Stochastic Particle Methods for Linear Tangent Filtering Equations

Frédéric Cérou and François LeGland
IRISA / INRIA, Campus de Beaulieu, 35042 Rennes Cédex, France

Nigel J. Newton
Department of Electronic Systems Engineering
University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England

This paper is dedicated to Alain Bensoussan, on the occasion of his 60th birthday.

Abstract : In this paper, an efficient particle approximation is proposed for the directional derivative of the optimal filter w.r.t. the drift coefficient in a partially observed diffusion process, under the assumption that variations of the drift coefficient are considered only in directions which are affected by nondegenerate noise.

Keywords : diffusion process, nonlinear filtering, reference probability, particle approximation.

1 Introduction

A common theme in many of the important contributions of Alain Bensoussan to stochastic control and nonlinear filtering, is the systematic use of the reference probability approach, see for example the monograph [1]. In this paper, the same very powerful approach is used to obtain a stochastic representation for the directional derivative of the optimal filter w.r.t. the drift coefficient of a partially observed diffusion process, under the assumption that variations of the drift coefficient are considered in directions which are affected by nondegenerate noise, see Assumptions A and B below. This representation shows that the directional derivative of the optimal filter, i.e. the solution of the linear tangent filtering equation, is actually a signed measure absolutely continuous w.r.t. the optimal filter, for which an efficient stochastic particle approximation can be obtained.

This work is motivated by applications in statistics. It is well–known that the likelihood function for the estimation of some unknown parameter in the drift coefficient of a partially observed diffusion process, and many other statistics such as the conditional least squares criterion, can be expressed in terms of the optimal filter. Likewise, the score function, i.e. the derivative of the log–likelihood function w.r.t. the parameter, and many other estimating functions, can be expressed in terms of the derivative of the optimal filter w.r.t. the parameter. Some details about a particular application in statistics can be found in Cérou and LeGland [3], where the method is applied to the solution of the *residual generation* problem, for the purpose of model validation and change detection.

2 Partially observed model

On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider the following model : The hidden state process $\{X_t\,,\, 0 \leq t \leq T\}$ is the solution of the following stochastic differential equation (SDE) on $\mathbb{R}^m$

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t \ , \qquad X_0 \sim \mu_0(dx) \ , \tag{1}$$

where $\{W_t\,,\, 0 \leq t \leq T\}$ is a r–dimensional Wiener process, $1 \leq r \leq m$, with identity covariance matrix, independent of the initial state X_0. The coefficients b and σ are locally Lipschitz continuous on $\mathbb{R}^m$, and the following (sufficient) condition for nonexplosion holds : for any $x \in \mathbb{R}^m$

$$x^*\, b(x) + \tfrac{1}{2}\, \text{trace}\; a(x) \leq K\, (1 + |x|^2) \ ,$$

where $a = \sigma\, \sigma^*$. These conditions ensure the existence of a unique strong solution of equation (1).

At discrete time instants

$$0 = t_0 < \cdots < t_n = T$$

a d–dimensional noisy observation z_k of the state X_{t_k} becomes available, with conditional probability distribution

$$\mathbb{P}[z_k \in dz \mid X_{t_k} = x] = g_k(x,z)\, dz \ .$$

The likelihood function for the estimation of the state X_{t_k} based on the observation z_k alone is defined by

$$\Psi_k(x) = g_k(x, z_k) \ .$$

The *memoryless channel* assumption holds, under which the observations $\{z_0, \cdots, z_n\}$ are independent given the corresponding states $\{X_{t_0}, \cdots, X_{t_n}\}$. This assumption holds for example in the case of observations in additive white noise, i.e.

$$z_k = h_k(X_{t_k}) + v_k \ ,$$

where $\{v_0, \cdots, v_n\}$ is a white noise sequence, independent of the states $\{X_{t_0}, \cdots, X_{t_n}\}$. In the special case of a Gaussian white noise sequence with identity covariance matrix, it holds that

$$\Psi_k(x) = (2\pi)^{-d/2}\, \exp\{ -\tfrac{1}{2}\, |z_k - h_k(x)|^2 \} \ .$$

The time–dependent transition probability kernel for the sampled Markov chain $\{X_{t_0}, \cdots, X_{t_n}\}$ is defined by

$$Q_k(x, dx') = \mathbb{P}[X_{t_k} \in dx' \mid X_{t_{k-1}} = x] \ ,$$

or equivalently

$$Q_k\, \phi(x) = \mathbb{E}[\phi(X_{t_k}) \mid X_{t_{k-1}} = x] \ , \tag{2}$$

for any test function ϕ defined on $\mathbb{R}^m$.

3 Optimal filtering equations and associated linear tangent equations

Let $\mathcal{Z}_k = (z_0, \cdots, z_k)$ denote the sequence of observations up to time t_k, and introduce the following conditional probability distributions of the state X_{t_k}

$$\mathbb{P}[X_{t_k} \in dx \mid \mathcal{Z}_{k-1}] = \mu_{k|k-1}(dx) \qquad \text{and} \qquad \mathbb{P}[X_{t_k} \in dx \mid \mathcal{Z}_k] = \mu_k(dx) \; .$$

The sequence $\{\mu_0, \cdots, \mu_n\}$ takes values in the space $\mathcal{P} = \mathcal{P}(\mathbb{R}^m)$ of probability distributions on $\mathbb{R}^m$. The transition from μ_{k-1} to μ_k is described by the following two steps :

$$\mu_{k-1} \xrightarrow{\text{prediction}} \mu_{k|k-1} = Q_k^* \, \mu_{k-1} \xrightarrow{\text{correction}} \mu_k = \Psi_k \cdot \mu_{k|k-1} \; ,$$

where $\cdot$ denotes the projective product. The following representation holds for the prediction step : for any probability distribution μ on $\mathbb{R}^m$, and any test function ϕ defined on $\mathbb{R}^m$

$$\langle Q_k^* \, \mu, \phi \rangle = \langle \mu, Q_k \, \phi \rangle = \int_{\mathbb{R}^m} \mathbb{E}[\phi(X_{t_k}) \mid X_{t_{k-1}} = x] \, \mu(dx) \; .$$

In the correction step, μ_k is given by the Bayes rule

$$\mu_k = \Psi_k \cdot \mu_{k|k-1} = \frac{\Psi_k \; \mu_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle} \; .$$

Remark 3.1 By construction, the prediction step conserves the total mass, whereas the correction step automatically returns a probability distribution, i.e. for any positive measure μ on $\mathbb{R}^m$

$$\langle Q_k^* \, \mu, 1 \rangle = \langle \mu, Q_k \, 1 \rangle = \langle \mu, 1 \rangle \; ,$$

and

$$\langle \Psi_k \cdot \mu, 1 \rangle = 1 \; ,$$

respectively.

Let $w_{k|k-1} = \partial \, \mu_{k|k-1}$ and $w_k = \partial \, \mu_k$ denote the directional derivative of the conditional probability distributions $\mu_{k|k-1}$ and μ_k respectively, w.r.t. the drift coefficient b. Derivability, with directional derivatives taking values in the linear tangent space to $\mathcal{P}$, i.e. in the space Σ of signed measures with zero mass, follows from Theorem 4.2 below. The transition from w_{k-1} to w_k is described by the following two steps, which are linear tangent to the prediction step and to the correction step respectively :

$$w_{k-1} \xrightarrow[\text{prediction}]{\text{linear tangent}} w_{k|k-1} = \partial \, [Q_k^* \, \mu_{k-1}] \xrightarrow[\text{correction}]{\text{linear tangent}} w_k = \partial \, [\Psi_k \cdot \mu_{k|k-1}] \; ,$$

In the linear tangent prediction step, $w_{k|k-1}$ is given by

$$w_{k|k-1} = Q_k^* \, w_{k-1} + [\partial \, Q_k^*] \, \mu_{k-1} \; . \tag{3}$$

In the linear tangent correction step, w_k is given by the linear tangent Bayes rule

$$
\begin{aligned}
w_k = \partial\,[\Psi_k \cdot \mu_{k|k-1}] &= \frac{\Psi_k \; w_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle} - \frac{\Psi_k \; \mu_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle} \, \frac{\langle w_{k|k-1}, \Psi_k \rangle}{\langle \mu_{k|k-1}, \Psi_k \rangle} \\
&= F_k(\mu_{k|k-1})\, w_{k|k-1} \;,
\end{aligned} \tag{4}
$$

where $w \mapsto F_k(\mu)\, w$ is the linear tangent map at point $\mu \in \mathcal{P}$ of the map $\mu \mapsto \Psi_k \cdot \mu$. Of course this map can be extended to any signed measure on $\mathbb{R}^m$, and the following identity holds

$$
F_k(\mu)\, \mu = 0 \qquad \text{hence} \qquad F_k(\mu)\, (w - c\, \mu) = F_k(\mu)\, w \;,
$$

for any scalar c. A practical consequence of this simple remark is the following.

Lemma 3.2 *Let μ be a probability distribution on $\mathbb{R}^m$, and let w be a signed measure on $\mathbb{R}^m$ with nonzero mass, i.e. $\langle w, 1 \rangle \neq 0$. The modified signed measure $w' = w - \langle w, 1 \rangle\, \mu$*

- *has zero mass, i.e. $\langle w', 1 \rangle = 0$,*
- *has the same image as w under the linear tangent map $F_k(\mu)$, i.e.*

$$
F_k(\mu)\, w' = F_k(\mu)\, w \;.
$$

Remark 3.3 By construction, the linear tangent prediction step conserves the total mass, whereas the linear tangent correction step automatically returns a signed measure with zero mass, i.e. for any positive measure μ on $\mathbb{R}^m$, and any signed measure w on $\mathbb{R}^m$

$$
\langle Q_k^*\, w, 1 \rangle + \langle [\partial\, Q_k]^*\, \mu, 1 \rangle = \langle w, Q_k\, 1 \rangle + \langle \mu, [\partial\, Q_k]\, 1 \rangle = \langle w, 1 \rangle \;,
$$

and

$$
\langle F_k(\mu)\, w, 1 \rangle = 0 \;,
$$

respectively.

4 Stochastic representation result

The next result provides a stochastic representation for the derivative $\partial\, Q_k$ of the transition probability kernel Q_k w.r.t. the drift coefficient b in the direction γ, where the coefficient γ is locally Lipschitz continuous on $\mathbb{R}^m$, and bounded (for simplicity).

Assumption A : For any $x \in \mathbb{R}^m$, the $m \times r$ matrix $\sigma(x)$ has full–rank r, and the $r \times r$ symmetric matrix $\sigma^*(x)\, \sigma(x)$ is uniformly definite positive, i.e. there exists a positive constant $\alpha > 0$, such that

$$
\sigma^*(x)\, \sigma(x) \geq \alpha\, I \;.
$$

Assumption B : For any $x \in \mathbb{R}^m$, the m–dimensional vector $\gamma(x)$ belongs to the range of the $m \times r$ matrix $\sigma(x)$, i.e. $\gamma(x) = \sigma(x)\, c(x)$ for some r–dimensional vector $c(x)$.

Remark 4.1 Under Assumption A, the r–dimensional vector $c(x)$ such that $\gamma(x) = \sigma(x)\, c(x)$ holds is unique, and

$$c(x) = [\sigma^*(x)\, \sigma(x)]^{-1}\, \sigma^*(x)\, \gamma(x) = \sigma^*(x)\, a^\ominus(x)\, \gamma(x) ,$$

where

$$a^\ominus(x) = \sigma(x)\, [\sigma^*(x)\, \sigma(x)]^{-2}\, \sigma^*(x)$$

is the pseudoinverse of $a(x) = \sigma(x)\, \sigma^*(x)$. In addition

$$|c(x)|^2 = \gamma^*(x)\, a^\ominus(x)\, \gamma(x) .$$

Notice also that

$$|\gamma(x)|^2 \geq \alpha\, |c(x)|^2 .$$

Theorem 4.2 *Under Assumptions A and B, the following representation holds*

$$[\partial\, Q_k]\, \phi(x) = \mathbb{E}[\phi(X_{t_k})\; S_{t_k,t_{k-1}} \mid X_{t_{k-1}} = x] , \tag{5}$$

for any test function ϕ defined on $\mathbb{R}^m$, where

$$S_{t,s} = \int_s^t c^*(X_u)\, dW_u = \int_s^t \gamma^*(X_u)\, a^\ominus(X_u)\, \sigma(X_u)\, dW_u .$$

A similar representation result has been obtained in Campillo and LeGland [2, Section 3.1], under the stronger assumption that the diffusion matrix $a = \sigma\, \sigma^*$ is invertible. The proof of Theorem 4.2 uses the reference probability approach, and follows the same lines as the proof of Proposition 3.1 in Fournié et al. [6], where the stronger assumption on the invertibility of the diffusion matrix $a = \sigma\, \sigma^*$ is made.

Proof. Let $\mathbb{P}^\varepsilon$ denote the probability measure on $(\Omega, \mathcal{F})$, absolutely continuous w.r.t. $\mathbb{P}$, with Radon–Nikodym derivative

$$Z^\varepsilon_{t,s} = \exp\{\varepsilon \int_s^t c^*(X_u)\, dW_u - \tfrac{1}{2}\, \varepsilon^2 \int_s^t |c(X_u)|^2\, du \} .$$

It follows from the Girsanov theorem that under $\mathbb{P}^\varepsilon$ the process

$$W^\varepsilon = W - \varepsilon \int_0^\bullet c(X_t)\, dt$$

is a r–dimensional Wiener process, with identity covariance matrix, independent of the initial state X_0. Substitution into (1) yields

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, [dW^\varepsilon_t + \varepsilon\, c(X_t)\, dt] = [b(X_t) + \varepsilon\, \gamma(X_t)]\, dt + \sigma(X_t)\, dW^\varepsilon_t ,$$

which is consistent with the statement in Liptser and Shiryayev [8, Section 7.6.4].

The transition probability kernel Q_k^ε corresponding to the perturbed drift $(b + \varepsilon\,\gamma)$ satisfies

$$Q_k^\varepsilon\,\phi(x) = \mathbb{E}^\varepsilon[\phi(X_{t_k}) \mid X_{t_{k-1}} = x] = \mathbb{E}[\phi(X_{t_k})\, Z^\varepsilon_{t_k,t_{k-1}} \mid X_{t_{k-1}} = x]\ ,$$

for any test function ϕ defined on $\mathbb{R}^m$, hence

$$\frac{1}{\varepsilon}\,[Q_k^\varepsilon - Q_k - \varepsilon\,\partial\,Q_k]\,\phi(x) = \mathbb{E}[\phi(X_{t_k})\,[\frac{1}{\varepsilon}\,(Z^\varepsilon_{t_k,t_{k-1}} - 1) - S_{t_k,t_{k-1}}] \mid X_{t_{k-1}} = x]\ ,$$

and the result follows, exactly as in the proof of [6, Proposition 3.1]. □

Proposition 4.3 *For any $k \geq 0$, the signed measures $w_{k|k-1}$ and w_k are absolutely continuous w.r.t. the probability distributions $\mu_{k|k-1}$ and μ_k respectively.*

PROOF. The property is true for $k = 0$, since $w_0 = 0$, and the proof proceeds by induction w.r.t. the time index k.

Assume that w_{k-1} is absolutely continuous w.r.t. μ_{k-1}. Clearly, $Q_k^*\,w_{k-1}$ is absolutely continuous w.r.t. $Q_k^*\,\mu_{k-1} = \mu_{k|k-1}$. On the other hand, it follows from the representation (5) that $[\partial\,Q_k]^*\,\mu_{k-1}$ is absolutely continuous w.r.t. $Q_k^*\,\mu_{k-1} = \mu_{k|k-1}$. Combining these two facts and using (3) yields that $w_{k|k-1}$ is absolutely continuous w.r.t. $\mu_{k|k-1}$.

Assume now that $w_{k|k-1}$ is absolutely continuous w.r.t. $\mu_{k|k-1}$. Using (4) yields immediately that w_k is absolutely continuous w.r.t. μ_k. □

Combining the representations (2) and (5) yields the following representation for the linear tangent prediction step : for any probability distribution μ on $\mathbb{R}^m$, any signed measure w on $\mathbb{R}^m$ absolutely continuous w.r.t. μ, and any test function ϕ defined on $\mathbb{R}^m$

$$\begin{aligned}\langle Q_k^*\,w, \phi\rangle + \langle[\partial\,Q_k]^*\,\mu, \phi\rangle &= \langle w, Q_k\,\phi\rangle + \langle\mu, [\partial\,Q_k]\,\phi\rangle \\ &= \int_{\mathbb{R}^m} \mathbb{E}[\phi(X_{t_k}) \mid X_{t_{k-1}} = x]\,\frac{dw}{d\mu}(x)\,\mu(dx) \\ &\quad + \int_{\mathbb{R}^m} \mathbb{E}[\phi(X_{t_k})\,S_{t_k,t_{k-1}} \mid X_{t_{k-1}} = x]\,\mu(dx)\ .\end{aligned} \tag{6}$$

5 Particle approximation of the optimal filter

A question that naturally arises is whether the optimal filtering equations have any practical use : since the Bayes rule is rather straightforward, the question reduces to find an efficient approximation scheme for the prediction step. For this purpose, a new class of approximate nonlinear filters has been recently proposed, under the name of *particle filters*, where the idea is to generate a sample $\{\xi^i_{k|k-1}\,,\ i \in I\}$ of i.i.d. random variables, called a *particle system*, with common probability distribution $Q_k^*\,\bar\mu_{k-1}$, where $\bar\mu_{k-1}$ is an approximation of μ_{k-1}, and to use the corresponding empirical probability distribution

$$\bar\mu_{k|k-1} = \frac{1}{|I|}\sum_{i \in I} \delta_{\xi^i_{k|k-1}}\ ,$$

as an approximation of $\mu_{k|k-1} = Q_k^* \, \mu_{k-1}$. The method is very easy to implement, even in high dimensional problems, since it is sufficient in principle to simulate independent sample paths of the hidden dynamical system. A major and earliest contribution in this field was made by Gordon, Salmond and Smith [7], which proposed to use sampling / importance resampling (SIR) techniques in the correction step : the positive effect of the resampling step is to automatically concentrate particles in regions of interest of the state space. A very complete account of the currently available mathematical results can be found in the survey paper by Del Moral and Miclo [4]. Theoretical and practical aspects can be found in the volume edited by Doucet, de Freitas and Gordon [5].

Algorithm : The particle approximation $\bar{\mu}_{k|k-1}$ to the optimal filter $\mu_{k|k-1}$ is completely characterized by the particle system $\{\xi^i_{k|k-1} \, , \, i \in I\}$. The transition from the particle system $\{\xi^i_{k|k-1} \, , \, i \in I\}$ to the particle system $\{\xi^i_{k+1|k} \, , \, i \in I\}$ is decribed by the following three steps :

(i) Correction : for all $i \in I$, compute the weight

$$\omega^i_k = \frac{\Psi_k(\xi^i_{k|k-1})}{\displaystyle\sum_{j \in I} \Psi_k(\xi^j_{k|k-1})} \; .$$

Then set

$$\bar{\mu}_k = \Psi_k \cdot \bar{\mu}_{k|k-1} = \sum_{i \in I} \omega^i_k \, \delta_{\xi^i_{k|k-1}} \; .$$

(ii) Resampling : independently for all $i \in I$, generate a random variable ξ^i_k with discrete probability distribution $\bar{\mu}_k$. A practical way to achieve this, is to generate i.i.d. random variables $\{\tau(i) \, , \, i \in I\}$ with values in I and with common probability distribution $\{\omega^i_k \, , \, i \in I\}$, and to set

$$\xi^i_k = \xi^{\tau(i)}_{k|k-1} \; ,$$

for all $i \in I$.

(iii) Prediction : independently for all $i \in I$, generate a random variable $\xi^i_{k+1|k}$ as the value taken at time t_{k+1} by the solution of the SDE

$$dX^i_t = b(X^i_t) \, dt + \sigma(X^i_t) \, dW^i_t \; , \qquad X^i_{t_k} = \xi^i_k \; .$$

Then set

$$\bar{\mu}_{k+1|k} = \frac{1}{|I|} \sum_{i \in I} \delta_{\xi^i_{k+1|k}} \; .$$

6 Particle approximation of the linear tangent optimal filter

Because the signed measures $w_{k|k-1}$ and w_k are absolutely continuous w.r.t. the probability distributions $\mu_{k|k-1}$ and μ_k respectively, see Proposition 4.3, it is natural to approximate $w_{k|k-1}$ and w_k by weighted empirical probability distributions associated with the same particle system $\{\xi^i_{k|k-1}\,,\, i \in I\}$ already used for the particle approximations of $\mu_{k|k-1}$ and μ_k. To be more specific, since

$$\bar{\mu}_{k|k-1} = \frac{1}{|I|} \sum_{i \in I} \delta_{\xi^i_{k|k-1}} \;,$$

then it is natural to set

$$\bar{w}_{k|k-1} = \frac{1}{|I|} \sum_{i \in I} \rho^i_{k|k-1}\, \delta_{\xi^i_{k|k-1}} \;,$$

so that the particle approximation $\bar{w}_{k|k-1}$ is absolutely continuous w.r.t. the particle approximation $\bar{\mu}_{k|k-1}$. Since the signed measure $w_{k|k-1}$ has zero mass, it is natural to ask that the particle approximation $\bar{w}_{k|k-1}$ has zero mass as well, which is obtained by requiring

$$\sum_{i \in I} \rho^i_{k|k-1} = 0 \;.$$

Algorithm : The particle approximations $\bar{\mu}_{k|k-1}$ and $\bar{w}_{k|k-1}$ to the optimal filter $\mu_{k|k-1}$ and its directional derivative $w_{k|k-1}$ w.r.t. the drift coefficient b are completely characterized by the particle and weight system $\{\xi^i_{k|k-1}, \rho^i_{k|k-1}\,,\, i \in I\}$. The transition from the particle and weight system $\{\xi^i_{k|k-1}, \rho^i_{k|k-1}\,,\, i \in I\}$ to the particle and weight system $\{\xi^i_{k+1|k}, \rho^i_{k+1|k}\,,\, i \in I\}$ is described by the following three steps :

(i) Correction : for all $i \in I$, compute the weights

$$\omega^i_k = \frac{\Psi_k(\xi^i_{k|k-1})}{\displaystyle\sum_{j \in I} \Psi_k(\xi^j_{k|k-1})} \;.$$

Then set

$$\bar{\mu}_k = \Psi_k \cdot \bar{\mu}_{k|k-1} = \sum_{i \in I} \omega^i_k\, \delta_{\xi^i_{k|k-1}} \;,$$

and

$$\bar{w}_k = F_k(\bar{\mu}_{k|k-1})\, \bar{w}_{k|k-1} = \sum_{i \in I} [\, \rho^i_{k|k-1} - \sum_{j \in I} \rho^j_{k|k-1}\, \omega^j_k \,]\; \omega^i_k\, \delta_{\xi^i_{k|k-1}} \;.$$

(ii) Resampling : independently for all $i \in I$, generate a random variable ξ^i_k with discrete probability distribution $\bar{\mu}_k$. A practical way to achieve this, is to generate i.i.d. random variables $\{\tau(i)\,,\, i \in I\}$ with values in I and with common probability distribution $\{\omega^i_k\,,\, i \in I\}$, and to set

$$\xi^i_k = \xi^{\tau(i)}_{k|k-1} \qquad \text{and} \qquad \rho^i_k = \rho^{\tau(i)}_{k|k-1} - \frac{1}{|I|} \sum_{j \in I} \rho^{\tau(j)}_{k|k-1} \;,$$

for all $i \in I$.

(iii) Prediction : independently for all $i \in I$, generate a random variable $(\xi^i_{k+1|k}, \Xi^i_{k+1|k})$ as the value taken at time t_{k+1} by the solution of the coupled SDE's

$$dX^i_t = b(X^i_t)\,dt + \sigma(X^i_t)\,dW^i_t \;, \qquad X^i_{t_k} = \xi^i_k \;,$$

$$dS^i_t = \gamma^*(X^i_t)\,a^\ominus(X^i_t)\,\sigma(X^i_t)\,dW^i_t \;, \qquad S^i_{t_k} = 0 \;.$$

Then set

$$\bar\mu_{k+1|k} = \frac{1}{|I|}\sum_{i\in I}\delta_{\xi^i_{k+1|k}} \qquad \text{and} \qquad \bar w_{k+1|k} = \frac{1}{|I|}\sum_{i\in I}\rho^i_{k+1|k}\,\delta_{\xi^i_{k+1|k}} \;,$$

where

$$\rho^i_{k+1|k} = \rho^i_k + \Xi^i_{k+1|k} \;,$$

for all $i \in I$.

Remark 6.1 If the approximations $\bar\mu_k$ and $\bar w_k$ are used in (6), then

$$\begin{aligned}&\langle Q^*_{k+1}\,\bar w_k, \phi\rangle + \langle[\partial\,Q_{k+1}]^*\,\bar\mu_k, \phi\rangle \\ &= \sum_{i\in I}\mathbb{E}[\phi(X_{t_{k+1}}) \mid X_{t_k} = \xi^i_{k|k-1}]\,[\rho^i_{k|k-1} - \sum_{j\in I}\rho^j_{k|k-1}\,\omega^j_k]\,\omega^i_k \\ &\quad + \sum_{i\in I}\mathbb{E}[\phi(X_{t_{k+1}})\,S_{t_{k+1},t_k} \mid X_{t_k} = \xi^i_{k|k-1}]\,\omega^i_k \;,\end{aligned}$$

for any test function ϕ defined on $\mathbb{R}^m$, which explains the steps (ii) and (iii).

Remark 6.2 In practice, it may happen that

$$\langle\bar w_{k|k-1}, 1\rangle = \frac{1}{|I|}\sum_{i\in I}\rho^i_{k-1} + \frac{1}{|I|}\sum_{i\in I}\Xi^i_{k|k-1} = \frac{1}{|I|}\sum_{i\in I}\Xi^i_{k|k-1} \neq 0 \;.$$

However, thanks to Lemma 3.2, it is harmless to use the following modified definition

$$\rho^i_{k|k-1} = \rho^i_{k-1} + [\,\Xi^i_{k|k-1} - \frac{1}{|I|}\sum_{j\in I}\Xi^j_{k|k-1}\,] \;,$$

for all $i \in I$.

7 Conclusion

The proposed particle approximation of the linear tangent optimal filter is especially attractive, since it uses the same particle system already used in the approximation of the optimal filter. Only one–dimensional weights are needed in addition.

A direction of future research is the derivation of error estimates for the particle approximation of the linear tangent optimal filter, as the number of particles goes to infinity. Concerning the statistical application to model validation and change detection, a direction of future research is the solution of the *residual evaluation* problem, using the local asymptotic approach, i.e. the characterization of the asymptotic behaviour of the residual under the null (nominal) hypothesis and under a contiguous alternative hypothesis, in the small noise asymptotics.

Acknowledgment

This work was partially supported by the Commission of the European Communities, under the TMR project *System Identification*, project number FMRX–CT98–0206, and under the IHP research training network *Statistical Methods for Dynamical Stochastic Models*, network number HPRN–CT2000–00100, and by the CNRS programme *Modélisation et Simulation Numérique*.

In addition, the authors gratefully acknowledge Dan Crişan, Pierre Del Moral and Arnaud Doucet for giving the opportunity to present a preliminary version of this work at the *Workshop on Practical and Theoretical Aspects of Particle Filters*, organized in Cambridge in December 1999.

References

[1] A. Bensoussan. *Stochastic Control of Partially Observable Systems.* Cambridge University Press, Cambridge, 1992.

[2] F. Campillo and F. Le Gland. MLE for the partially observed diffusions : direct maximization vs. the EM algorithm. *Stochastic Processes and their Applications*, 33(2):245–274, 1989.

[3] F. Cérou and F. Le Gland. Efficient particle filters for residual generation in partially observed SDE's. In *Proceedings of the 39th Conference on Decision and Control, Sydney 2000.* IEEE–CSS, Dec. 2000. Paper TuP08–3.

[4] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering. In J. Azéma, M. Émery, M. Ledoux, and M. Yor, editors, *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pages 1–145. Springer–Verlag, Berlin, 2000.

[5] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice.* Springer–Verlag, New York, 2001.

[6] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3(4):391–412, 1999.

[7] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear / non–Gaussian Bayesian state estimation. *IEE Proceedings, Part F*, 140(2):107–113, Apr. 1993.

[8] R. S. Liptser and A. N. Shiryayev. *Statistics of Random Processes I. General Theory*, volume 5 of *Applications of Mathematics.* Springer–Verlag, New York, 1977.