

where $\Pi^n_\cdot$ is defined in Equation (3.4) and $B_j := \sum_{i=1}^{j} V_i$. For an arbitrary measurable set C in the product path space of $\Psi^n(j,\cdot)$, we define the measure-valued random variable $Q^{n,N_n}(\cdot)$:

$$Q^{n,N_n}(C) = \frac{1}{N_n}\sum_{j=1}^{N_n} I_C(\Psi^n(j,\cdot)),$$

where $I_C(\Psi^n(j,\cdot))$ denotes the indicator function of the event $\Psi^n(j,\cdot) \in C$. By the compactness of the state space, one can show that the families $\{X_{j+\cdot}, j \in \mathbb{N}\}$, $\{\Pi^n_{j+\cdot}, n, j \in \mathbb{N}\}$, $\{Y_{j+\cdot} - Y_j, j \in \mathbb{N}\}$, $\{B_{j+\cdot} - B_j, j \in \mathbb{N}\}$ are tight and thus the sequence $\{E[Q^{n,N_n}(\cdot)]\}$ is tight. Therefore, the sequence $\{Q^{n,N_n}(\cdot)\}$ of measure-valued random variables is tight by Theorem 1.6.1 of Kushner (1990).

We need to determine the sample values Q^ω of any weak sense limit $Q(\cdot)$. For each ω, Q^ω induces a process

$$\Psi^\omega_\cdot = (X^\omega_\cdot, \Pi^\omega_\cdot, Y^\omega_\cdot, B^\omega_\cdot).$$

The proof of the stationarity of the $(X^\omega_\cdot, \Pi^\omega_\cdot)$ in Budhiraja and Kushner (1999) will work without any change for our problem. By virtue of the assumption that $\lim_{n\to\infty} n/N_n = 0$, one can similarly prove that the relationship between $X^\omega_\cdot$ and $Y^\omega_\cdot$ is characterized by (1.2) as in Theorem 5.1 of Budhiraja and Kushner (1999). Also, the proof that X^ω has the law of evolution of $X_\cdot$ for almost all ω is the same as in Budhiraja and Kushner (1999).

Let $p^0(t,x,y)$ be the transition density function of X^0_t. By Theorem 2.4.4 of Davies (1989), we know that there exists $c > 0$ such that

$$0 \le p^0(t,x,y) \le ct^{-\frac{d}{2}}$$

almost everywhere on $D \times D$ for all $0 < t < 1$. Then, $(T_t^0)_{t>0}$, the L^2-semigroup associated with the Dirichlet form $\mathcal{E}^0$, is ultracontractive (cf. Lemma 2.1.2 of Davies (1989)). By the eigenfunction expansion of $p^0(t, x, y)$ (cf. Theorem 2.1.4 of Davies (1989)) and the argument as in the proof of Theorem 2.4 of Bass and Hsu (1991), one can see that there exist $T > 0$ and $c' > 0$ such that

$$\left| p^0(t, x, y) - \frac{1}{\mu(D)} \right| \leq e^{-c't}$$

almost everywhere on $D \times D$ for all $t \geq T$. In other words, $p^0(t, x, y)$ approaches the stationary distribution uniformly and exponentially. So by Equation (2.2) and the Remarks after Theorem 7.2 of Budhiraja and Kushner (1999), we conclude that the filter $\{\Pi_j\}$ forgets its initial condition. Thus, the signal-filter pair $\{X_{t_j}, \Pi_j\}$ has a unique probability invariant measure by Theorem 7.1 of Budhiraja and Kushner (1999) [cf. also Budhiraja (2003)].

By Theorem 1 and Equation (2.2) (cf. also the remark below A.2.1 of Budhiraja and Kushner (2000a) , one can see that for any sequence $\{\nu^n\}$ of probability measures on $\overline{D}$ converging weakly to some probability measure ν on $\overline{D}$, $(X_0^n, X_{t_1}^n)$ with the initial distribution ν^n converges weakly to (X_0, X_{t_1}) with the initial distribution ν. Following the idea of Theorem 2.1 of Budhiraja and Kushner (2000b), we can thus establish the representation (3.3) for almost all ω. Furthermore, we obtain as in Budhiraja and Kushner (1999) that

$$\lim_{n\to\infty} \frac{1}{N_n} \sum_{j=1}^{N_n} \left[\langle \Pi_j^n, \phi \rangle - \phi(X_{t_j}) \right]^2 = \lim_{n\to\infty} \frac{1}{N_n} \sum_{j=1}^{N_n} \left[\langle \Pi_j, \phi \rangle - \phi(X_{t_j}) \right]^2$$

in probability P for any $\phi \in C(\overline{D})$. The proof is therefore done by Theorem 5.2 of Budhiraja and Kushner (2000b).

Acknowledgments

The authors thank Xinjian Ma for providing the simulation results and gratefully acknowledge support from Lockheed Martin Naval Electronics and Surveillance Systems, NSERC, Lockheed Martin Canada, and PIMS through sponsorship of the MITACS-PINTS centre of excellence. Also, the authors are grateful to the anonymous referee for a careful reading of the manuscript resulting in many constructive comments.

References

Ballantyne D.J., Chan H.Y. and Kouritzin M.A. (2000). A novel branching particle method for tracking. *Signal and Data Processing of Small Targets 2000, Proceedings of SPIE* **4048** 277-287.

Bass R.F. and Hsu P. (1991). Some potential theory for reflecting Brownian motion in Hölder and Lipschitz domains. *Ann. Probab.* **19** 486–508.

Budhiraja A. (2003). Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter. *Ann. Inst. H. Poincaré Probab. Statist.* **39** 919–941.

Budhiraja A. and Kushner H. (1999). Approximation and limit results for nonlinear filters over an infinite time interval. *SIAM J. Cont. Opt.* **37** 1946–1979.

Budhiraja A. and Kushner H. (2000a) Approximation and limit results for nonlinear filters over an infinite time interval. II. Random sampling algorithms. *SIAM J. Cont. Opt.* **38** 1874–1908.

Budhiraja A. and Kushner H. (2000b). Monte Carlo algorithms and asymptotic problems in nonlinear filtering. In *Stochastics in Finite and Infinite Dimensions* 59–87. Trends Math., Birkhäuser, Boston.

Chow P.L., Khasminskii R. and Liptser, R. (1997). Tracking of signal and its derivatives in Gaussian white noise. *Stochastic Process. Appl.* **69** 259–273.

Chow P.L., Khasminskii R. and Liptser, R. (2001). On estimation of time dependent spatial signal in Gaussian white noise. *Stochastic Process. Appl.* **96** 161–175.

Davies E.B. (1989). *Heat Kernels and Spectral Theory.* Cambridge University Press.

Elliott R.J. (2001). *Financial Filtering.* Lecture Note. http://www.fields.utoronto.ca/programs/scientific/98-99/probability/graduate_courses/elliott2.ps

Freidlin M. (1985). *Functional Integration and Partial Differential Equations.* Princeton University Press.

Fukushima M., Oshima Y. and Takeda M. (1994). *Dirichlet Forms and Symmetric Markov Processes.* Walter de Gruyter, Berlin-New York.

Fukushima M. and Tomisaki M. (1996). Construction and decomposition of reflecting diffusions on Lipschitz domains with Hölder cusps. *Probab. Theory Related Fields* **106** 521–557.

Kouritzin M.A., Long H.W. and Sun W. (2003). Nonlinear filtering for diffusions in random environments. *J. Theoret. Probab.* **16** 1-20.

Kouritzin M.A., Rémillard B. and Chan, C. (2001). Parameter estimation for filtering problems with stable noise. *Proceedings of 4th Annual Conference on Information Fusion* Vol. I WeB127-WeB130.

Kunita H. (1971). Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal.* **1** 365-393.

Kunita H. (1982). Stochastic partial differential equations connected with nonlinear filtering. *Nonlinear Filtering and Stochastic Control. Lect. Notes Math.* Springer, **972** 100-169.

Kushner H. (1990). *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems.* Birkhäuser, Boston-Basel-Berlin.

Ledoux M. (2001). Logarithmic Sobolev inequalities for unbounded spin systems revisited. *Séminaire de Probabilités XXXV. Lect. Notes Math.* Springer, **1755** 1167-194.

Lototsky S., Mikulevicius R. and Rozovskii B.L. (1997). Nonlinear filtering revisited: a spectral approach. *SIAM J. Cont. Opt.* **35** 435-461.

Ocone D. (1983). Multiple integral expansions for nonlinear filtering. *Stochastics* **10** 1-30.

Rosenblatt M. (1971). *Markov Processes. Structure and Asymptotic Behavior.* Springer, Berlin-Heidelberg-New York.

Tanaka H. (1995). Diffusion processes in random environments. *Proceedings of the International Congress of Mathematicians*, Birkhäuser, 1047–1054.

Zygmund A. (1959) *Trigonometric Series, Vol. II.* Cambridge Univ. Press, New York.