

Efficient Particle Methods for Residual Generation in Partially Observed SDE's*

Frédéric Cérou and François LeGland
IRISA / INRIA
Campus de Beaulieu
35042 RENNES Cédex, France
e-mail : {fcerou, legland}@irisa.fr

Abstract

In this paper, the problem of detecting a change in the drift coefficient of a partially observed stochastic differential equation is addressed. The score function, evaluated at the nominal value, is used as the residual, and only the problem of residual generation is considered. In the special case where the drift coefficient depends on the parameter only in directions that are affected by nondegenerate noise, an efficient numerical approximation of the residual is proposed, using particle filters. The more complicated problem of residual evaluation will be considered elsewhere, under the small noise asymptotics.

Keywords : residual generation, stochastic differential equation, nonlinear filter, particle filter, likelihood function, score function.

1 Introduction

The purpose of this paper is to design statistical tests to decide whether $\theta = \alpha$, corresponding to a nominal behaviour of the system, or $\theta \neq \alpha$, on the basis of observations (z_0, \dots, z_n) , in the following parametric model.

- Under \mathbb{P}^θ , 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

$$dX_t = b_\theta(X_t) dt + \sigma(X_t) dW_t, \quad (1)$$

with $X_0 \sim \mu_0(dx)$ and 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 .

*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*.

- At discrete time instants

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

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

$$\mathbb{P}^\theta[z_k \in dz \mid X_{t_k} = x] = g_k(x, z) dz, \quad (2)$$

independent of θ . 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 here, under which the observations $\{z_0, \dots, z_n\}$ are i.i.d. given the corresponding states $\{X_{t_0}, \dots, X_{t_n}\}$. 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 $\{v_0, \dots, v_n\}$ is a white noise sequence, independent of the states $\{X_{t_0}, \dots, X_{t_n}\}$. In the special case of a Gaussian white noise sequence with identity covariance matrix, it holds

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

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

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

or equivalently

$$Q_k^\theta \phi(x) = \mathbb{E}^\theta[\phi(X_{t_k}) \mid X_{t_{k-1}} = x], \quad (3)$$

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

The first (nonlocal) approach to detect a change, is to use a likelihood ratio test. For each value of the parameter the likelihood function can be expressed in terms

of the prediction filter, i.e. the probability distribution of the hidden state given past observations, and a numerical approximation is proposed here, using particle filters. The behaviour of the log-likelihood ratio under the null (nominal) hypothesis or under the alternative hypothesis, could be derived in the small noise asymptotics, as in James and LeGland [10], see also Campillo and LeGland [4] where the change detection problem is considered.

Following Benveniste, Basseville and Moustakides [2], see also Zhang, Basseville and Benveniste [12] and Basseville [1], the second (local) approach consists of the following two steps. The first step, called *residual generation*, is to propose a statistics, called the *residual*, which depends only the observations and on the nominal value of the parameter, and which ideally should be close to zero under the null hypothesis, and significantly different from zero under the alternative hypothesis. The second step, called *residual evaluation*, is to study the asymptotic behaviour of the residual under the null hypothesis and under a contiguous alternative hypothesis, and to design a simple test based on the residual. In this paper, only the problem of residual generation is considered, and the score function, i.e. the derivative of the log-likelihood function w.r.t. the parameter, evaluated at the nominal value, is used as the residual. In the special case where the drift coefficient depends on the parameter only in directions that are affected by nondegenerate noise, see Assumptions A and B below, an efficient numerical approximation is proposed, using particle filters. The asymptotic behaviour of the residual under the null (nominal) hypothesis or under a contiguous alternative hypothesis, could be derived in the small noise asymptotics, and will be considered elsewhere.

2 Optimal filtering equations and their linear tangent equations

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

$$\mathbb{P}^\theta[X_{t_k} \in dx \mid \mathcal{Z}_{k-1}] = \mu_{k|k-1}^\theta(dx),$$

$$\mathbb{P}^\theta[X_{t_k} \in dx \mid \mathcal{Z}_k] = \mu_k^\theta(dx).$$

The sequence $\{\mu_0^\theta, \dots, \mu_n^\theta\}$ 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}^θ to μ_k^θ is described by the following two steps :

$$\begin{aligned} \mu_{k-1}^\theta &\xrightarrow{\text{prediction}} \mu_{k|k-1}^\theta = [Q_k^\theta]^* \mu_{k-1}^\theta \\ &\xrightarrow{\text{correction}} \mu_k^\theta = \Psi_k \cdot \mu_{k|k-1}^\theta, \end{aligned}$$

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

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

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

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

Remark 2.1 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 ,

$$\langle [Q_k^\theta]^* \mu, 1 \rangle = \langle \mu, Q_k^\theta 1 \rangle = \langle \mu, 1 \rangle,$$

and

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

respectively.

Let $w_{k|k-1}^\theta = \partial \mu_{k|k-1}^\theta$ and $w_k^\theta = \partial \mu_k^\theta$ denote the derivative of the conditional probability distributions $\mu_{k|k-1}^\theta$ and μ_k^θ respectively, w.r.t. the p -dimensional parameter θ . The sequence $\{w_0^\theta, \dots, w_n^\theta\}$ takes values in the linear tangent space to \mathcal{P} , i.e. in the space Σ of signed measures with zero mass. The transition from w_{k-1}^θ to w_k^θ is described by the following two steps, which are linear tangent to the prediction step and to the correction step respectively :

$$\begin{array}{c} \text{linear tangent} \\ \text{prediction} \\ w_{k-1}^\theta \longrightarrow w_{k|k-1}^\theta = \partial [[Q_k^\theta]^* \mu_{k-1}^\theta] \end{array}$$

$$\begin{array}{c} \text{linear tangent} \\ \text{correction} \\ \longrightarrow w_k^\theta = \partial [\Psi_k \cdot \mu_{k|k-1}^\theta], \end{array}$$

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

$$w_{k|k-1}^\theta = [Q_k^\theta]^* w_{k-1}^\theta + [\partial Q_k^\theta]^* \mu_{k-1}^\theta. \quad (4)$$

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

$$\begin{aligned} w_k^\theta &= \partial [\Psi_k \cdot \mu_{k|k-1}^\theta] \\ &= \frac{\Psi_k w_{k|k-1}^\theta}{\langle \mu_{k|k-1}^\theta, \Psi_k \rangle} - \frac{\Psi_k \mu_{k|k-1}^\theta}{\langle \mu_{k|k-1}^\theta, \Psi_k \rangle} \frac{\langle w_{k|k-1}^\theta, \Psi_k \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k \rangle} \\ &= F_k(\mu_{k|k-1}^\theta) w_{k|k-1}^\theta, \end{aligned} \quad (5)$$

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

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

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

Lemma 2.2 *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 2.3 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

$$\begin{aligned} & \langle [Q_k^\theta]^* w, 1 \rangle + \langle [\partial Q_k^\theta]^* \mu, 1 \rangle \\ &= \langle w, Q_k^\theta 1 \rangle + \langle \mu, [\partial Q_k^\theta] 1 \rangle = \langle w, 1 \rangle , \end{aligned}$$

and

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

respectively.

3 Stochastic representation result

The next result provides a stochastic representation for the derivative ∂Q_k^θ of the transition probability kernel Q_k^θ w.r.t. the parameter θ .

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.

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

Remark 3.1 Under Assumption A, the r -dimensional vector $c_\theta(x)$ such that $\partial b_\theta(x) = \sigma(x) c_\theta(x)$ holds is unique, and

$$\begin{aligned} c_\theta(x) &= [\sigma^*(x) \sigma(x)]^{-1} \sigma^*(x) [\partial b_\theta(x)] \\ &= \sigma^*(x) a^\ominus(x) [\partial b_\theta(x)] , \end{aligned}$$

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

$$|c_\theta(x)|^2 = [\partial b_\theta(x)]^* a^\ominus(x) [\partial b_\theta(x)] .$$

Theorem 3.2 *Under Assumptions A and B, the following representation holds*

$$[\partial Q_k^\theta] \phi(x) = \mathbb{E}^\theta [\phi(X_{t_k}) S_k^\theta | X_{t_{k-1}} = x] , \quad (6)$$

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

$$\begin{aligned} S_k^\theta &= \int_{t_k}^{t_{k-1}} [c_\theta(X_t)]^* dW_t \\ &= \int_{t_k}^{t_{k-1}} [\partial b_\theta(X_t)]^* a^\ominus(X_t) \sigma(X_t) dW_t . \end{aligned}$$

A similar representation result has been obtained in Campillo and LeGland [3, Section 3.1], under the stronger assumption that the diffusion matrix $a = \sigma \sigma^*$ is invertible. The proof of Theorem 3.2 can be found in Cérou, Le Gland and Newton [5] : it uses the reference probability approach, and a result stated in Liptser and Shirayev [11, Section 7.6.4], and follows the same lines as the proof of Proposition 3.1 in Fournié et al. [8], where the stronger assumption on the invertibility of the diffusion matrix $a = \sigma \sigma^*$ is made. As a consequence of Theorem 3.2, the following result is obtained.

Proposition 3.3 *For any $k \geq 0$, the signed measures $w_k^\theta|_{k-1}$ and w_k^θ are absolutely continuous w.r.t. the probability distributions $\mu_k^\theta|_{k-1}$ and μ_k^θ respectively.*

Combining the representations (3) and (6) 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

$$\begin{aligned} & \langle [Q_k^\theta]^* w, \phi \rangle + \langle [\partial Q_k^\theta]^* \mu, \phi \rangle \\ &= \langle w, Q_k^\theta \phi \rangle + \langle \mu, [\partial Q_k^\theta] \phi \rangle \\ &= \int_{\mathbb{R}^m} \mathbb{E}^\theta [\phi(X_{t_k}) | X_{t_{k-1}} = x] \frac{dw}{d\mu}(x) \mu(dx) \quad (7) \\ &+ \int_{\mathbb{R}^m} \mathbb{E}^\theta [\phi(X_{t_k}) S_k^\theta | X_{t_{k-1}} = x] \mu(dx) . \end{aligned}$$

4 Particle filter implementation of the log-likelihood function

The log-likelihood function for the estimation of the parameter θ on the basis of observations (z_0, \dots, z_n) is given by

$$\begin{aligned} \ell(\theta) &= \sum_{k=0}^n \log \int_{\mathbb{R}^m} g_k(x, z_k) \mathbb{P}^\theta[X_{t_k} \in dx \mid \mathcal{Z}_{k-1}] \\ &= \sum_{k=0}^n \log \int_{\mathbb{R}^m} \Psi_k(x) \mu_{k|k-1}^\theta(dx) \\ &= \sum_{k=0}^n \log \langle \mu_{k|k-1}^\theta, \Psi_k \rangle. \end{aligned}$$

A question that naturally arises is whether the optimal filtering equations have any practical use here : 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 $\{\xi_k^i|_{k-1}, i \in I\}$ of i.i.d. random variables, called a *particle system*, with common probability distribution $[Q_k^\theta]^* \bar{\mu}_{k-1}^\theta$, where $\bar{\mu}_{k-1}^\theta$ is an approximation of μ_{k-1}^θ , and to use the corresponding empirical probability distribution

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

as an approximation of $\mu_{k|k-1}^\theta = [Q_k^\theta]^* \mu_{k-1}^\theta$. 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 [9], 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 [6]. Theoretical and practical aspects can be found in the volume edited by Doucet, de Freitas and Gordon [7].

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

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

$$\omega_k^i = \frac{\Psi_k(\xi_k^i|_{k-1})}{\sum_{j \in I} \Psi_k(\xi_k^j|_{k-1})}.$$

Then set

$$\bar{\mu}_k^\theta = \Psi_k \cdot \bar{\mu}_{k|k-1}^\theta = \sum_{i \in I} \omega_k^i \delta_{\xi_k^i|_{k-1}}.$$

(ii) Resampling : independently for all $i \in I$, generate a random variable ξ_k^i with discrete probability distribution $\bar{\mu}_k^\theta$. 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 $\{\omega_k^i, i \in I\}$, and to set

$$\xi_k^i = \xi_{k|k-1}^{\tau(i)},$$

for all $i \in I$.

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

$$dX_t^i = b_\theta(X_t^i) dt + \sigma(X_t^i) dW_t^i,$$

with $X_{t_k}^i = \xi_k^i$. Then set

$$\bar{\mu}_{k+1|k}^\theta = \frac{1}{|I|} \sum_{i \in I} \delta_{\xi_{k+1}^i|_k}.$$

As a result, the particle filter approximation of the log-likelihood function is given by

$$\begin{aligned} \ell(\theta) &\approx \sum_{k=0}^n \log \langle \bar{\mu}_{k|k-1}^\theta, \Psi_k \rangle \\ &= \sum_{k=0}^n \log \left[\frac{1}{|I|} \sum_{i \in I} \Psi_k(\xi_k^i|_{k-1}) \right]. \end{aligned}$$

5 Particle filter implementation of the score function

By definition, the score function is the derivative of the log-likelihood function w.r.t. the parameter θ , i.e.

$$\partial \ell(\theta) = \sum_{k=0}^n \frac{\langle w_{k|k-1}^\theta, \Psi_k \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k \rangle}.$$

Because the signed measures $w_{k|k-1}^\theta$ and w_k^θ are absolutely continuous w.r.t. the probability distributions $\mu_{k|k-1}^\theta$ and μ_k^θ respectively, see Proposition 3.3, it is natural to approximate $w_{k|k-1}^\theta$ and w_k^θ by weighted empirical probability distributions associated with the same

particle system $\{\xi_{k|k-1}^i, i \in I\}$ already used for the particle approximations of $\mu_{k|k-1}^\theta$ and μ_k^θ . To be more specific, since

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

then it is natural to set

$$\bar{w}_{k|k-1}^\theta = \frac{1}{|I|} \sum_{i \in I} \rho_{k|k-1}^i \delta_{\xi_{k|k-1}^i},$$

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

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

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

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

$$\omega_k^i = \frac{\Psi_k(\xi_{k|k-1}^i)}{\sum_{j \in I} \Psi_k(\xi_{k|k-1}^j)}.$$

Then set

$$\bar{\mu}_k^\theta = \Psi_k \cdot \bar{\mu}_{k|k-1}^\theta = \sum_{i \in I} \omega_k^i \delta_{\xi_{k|k-1}^i},$$

and

$$\begin{aligned} \bar{w}_k^\theta &= F_k(\bar{\mu}_{k|k-1}^\theta) \bar{w}_{k|k-1}^\theta \\ &= \sum_{i \in I} \left[\rho_{k|k-1}^i - \sum_{j \in I} \rho_{k|k-1}^j \omega_k^j \right] \omega_k^i \delta_{\xi_{k|k-1}^i}. \end{aligned}$$

(ii) Resampling : independently for all $i \in I$, generate a random variable ξ_k^i with discrete probability distribution $\bar{\mu}_k^\theta$. 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 $\{\omega_k^i, i \in I\}$, and to set

$$\xi_k^i = \xi_{k|k-1}^{\tau(i)},$$

$$\rho_k^i = \rho_{k|k-1}^{\tau(i)} - \frac{1}{|I|} \sum_{j \in I} \rho_{k|k-1}^{\tau(j)},$$

for all $i \in I$.

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

$$dX_t^i = b_\theta(X_t^i) dt + \sigma(X_t^i) dW_t^i,$$

$$dS_t^i = [\partial b_\theta(X_t^i)]^* a^\ominus(X_t^i) \sigma(X_t^i) dW_t^i,$$

with $X_{t_k}^i = \xi_k^i$ and $S_{t_k}^i = 0$. Then set

$$\bar{\mu}_{k+1|k}^\theta = \frac{1}{|I|} \sum_{i \in I} \delta_{\xi_{k+1|k}^i},$$

$$\bar{w}_{k+1|k}^\theta = \frac{1}{|I|} \sum_{i \in I} \rho_{k+1|k}^i \delta_{\xi_{k+1|k}^i},$$

where

$$\rho_{k+1|k}^i = \rho_k^i + \Xi_{k+1|k}^i,$$

for all $i \in I$.

Remark 5.1 If the approximations $\bar{\mu}_k^\theta$ and \bar{w}_k^θ are used in (7), then

$$\begin{aligned} &\langle [Q_{k+1}^\theta]^* \bar{w}_k^\theta, \phi \rangle + \langle [\partial Q_{k+1}^\theta]^* \bar{\mu}_k^\theta, \phi \rangle \\ &= \sum_{i \in I} \mathbb{E}^\theta [\phi(X_{t_{k+1}}) | X_{t_k} = \xi_{k|k-1}^i] \\ &\quad \left[\rho_{k|k-1}^i - \sum_{j \in I} \rho_{k|k-1}^j \omega_k^j \right] \omega_k^i \\ &\quad + \sum_{i \in I} \mathbb{E}^\theta [\phi(X_{t_{k+1}}) S_k^\theta | X_{t_k} = \xi_{k|k-1}^i] \omega_k^i, \end{aligned}$$

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

Remark 5.2 In practice, it may happen that

$$\begin{aligned} \langle \bar{w}_{k|k-1}^\theta, 1 \rangle &= \frac{1}{|I|} \sum_{i \in I} \rho_{k|k-1}^i + \frac{1}{|I|} \sum_{i \in I} \Xi_{k|k-1}^i \\ &= \frac{1}{|I|} \sum_{i \in I} \Xi_{k|k-1}^i \neq 0. \end{aligned}$$

However, thanks to Lemma 2.2, it is harmless to use the following modified definition

$$\rho_{k|k-1}^i = \rho_{k-1}^i + \left[\Xi_{k|k-1}^i - \frac{1}{|I|} \sum_{j \in I} \Xi_{k|k-1}^j \right],$$

for all $i \in I$.

As a result, the particle filter approximation of the score function is given by

$$\begin{aligned} \partial \ell(\theta) &\approx \sum_{k=0}^n \frac{\langle \bar{w}_k^\theta |_{k-1}, \Psi_k \rangle}{\langle \bar{\mu}_k^\theta |_{k-1}, \Psi_k \rangle} \\ &= \sum_{k=0}^n \frac{\sum_{i \in I} \rho_{k|k-1}^i \Psi_k(\xi_{k|k-1}^i)}{\sum_{i \in I} \Psi_k(\xi_{k|k-1}^i)} \\ &= \sum_{k=0}^n \left[\sum_{i \in I} \rho_{k|k-1}^i \omega_k^i \right]. \end{aligned}$$

6 Conclusion

An efficient algorithm has been proposed to compute the log-likelihood function and the score function in the parametric model defined by (1) and (2), under the assumption that the drift coefficient depends on the parameter only in directions that are affected by nondegenerate noise. The algorithm for the score function is especially attractive in the sense that only one m -dimensional particle system is propagated, and one scalar set of weights is propagated for each of the p components of the score function. In particular, the algorithm avoids the use of the gradient system associated with equation (1), where an additional m -dimensional SDE, coupled with equation (1), would be considered for each of the p components of the parameter.

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

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] M. Basseville. On-board component fault detection and isolation using the statistical local approach. *Automatica*, 34(11):1391–1416, Nov. 1998.
- [2] A. Benveniste, M. Basseville, and G. V. Moustakides. The asymptotic local approach to change detection and model validation. *IEEE Transactions on Automatic Control*, AC-32(7):583–592, July 1987.
- [3] 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.
- [4] F. Campillo and F. Le Gland. Threshold selection in the GLR test for change detection in partially observed diffusion processes. In *Proceedings of the 2nd European Control Conference, Groningen 1993*, pages 134–138, July 1993.
- [5] F. Cérou, F. Le Gland, and N. Newton. Stochastic particle methods for linear tangent filtering equations. In *Optimal Control and PDE's — Innovations and Applications : A Volume in Honor of A. Bensoussan*. IOS Press, Amsterdam, 2000.
- [6] 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.
- [7] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer–Verlag, New York, 2001.
- [8] 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.
- [9] 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.
- [10] M. R. James and F. Le Gland. Consistent parameter estimation for partially observed diffusions with small noise. *Applied Mathematics & Optimization*, 32(1):47–72, July/August 1995.
- [11] R. S. Liptser and A. N. Shirayev. *Statistics of Random Processes I. General Theory*, volume 5 of *Applications of Mathematics*. Springer–Verlag, New York, 1977.
- [12] Q. Zhang, M. Basseville, and A. Benveniste. Early warning of slight changes in systems. *Automatica*, 30(1 (Special issue on Statistical Methods in Signal Processing and Control)):95–113, Jan. 1994.