A PARTICLE IMPLEMENTATION OF THE RECURSIVE MLE FOR PARTIALLY OBSERVED DIFFUSIONS

Arnaud Guyader * **François Le Gland** ** **Nadia Oudjane** ***

* *Université de Haute Bretagne, 35043 RENNES Cédex, France*
** *IRISA / INRIA, 35042 RENNES Cédex, France*
*** *EDF, Division R&D, 92141 CLAMART Cédex, France*

Abstract In this paper, the problem of identifying a hidden Markov model (HMM) with general state space, e.g. a partially observed diffusion process, is considered. A particle implementation of the recursive maximum likelihood estimator for a parameter in the transition kernel of the Markov chain is presented. The key assumption is that the derivative of the transition kernel w.r.t. the parameter has a probabilistic interpretation, suitable for Monte Carlo simulation. Examples are given to show that this assumption is satisfied in quite general situations. As a result, the linear tangent filter, i.e. the derivative of the filter w.r.t. the parameter, is absolutely continuous w.r.t. the filter and the idea is to jointly approximate the (prediction) filter and its derivative with the empirical probability distribution and with a weighted empirical distribution associated with the same and unique particle system. Application to the identification of a stochastic volatility model is presented.

Keywords: hidden Markov model, stochastic volatility model, nonlinear filter, linear tangent filter, particle filter, recursive MLE.

1. HIDDEN MARKOV MODEL

The state sequence $\{X_k \, , \, k \geq 0\}$ is a Markov chain taking values in the space E, with transition kernel $Q(x, dx')$, i.e.

$$\mathbb{P}[X_{k+1} \in dx' \mid X_k = x] = Q(x, dx') \; .$$

The kernel $Q(x, dx')$ could depend on a parameter, that should be either estimated, or monitored (i.e. changes w.r.t. a nominal value should be detected), however the dependence w.r.t. the parameter is not written explicitly, so as to avoid intricated notations. The following assumption is made

It is *easy* to *simulate* a r.v. X with probability distribution $Q(x, dx')$, even though the analytical expression of the kernel $Q(x, dx')$ is not known, or is so complicated that it is *pratically impossible* to *compute* such integrals as

$$Q \, \phi(x) = \int_E Q(x, dx') \, \phi(x') \; ,$$

or

$$Q \, \mu(dx') = \int_E \mu(dx) \, Q(x, dx') \; .$$

This is the case for instance if the Markov chain $\{X_k \, , \, k \geq 0\}$ is obtained by sampling a diffusion process $\{X'_t \, , \, t \geq 0\}$ at discrete time instants $\{t_k \, , \, k \geq 0\}$, i.e. if $X_k = X'_{t_k}$, with

$$dX'_t = b(X'_t) \, dt + \sigma(X'_t) \, dW_t \; , \qquad (1)$$

where $\{W_t \, , \, t \geq 0\}$ is a Brownian motion. In this case, to *simulate* a r.v. with probability distribution $Q(x, dx')$ simply reduces to simulate (with an appropriate numerical discretisation scheme) the solution at time t_{k+1} of the stochastic differential equation (1) starting from the initial condition $X'_{t_k} = x$ at time t_k.

[1] This work was partially supported by the CNRS, under the *MathSTIC* project *Chaînes de Markov Cachées et Filtrage Particulaire*, and under the *AS–STIC* project *Méthodes Particulaires* (AS 67).

The state sequence $\{X_k\,,\,k\geq 0\}$ is not observed, but instead an observation sequence $\{Y_k\,,\,k\geq 0\}$ is available, which has the following property : given the hidden states $\{X_k\,,\,k\geq 0\}$, the observations $\{Y_k\,,\,k\geq 0\}$ are mutually independent, and the conditional probability distribution of Y_k depends only on the hidden state X_k at the same time instant, and by definition

$$\mathbb{P}[Y_k\in dy\mid X_k=x]=g(x,y)\,\lambda(dy)\;,$$

and

$$\Psi_k(x)=g(x,Y_k)\;.$$

Notice that when x varies, all the conditional probability distributions $\mathbb{P}[Y_k\in dy\mid X_k=x]$ are assumed absolutely continuous w.r.t. a nonnegative measure $\lambda(dy)$ which does not depend on x (with densities $g(x,y)$ which do depend on x). This *memoryless channel* assumption is satisfied for instance in the case where the hidden state is observed in an additive white noise sequence, not necessarily Gaussian, i.e. in the case where the observation Y_k is related to the hidden state X_k by the relation

$$Y_k=h(X_k)+V_k\;,$$

where $\{V_k\,,\,k\geq 0\}$ is a white noise sequence (i.e. a sequence of mutually independent r.v.'s) with probability distribution $q(v)\,dv$, independent of $\{X_k\,,\,k\geq 0\}$. In this case

$$\mathbb{P}[Y_k\in dy\mid X_k=x]=q(y-h(x))\,dy\;,$$

and

$$\Psi_k(x)=q(Y_k-h(x))\;.$$

The memoryless channel assumption is also satisfied in the case where the covariance of the observation noise depends on the hiden state, i.e. in the case where

$$Y_k=r(X_k)\,V_k\;,$$

where $\{V_k\,,\,k\geq 0\}$ is a white noise sequence with probability distribution $q(v)\,dv$, independent of $\{X_k\,,\,k\geq 0\}$. In this case, provided the matrix $r(x)$ is invertible for any $x\in\mathbb{R}^m$, it holds

$$\mathbb{P}[Y_k\in dy\mid X_k=x]=\frac{q([r(x)]^{-1}\,y)}{\det r(x)}\,dy\;,$$

and

$$\Psi_k(x)=\frac{q([r(x)]^{-1}\,Y_k)}{\det r(x)}\;.$$

2. PARTICLE APPROXIMATION OF THE FILTER

Given observations, the objective is to estimate the hidden states, and to this effect the probability distributions

$$\mu_k(dx)=\mathbb{P}[X_k\in dx\mid Y_0,\cdots,Y_k]\;,$$

and

$$\mu_{k|k-1}(dx)=\mathbb{P}[X_k\in dx\mid Y_0,\cdots,Y_{k-1}]\;,$$

are introduced. The evolution of the sequence $\{\mu_k\,,\,k\geq 0\}$ taking values in the space of probability distributions on E, is very easily described by the following steps

$$\mu_{k-1}\xrightarrow{\text{prediction}}\mu_{k|k-1}=Q\,\mu_{k-1}$$

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

where

$$\mu_{k|k-1}(dx')=Q\,\mu_{k-1}(dx')=\int_E\mu_{k-1}(dx)\,Q(x,dx')\;,$$

can happen to be *difficult* (if not just impossible) to *compute*, and where $\cdot$ denotes the projective product, i.e.

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

In view of the key assumption that it is on the other hand *easy* to *simulate* r.v.'s with probability distribution $Q(x,dx')$, the idea is to approximate the predictor $\mu_{k|k-1}$ with the empirical probability distribution associated with an N–sample, i.e.

$$\mu_{k|k-1}\approx\mu^N_{k|k-1}=\frac{1}{N}\sum_{i=1}^N\delta_{\xi^i_{k|k-1}}\;.$$

This approximation is completely characterized by the set $\{\xi^i_{k|k-1}\,,\,i=1,\cdots,N\}$ of particles, and the algorithm is completely described by the mechanism which builds $\{\xi^i_{k+1|k}\,,\,i=1,\cdots,N\}$ from $\{\xi^i_{k|k-1}\,,\,i=1,\cdots,N\}$. This mechanism is as follows :

(i) the correction step is applied *exactly* to $\mu^N_{k|k-1}$, which results in

$$\mu^N_k=\Psi_k\cdot\mu^N_{k|k-1}=\sum_{i=1}^N\frac{\Psi_k(\xi^i_{k|k-1})\,\delta_{\xi^i_{k|k-1}}}{\displaystyle\sum_{j=1}^N\Psi_k(\xi^j_{k|k-1})}=\sum_{i=1}^N\omega^i_k\,\delta_{\xi^i_{k|k-1}}\;,$$

i.e. particles $\{\xi^i_{k|k-1}\,,\,i=1,\cdots,N\}$ are now weighted, with weights $\{\omega^i_k\,,\,i=1,\cdots,N\}$ which are more heavy for those particles which are more *consistent* with the current observation Y_k,

(ii) instead of trying to *compute* $Q\,\mu^N_k$, the following particle approximation

$$\mu^N_{k+1|k}=S^N(Q\,\mu^N_k)=\frac{1}{N}\sum_{i=1}^N\delta_{\xi^i_{k+1|k}}\;,$$

is used, where the r.v.'s $\{\xi^i_{k+1|k}\,,\,i=1,\cdots,N\}$ form an N–sample with probability distribution

$Q\,\mu_k^N$, which can be easily achieved in the following manner : independently for any $i = 1, \cdots, N$

$$\xi_k^i \sim \mu_k^N(dx) \; ,$$

which easy, since the probability distribution μ_k^N is discrete, and

$$\xi_{k+1|k}^i \sim Q(\xi_k^i, dx') \; ,$$

which is easy, by assumption.

3. LINEAR TANGENT KERNEL / EXTENDED KERNEL, ETC.

If the transition kernel $Q(x, dx')$ depends on a parameter, then the filter μ_k depends also on the parameter, and one would like to compute the linear tangent filter w_k, i.e. the derivative of the filter μ_k w.r.t. the parameter. To this end, one needs first to study the linear tangent kernel $\Gamma(x, dx')$, i.e. the derivative of the transition kernel $Q(x, dx')$ w.r.t. the parameter, and the following assumption is made

Assumprion AC : The following probabilistic representation holds for the linear tangent kernel $\Gamma(x, dx')$

$$\Gamma\,\phi(x) = \int_E \Gamma(x, dx')\,\phi(x')$$

$$= \mathbb{E}[\phi(X_{k+1})\,\Xi_{k+1} \mid X_k = x] \; ,$$

where $\{\,(X_k, \Xi_k)\,,\; k \geq 0\}$ is a Markov chain taking values in the product space $E \times F$, such that

$$\mathbb{P}[X_{k+1} \in dx', \Xi_{k+1} \in ds' \mid X_k = x, \Xi_k = s]$$

$$= \mathbb{P}[X_{k+1} \in dx', \Xi_{k+1} \in ds' \mid X_k = x]$$

$$= K(x, dx', ds') \; .$$

The following assumption, which extends the similar assumption introduced in Section 1, is made

It is *easy* to *simulate* a r.v. (X, Ξ) with probability distribution $K(x, dx', ds')$, even though the analytical expression of the kernel $K(x, dx', ds')$ is not known, or is so complicated that it is *pratically impossible* to *compute* such integrals as

$$\Gamma\,\phi(x) = \int_{E \times F} s'\,\phi(x')\,K(x, dx', ds') \; ,$$

or

$$\Gamma\,\mu(dx') = \int_{E \times F} \mu(dx)\,s'\,K(x, dx', ds') \; .$$

Example 3.1. Let the Markov chain $\{X_k\,,\; k \geq 0\}$ taking values in $E = \mathbb{R}^m$, be defined by

$$X_{k+1} = f(X_k) + g(X_k)\,W_k \; ,$$

where only the function f depends on the parameter, and where $\{W_k\,,\; k \geq 0\}$ is a sequence of independent r.v.'s taking values in $\mathbb{R}^q$ with probability distribution $p(w)\,dw$. In the simple case where $g(x) = I$ for any $x \in \mathbb{R}^m$, the transition kernel $Q(x, dx')$ is given by

$$Q(x, dx') = p(x' - f(x))\,dx' \; ,$$

and one can show directly that

$$\Gamma(x, dx') = -p'(x' - f(x))\;\partial f(x)\,dx'$$

$$= \frac{-p'}{p}(x' - f(x))\;\partial f(x)\;Q(x, dx') \; ,$$

where ∂f denotes the derivative of the function f w.r.t. the parameter. It follows that

$$\Gamma\,\phi(x) = \int_E \Gamma(x, dx')\,\phi(x')$$

$$= \int_E \phi(x')\;\frac{-p'}{p}(x' - f(x))\;\partial f(x)\;Q(x, dx') \; ,$$

i.e. Assumption AC is satisfied, with

$$\Xi_{k+1} = \frac{-p'}{p}(W_k)\;\partial f(X_k)$$

This result generalizes to the case where for any $x \in \mathbb{R}^m$ the matrix $g(x)$ has full rank, and the vector $\partial f(x)$ belongs to the range of $g(x)$.

Notice that in the above example, the r.v. Ξ_{k+1} depends only on (X_k, W_k), in which case it does not seem necessary to simulate Ξ_{k+1} in addition to W_k. Taking into account that the matrix $g(X_k)$ has full rank, it is even possible to express W_k in terms of (X_k, X_{k+1}), and finally the r.v. $\Xi_{k+1} = I(X_k, X_{k+1})$ depends only on (X_k, X_{k+1}). This apparently very particular situation is actually very general, as the following result shows.

Lemma 3.2. Under Assumption AC

$$\Gamma(x, dx') = I(x, x')\;Q(x, dx') \; ,$$

with

$$I(x, x') = \mathbb{E}[\Xi_{k+1} \mid X_k = x, X_{k+1} = x'] \; ,$$

for any $x, x' \in E$.

However, and as the following two examples show, there exist situations where (i) the existence of the function I does not imply that an easy–to–compute explicit expression exists, whereas in opposition (ii) the joint simulation of (X_{k+1}, Ξ_{k+1}) is easy.

Example 3.3. Let the Markov chain $\{X_k\,,\; k \geq 0\}$ taking values in $E = \mathbb{R}^m$, be defined by

$$X_{k+1} = f(X_k) + c\,g(X_k)\,W_k \; ,$$

where $\{W_k\,,\; k \geq 0\}$ is a sequence of independent r.v.'s taking values in $\mathbb{R}^q$ with probability distribution $p(w)\,dw$. The transition kernel $Q(x, dx')$ satisfies

$$Q\,\phi(x) = \int_{\mathbb{R}^q} \phi(f(x) + c\,g(x)\,w)\; p(w)\,dw$$
$$= \int_{\mathbb{R}^q} \phi(f(x) + g(x)\,u)\; p(\frac{u}{c})\,\frac{du}{c^q}\;,$$

for any test function ϕ defined on E. Differentiating w.r.t. the parameter c yields

$$\Gamma\,\phi(x) = \int_{\mathbb{R}^q} \phi(f(x) + g(x)\,u)$$
$$[\,-p'(\frac{u}{c})\,\frac{u}{c^2}\,\frac{du}{c^q} - p(\frac{u}{c})\,\frac{q\,du}{c^{q+1}}\,]$$
$$= \int_{\mathbb{R}^q} \phi(f(x) + c\,g(x)\,w)$$
$$\frac{1}{c}\,[\,\frac{-p'}{p}(w)\,w - q\,]\;p(w)\,dw\;,$$

i.e. Assumption AC is satisfied, with

$$\Xi_{k+1} = \frac{1}{c}\,[\,\frac{-p'}{p}(W_k)\,W_k - q\,]\;,$$

and

$$\Xi_{k+1} = \frac{1}{c}\,(\,|W_k|^2 - q\,)\;,$$

in the special case where W_k is a zero mean Gaussian r.v. with identity covariance matrix. This result holds without any assumption on the matrix $g(X_k)$ which in any case does not appear in the expression of Ξ_{k+1}, and unless the matrix $g(X_k)$ has full rank, it is not possible in general to express W_k in terms of (X_k, X_{k+1}).

Example 3.4. Let the Markov chain $\{X_k\,,\,k \geq 0\}$ be defined by sampling at discrete time instants $\{t_k\,,\,k \geq 0\}$ a diffusion process $\{X'_t\,,\,t \geq 0\}$, i.e. $X_k = X'_{t_k}$, with

$$dX'_t = b(X'_t)\,dt + \sigma(X'_t)\,dW_t\;,$$

where only the drift function b depends on the parameter, and where $\{W_t\,,\,t \geq 0\}$ is a Brownian motion. If for any $x \in \mathbb{R}^m$, the matrix $\sigma(x)$ has full rank, and the vector $\partial b(x)$ belongs to the range of $\sigma(x)$, then Assumption AC is satisfied, with

$$\Xi_{k+1} = \int_{t_k}^{t_{k+1}} [\,\partial b(X'_t)\,]^*$$
$$\sigma(X'_t)\;[\sigma^*(X'_t)\,\sigma(X'_t)]^{-1}\;dW_t\;,$$

where ∂b denotes the derivative of the drift function b w.r.t. the parameter, see (Cérou *et al.*, 2001) or (Fournié *et al.*, 1999) for the simpler case where for any $x \in \mathbb{R}^m$ the matrix $\sigma(x)$ is invertible. It is easy (with an appropriate numerical discretization scheme) to jointly simulate (X_{k+1}, Ξ_{k+1}), but in opposition there does not exist in general a simple analytical expression for

$$I(x,x') = \mathbb{E}[\Xi_{k+1} \mid X_k = x, X_{k+1} = x']$$
$$= \mathbb{E}[\int_{t_k}^{t_{k+1}} [\,\partial b(X'_t)\,]^*$$
$$\sigma(X'_t)\;[\sigma^*(X'_t)\,\sigma(X'_t)]^{-1}\;dW_t \mid$$
$$X'_{t_k} = x, X'_{t_{k+1}} = x']\;.$$

By definition

$$\Gamma\,\phi(x) = \int_{E\times F} \phi(x')\;s'\,K(x, dx', ds')\;,$$

and

$$Q\,\phi(x) = \int_{E\times F} \phi(x')\;K(x, dx', ds')\;,$$

hence

$$\Gamma(x, dx') = \int_F s'\,K(x, dx', ds')\;,$$

and

$$Q(x, dx') = \int_F K(x, dx', ds')\;.$$

On the product space $E \times E \times F$, define the projection $\pi_0\;:\;(x, x', s') \longmapsto x$ on the (first) space E, the projection $\pi\;:\;(x, x', s') \longmapsto x'$ on the (second) space E and the projection $\pi_F\;:\;(x, x', s') \longmapsto s'$ on the auxiliary space F. For any probability distribution μ on the space E, the probability distribution $\mu \otimes K$ is defined on the product space $E \times E \times F$ by

$$(\mu \otimes K)(dx, dx', ds') = \mu(dx)\;K(x, dx', ds')\;.$$

It follows that

$$Q\,\mu(dx') = \int_{E\times F} \mu(dx)\,K(x, dx', ds')$$
$$= (\mu \otimes K) \circ \pi^{-1}(dx')\;,$$

and

$$\Gamma\,\mu(dx') = \int_{E\times F} \mu(dx)\,s'\,K(x, dx', ds')$$
$$= \int_{E\times F} \pi_F(x, x', s')\,(\mu \otimes K)(dx, dx', ds')$$
$$= (\pi_F\,(\mu \otimes K)) \circ \pi^{-1}(dx')\;,$$

and if the finite signed measure w is absolutely continuous w.r.t. μ, then

$$Q\,w(dx') = \int_{E\times F} w(dx)\, K(x,dx',ds')$$

$$= \int_{E\times F} \frac{dw}{d\mu}(x)\, \mu(dx)\, K(x,dx',ds')$$

$$= \int_{E\times F} (\frac{dw}{d\mu} \circ \pi_0)(x,x',s')$$

$$(\mu \otimes K)(dx,dx',ds')$$

$$= ((\frac{dw}{d\mu} \circ \pi_0)\ (\mu \otimes K)) \circ \pi^{-1}(dx')\ ,$$

i.e.

$$Q\,\mu = (\mu \otimes K) \circ \pi^{-1}\ ,$$

and

$$\Gamma\,\mu = (\pi_F\ (\mu \otimes K)) \circ \pi^{-1}\ ,$$

and

$$(\, w \ll \mu \quad \Longrightarrow \quad Q\,w = ((\frac{dw}{d\mu}\circ\pi_0)\ (\mu\otimes K))\circ\pi^{-1}\,)\ .$$

Lemma 3.5. Under Assumption AC, $\Gamma\,\mu \ll Q\,\mu$ for any probability distribution μ on E, with Radon–Nikodym derivative (which depends on μ)

$$\frac{d(\Gamma\,\mu)}{d(Q\,\mu)}(x') = \mathbb{E}_\mu[\Xi_{n+1} \mid X_{n+1} = x']\ .$$

For completeness, the following elementary property is recalled

Lemma 3.6. If the finite signed measure w is absolutely continuous w.r.t. the probability distribution μ, then $Q\,w \ll Q\,\mu$, with Radon–Nikodym derivative

$$\frac{d(Q\,w)}{d(Q\,\mu)}(x') = \mathbb{E}_\mu[\frac{dw}{d\mu}(X_k) \mid X_{k+1} = x']\ .$$

The explicit expression of the Radon–Nikodym derivatives will not be used in the sequel : only the qualitative properties

$$\Gamma\,\mu \ll Q\,\mu\ ,$$

and

$$(\, w \ll \mu \quad \Longrightarrow \quad Q\,w \ll Q\,\mu\,)\ ,$$

will be used.

By definition

$$F_k(\mu)\,w = \frac{\Psi_k\,w}{\langle \mu, \Psi_k\rangle} - \frac{\langle w, \Psi_k\rangle}{\langle \mu, \Psi_k\rangle}\ \frac{\Psi_k\,\mu}{\langle \mu, \Psi_k\rangle}\ ,$$

is the derivative at point μ and in the direction w, of the mapping $\mu \longmapsto \Psi_k \cdot \mu$. The following elementary property holds

Lemma 3.7. If the finite signed measure w is absolutely continuous w.r.t. the probability distribution μ, then $F_k(\mu)\,w \ll \Psi_k \cdot \mu$, with Radon–Nikodym derivative

$$\frac{d(F_k(\mu)\,w)}{d(\Psi_k \cdot \mu)}(x) = \frac{dw}{d\mu}(x) - \langle \Psi_k \cdot \mu, \frac{dw}{d\mu}\rangle\ .$$

4. PARTICLE APPROXIMATION OF SOME FINITE SIGNED MEASURES

With the notations of the previous section, it easily seen that the probability distribution $Q\,\mu$ and the finite signed measures $\Gamma\,\mu$ and $Q\,w$ can be put in the general form $(r\,(\mu \otimes K)) \circ \pi^{-1}$ for some appropriate choice of the weight function r, namely $r \equiv 1$, $r = \pi_F$ and $r = \dfrac{dw}{d\mu} \circ \pi_0$ respectively. The weighted particle approximation of a finite signed measure of the general form $r\,(\mu \otimes K)$ is defined by

$$r\,(\mu \otimes K) \approx r\,S^N(\mu \otimes K)$$

$$= \frac{1}{N} \sum_{i=1}^N r(\xi_0^i, \xi^i, \Xi^i)\ \delta_{(\xi_0^i, \xi^i, \Xi^i)}\ ,$$

where the r.v.'s $\{\, \xi_0^i, \xi^i, \Xi^i\ ,\ i = 1, \cdots, N\}$ form an N–sample with probability distribution $\mu \otimes K$, which can be easily achieved in the following manner : independently for any $i = 1, \cdots, N$

$$\xi_0^i \sim \mu(dx)\ ,$$

and

$$(\xi^i, \Xi^i) \sim K(\xi_0^i, dx', ds')\ ,$$

and the corresponding particle approximation for the marginal measure $(r\,(K \otimes \mu)) \circ \pi^{-1}$ is defined by

$$(r\,(\mu \otimes K)) \circ \pi^{-1} \approx (r\,S^N(\mu \otimes K)) \circ \pi^{-1}$$

$$= \frac{1}{N} \sum_{i=1}^N r(\xi_0^i, \xi^i, \Xi^i)\ \delta_{\xi^i}\ .$$

In particular for the weight functions $r \equiv 1$, $r = \pi_F$ and $r = \dfrac{dw}{d\mu} \circ \pi_0$, it holds

$$Q\,\mu = (\mu \otimes K) \circ \pi^{-1} \approx \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i}\ ,$$

$$\Gamma\,\mu = (\pi_F\ (\mu \otimes K)) \circ \pi^{-1} \approx \frac{1}{N} \sum_{i=1}^N \Xi^i\ \delta_{\xi^i}\ ,$$

and

$$Q\,w = ((\frac{dw}{d\mu}\circ\pi_0)\ (\mu\otimes K))\circ\pi^{-1} \approx \frac{1}{N} \sum_{i=1}^N \frac{dw}{d\mu}(\xi_0^i)\ \delta_{\xi^i}\ ,$$

respectively. For any test function ϕ defined on E, it holds

$$\sup_{\|\phi\|=1} \mathbb{E}|\ \frac{1}{N} \sum_{i=1}^N \phi(\xi^i) - \langle Q\,\mu, \phi\rangle\ | \le \frac{1}{\sqrt{N}}\ ,$$

$$\sup_{\|\phi\|=1} \mathbb{E}|\ \frac{1}{N} \sum_{i=1}^N \Xi^i\ \phi(\xi^i) - \langle \Gamma\,\mu, \phi\rangle\ |$$

$$\le \frac{1}{\sqrt{N}}\ \{\sup_{x \in E} \int_{E\times F} |s'|^2\ K(x,dx',ds')\,\}^{1/2}\ ,$$

and

$$\sup_{\|\phi\|=1} \mathbb{E}| \frac{1}{N} \sum_{i=1}^N \frac{dw}{d\mu}(\xi_0^i)\; \phi(\xi^i) - \langle Q\, w, \phi\rangle \,| \leq \frac{1}{\sqrt{N}}\; \{ \int_E |\frac{dw}{d\mu}(x)|^2\, \mu(dx)\, \}^{1/2} \;,$$

respectively.

5. JOINT PARTICLE APPROXIMATION OF THE FILTER AND THE LINEAR TANGENT FILTER

Recall that the evolution of the sequence $\{\mu_k \,,\, k \geq 0\}$ taking values in the space of probability distributions on E, is described by the following two steps

$$\mu_{k-1} \xrightarrow{\text{prediction}} \mu_{k|k-1} = Q\, \mu_{k-1}$$

$$\xrightarrow{\text{correction}} \mu_k = \Psi_k \cdot \mu_{k|k-1} \;.$$

If w_k denotes at each time instant the linear tangent filter, i.e. the derivative of the filter μ_k w.r.t. the parameter, then the evolution of the sequence $\{w_k \,,\, k \geq 0\}$ taking values in the linear tangent space to the space of probability distributions on E, i.e. taking values in the space of finite signed measures on E with zero total mass, is described by the following two steps, which are linear tangent versions of the prediction step and correction step respectively

$$w_{k-1} \xrightarrow[\text{}]{\text{linear tangent prediction}} w_{k|k-1} = Q\, w_{k-1} + \Gamma\, \mu_{k-1}$$

$$\xrightarrow{\text{linear tangent correction}} w_k = F_k(\mu_{k|k-1})\, w_{k|k-1} \;.$$

Under Assumption AC, it is easily seen by induction, and using Lemmas 3.5, 3.6 and 3.7, that at each time instant $w_{k|k-1} \ll \mu_{k|k-1}$ and $w_k \ll \mu_k$.

In view of this absolute continuity property, and of the key assumption that it is *easy* to *simulate* r.v.'s with probability distribution $K(x, dx', ds')$, the idea is to jointly approximate the predictor $\mu_{k|k-1}$ and its derivative $w_{k|k-1}$ w.r.t. the parameter with the empirical probability distribution and a weighted empirical distribution associated with the same and unique N–sample, i.e.

$$\mu_{k|k-1} \approx \mu^N_{k|k-1} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_{k|k-1}} \;,$$

and

$$w_{k|k-1} \approx w^N_{k|k-1} = \frac{1}{N} \sum_{i=1}^N \rho^i_{k|k-1}\; \delta_{\xi^i_{k|k-1}} \;.$$

With this definition $w^N_{k|k-1} \ll \mu^N_{k|k-1}$, with Radon–Nikodym derivative

$$r^N_{k|k-1}(x) = \frac{dw^N_{k|k-1}}{d\mu^N_{k|k-1}}(x) = \frac{1}{|I^N_{k|k-1}(x)|} \sum_{i \in I^N_{k|k-1}(x)} \rho^i_{k|k-1} \;,$$

where $I^N_{k|k-1}(x) = \{ i = 1, \cdots, N \,:\, \xi^i_{k|k-1} = x\}$, for any x in the support supp $\mu^N_{k|k-1}$ of the discrete probability distribution $\mu^N_{k|k-1}$. Indeed

$$\mu^N_{k|k-1} = \frac{1}{N} \sum_{x \in \text{supp } \mu^N_{k|k-1}} |I^N_{k|k-1}(x)|\; \delta_x \;,$$

and

$$w^N_{k|k-1} = \frac{1}{N} \sum_{x \in \text{supp } \mu^N_{k|k-1}} [\sum_{i \in I^N_{k|k-1}(x)} \rho^i_{k|k-1}\,]\; \delta_x \;.$$

Notice that in most cases, the particle locations $\{\xi^i_{k|k-1} \,,\, i = 1, \cdots, N\}$ happen to be all distinct, and the much simpler relation

$$r^N_{k|k-1}(\xi^i_{k|k-1}) = \rho^i_{k|k-1} \;,$$

holds for any $i = 1, \cdots, N$.

This approximation is completely characterized by the set $\{\xi^i_{k|k-1}, \rho^i_{k|k-1} \,,\, i = 1, \cdots, N\}$ of particles and weights, and the algorithm is completely described by the mechanism which builds $\{\xi^i_{k+1|k}, \rho^i_{k+1|k} \,,\, i = 1, \cdots, N\}$ from $\{\xi^i_{k|k-1}, \rho^i_{k|k-1} \,,\, i = 1, \cdots, N\}$. This mechanism is as follows :

(i) the correction step is applied *exactly* to $\mu^N_{k|k-1}$, which results in

$$\mu^N_k = \Psi_k \cdot \mu^N_{k|k-1} = \sum_{i=1}^N \frac{\Psi_k(\xi^i_{k|k-1})\; \delta_{\xi^i_{k|k-1}}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi^j_{k|k-1})} = \sum_{i=1}^N \omega^i_k\; \delta_{\xi^i_{k|k-1}} \;,$$

as previously, and the linear tangent correction step is applied *exactly* to $w^N_{k|k-1}$, which results in

$$w^N_k = F_k(\mu^N_{k|k-1})\, w^N_{k|k-1} = [\, r^N_{k|k-1} - \langle \mu^N_k, r^N_{k|k-1}\rangle\,]\; \mu^N_k \;,$$

(ii) instead of trying to *compute* $Q\, \mu^N_k$, the following particle approximation

$$\mu^N_{k+1|k} = S^N(Q\, \mu^N_k) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_{k+1|k}} \;,$$

is used as previously, instead of trying to *compute*

$$Q\, w_k^N = Q(r_{k|k-1}^N\, \mu_k^N) - \langle \mu_k^N, r_{k|k-1}^N \rangle\; Q\, \mu_k^N$$

$$= ((r_{k|k-1}^N \circ \pi_0)\; (\mu_k^N \otimes Q)) \circ \pi^{-1}$$

$$- \langle \mu_k^N, r_{k|k-1}^N \rangle\; Q\, \mu_k^N \;,$$

the following weighted particle approximation

$$(\, (r_{k|k-1}^N \circ \pi_0)\; S^N(\mu_k^N \otimes Q)\,) \circ \pi^{-1}$$

$$- \langle S^N(\mu_k^N), r_{k|k-1}^N \rangle\; S^N(Q\, \mu_k^N)$$

$$= \frac{1}{N} \sum_{i=1}^N [\, r_{k|k-1}^N(\xi_k^i)$$

$$- \frac{1}{N} \sum_{j=1}^N r_{k|k-1}^N(\xi_k^j)\,]\; \delta_{\xi_{k+1|k}^i} \;,$$

is used, and instead of trying to *compute*

$$\Gamma\, \mu_k^N = (\pi_F\, (\mu_k^N \otimes K)) \circ \pi^{-1} \;,$$

the following weighted particle approximation

$$(\pi_F\;\, S^N(\mu_k^N \otimes K)) \circ \pi^{-1} = \frac{1}{N} \sum_{i=1}^N \Xi_{k+1}^i\; \delta_{\xi_{k+1|k}^i} \;,$$

is used, hence finally the weighted particle approximation

$$w_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N [\, r_{k|k-1}^N(\xi_k^i) + \Xi_{k+1}^i$$

$$- \frac{1}{N} \sum_{j=1}^N r_{k|k-1}^N(\xi_k^j)\,]\; \delta_{\xi_{k+1|k}^i}$$

$$= \frac{1}{N} \sum_{i=1}^N \rho_{k+1|k}^i\; \delta_{\xi_{k+1|k}^i} \;,$$

where the r.v.'s $\{\, \xi_k^i, \xi_{k+1|k}^i, \Xi_{k+1}^i \;,\; i = 1, \cdots, N\}$ form an N–sample with probability distribution $\mu_k^N \otimes K$, which can be easily achieved in the following manner : independently for any $i = 1, \cdots, N$

$$\xi_k^i \sim \mu_k^N(dx) \;,$$

which is easy, since the probability distribution μ_k^N is discrete, and

$$(\xi_{k+1|k}^i, \Xi_{k+1}^i) \sim K(\xi_k^i, dx', ds') \;,$$

which is easy, by assumption.

6. PARTICLE FILTER IMPLEMENTATION OF THE RECURSIVE MLE

In this section, the parameter is denoted by θ and dependence w.r.t. the parameter appears explicitly in the notation for the transition kernel $Q^\theta(x, dx')$, and for the linear tangent kernel $K^\theta(x, dx', ds')$. It is well–known that in such a parametric model, the log–likelihood function for the estimation of the parameter θ can be written as

$$\ell_n^\theta = \frac{1}{n} \sum_{k=0}^n \log \langle \mu_{k|k-1}^\theta, \Psi_k \rangle \;,$$

and the score fuction, i.e. the derivative of the log–likelihood function w.r.t. the parameter, can be written as

$$\partial\, \ell_n^\theta = \frac{1}{n} \sum_{k=0}^n \frac{\langle w_{k|k-1}^\theta, \Psi_k \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k \rangle} \;,$$

where the filter $\{\mu_k^\theta \;,\; k \geq 0\}$ and the linear tangent filter $\{w_k^\theta \;,\; k \geq 0\}$ satisfy

$$\mu_{k-1}^\theta \xrightarrow{\text{prediction}} \mu_{k|k-1}^\theta = Q^\theta\, \mu_{k-1}^\theta$$

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

and

$$w_{k-1}^\theta \xrightarrow[\text{}]{\substack{\text{linear tangent} \\ \text{prediction}}} w_{k|k-1}^\theta = Q^\theta\, w_{k-1}^\theta + \Gamma^\theta\, \mu_{k-1}^\theta$$

$$\xrightarrow{\substack{\text{linear tangent} \\ \text{correction}}} w_k^\theta = F_k(\mu_{k|k-1}^\theta)\, w_{k|k-1}^\theta \;,$$

respectively.

Monitoring the parametric model, i.e. detecting a small change from a nominal value, corresponding to the normal behaviour of the system, has been addressed in (Cérou and Le Gland, 2000). Another question is to identify the parametric model, and it is natural to consider the recursive MLE, which is defined by the following relation

$$\widehat{\theta}_k = \widehat{\theta}_{k-1} + \gamma_k\; \frac{\langle \widehat{w}_{k|k-1}, \Psi_k \rangle}{\langle \widehat{\mu}_{k|k-1}, \Psi_k \rangle} \;, \qquad (2)$$

where typically $\gamma_k \simeq k^{-2/3}$, and the averaged estimator (which achieves the minimum variance of the estimation error) is obtained by post–processing

$$\overline{\theta}_k = \overline{\theta}_{k-1} + \frac{1}{k}\, (\widehat{\theta}_k - \overline{\theta}_{k-1}) \;.$$

Here, the adaptive filter $\{\widehat{\mu}_k \;,\; k \geq 0\}$ and the adaptive linear tangent filter $\{\widehat{w}_k, \;,\; k \geq 0\}$ satisfy the same equations as the filter and the linear tangent filter respectively, in which the value of the parameter is adapted at each time instant according to equation (2), i.e.

$$\widehat{\mu}_{k-1} \xrightarrow{\substack{\text{adaptive} \\ \text{prediction}}} \widehat{\mu}_{k|k-1} = Q^{\widehat{\theta}_{k-1}}\, \widehat{\mu}_{k-1}$$

$$\xrightarrow{\text{correction}} \widehat{\mu}_k = \Psi_k \cdot \widehat{\mu}_{k|k-1} \;,$$

and

$$\widehat{w}_{k-1} \xrightarrow[\text{prediction}]{\substack{\text{adaptive}\\ \text{linear tangent}}} \widehat{w}_{k|k-1} = Q^{\widehat{\theta}_{k-1}}\, \widehat{w}_{k-1} + \Gamma^{\widehat{\theta}_{k-1}}\, \widehat{\mu}_{k-1}$$

$$\xrightarrow[\text{correction}]{\text{linear tangent}} \widehat{w}_k = F_k(\widehat{\mu}_{k|k-1})\, \widehat{w}_{k|k-1} \;,$$

respectively. The particle implementation of the recursive MLE is

$$\widehat{\theta}_k^N = \widehat{\theta}_{k-1}^N + \gamma_k \, [\sum_{i=1}^N \rho_{k|k-1}^i \, \omega_k^i \,] \;,$$

and

$$\overline{\theta}_k^N = \overline{\theta}_{k-1}^N + \frac{1}{k}\,(\widehat{\theta}_k^N - \overline{\theta}_{k-1}^N) \;,$$

and the corresponding algorithm is described in Table 1.

The mathematical analysis of the asymptotic properties of the estimator $\widehat{\theta}_k^N$ as $k \to \infty$ and $N \to \infty$ is far beyond the scope of this paper, and would rely on joint stability properties of the filter and the linear tangent filter, which is a very difficult question. Even the asymptotic properties of the estimator $\widehat{\theta}_k$ as $k \to \infty$ are difficult to prove, unless some mixing assumption holds for the transition kernels $Q^\theta(x, dx')$ and the linear tangent kernels $\Gamma^\theta(x, dx')$, which practically implies that the state–space E should be compact, see e.g. (Douc and Matias, 2001) where only the non–recursive MLE is studied.

7. APPLICATION TO A STOCHASTIC VOLATILITY MODEL

The following stochastic volatility model, with mean–reverting hidden diffusion

$$dV_t = a\,(b - V_t)\,dt + c\,V_t\,dW'_t \;, \qquad V_0 > 0$$

$$dY'_t = \sqrt{V_t}\,dB'_t \;,$$

where $\{W'_t \,,\, t \geq 0\}$ and $\{B'_t \,,\, t \geq 0\}$ are independent Brownian motions, has been considered by (Genon-Catalot *et al.*, 2000), yes

and by (Sørensen, 2000). An alternate discrete–time observation model is considered in the present paper, in which $X_k = V_{t_k}$ and

$$Y_k = \sqrt{X_k}\, B_k \;,$$

where $\{B_k \,,\, k \geq 0\}$ is a Gaussian white noise sequence independent of $\{W_t \,,\, t \geq 0\}$, hence

$$\Psi_k(x) = \frac{1}{\sqrt{x}}\, \exp\{-\frac{Y_k^2}{2x}\} \;, \qquad x > 0 \;.$$

It follows from Example 3.4 and from (Fournié *et al.*, 1999) that Assumption AC is satisfied, with

$$\Xi_{k+1} = (\Xi_{k+1}^a \,,\, \Xi_{k+1}^b \,,\, \Xi_{k+1}^c)$$

$$= (\int_{t_k}^{t_{k+1}} \frac{b - V_t}{c\,V_t}\, dW'_t \,,\, \int_{t_k}^{t_{k+1}} \frac{a}{c\,V_t}\, dW'_t \,,\, \frac{1}{c}\,(|\frac{\Delta W'_k}{\sqrt{\Delta_k}}|^2 - 1) - \Delta W'_k \,) \;,$$

where $\Delta_k = t_{k+1} - t_k$ and $\Delta W'_k = W'_{t_{k+1}} - W'_{t_k}$.

Instead of the sampled version $X_k = V_{t_k}$ of the continuous–time hidden diffusion, an approximate Markov chain $\{X_k \,,\, k \geq 0\}$ could be used, based on a Euler (or an alternate splitting–up) scheme, i.e.

$$X_k = (1 - a\,\Delta_k)\,X_{k-1} + a\,b\,\Delta_k + c\,X_{k-1}\,W_k \;,$$

where $\Delta_k = t_{k+1} - t_k$ and where $\{W_k \,,\, k \geq 0\}$ is a Gaussian white noise sequence with variance Δ_k. In this case, it follows from Examples 3.1 and 3.3 that Assumption AC is satisfied again, with

$$\Xi_{k+1} = (\Xi_{k+1}^a \,,\, \Xi_{k+1}^b \,,\, \Xi_{k+1}^c)$$

$$= (\frac{b - X_k}{c\,X_k}\, W_k \,,\, \frac{a}{c\,X_k}\, W_k \,,\, \frac{1}{c}\,(|W_k|^2 - 1)\,) \;.$$

Numerical results will be presented in the final version of the paper.

8. REFERENCES

Cérou, Frédéric and François Le Gland (2000). 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. pp. 1200–1205.

Cérou, Frédéric, François Le Gland and Nigel J. Newton (2001). Stochastic particle methods for linear tangent filtering equations. In: *Optimal Control and Partial Differential Equations. In honour of professor Alain Bensoussan's 60th birthday* (José-Luis Menaldi, Edmundo Rofman and Agnès Sulem, Eds.). pp. 231–240. IOS Press. Amsterdam.

Douc, Randal and Catherine Matias (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7**(3), 381–420.

Fournié, Éric, Jean-Michel Lasry, Jérôme Lebuchoux, Pierre-Louis Lions and Nizar Touzi (1999). Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics* **3**(4), 391–412.

Genon-Catalot, Valentine, Thierry Jeantheau and Catherine Laredo (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**(6), 1051–1079.

Sørensen, Helle (2000). Inference for Diffusion Processes and Stochastic Volatility Models. Ph.D. Thesis. Department of Statistics and Operations Research, University of Copenhagen. Copenhagen.

initialization : $k = 0$, choose $\widehat{\theta}^N_{-1}$, and independently for any $i = 1, \cdots, N$, simulate

$$\xi^i_{0|-1} \sim \mu_0(dx) \qquad \text{and set} \qquad \rho^i_{0|-1} = 0 \; ,$$

selection : for any $i = 1, \cdots, N$, compute

$$\omega^i_k \propto \Psi_k(\xi^i_{k|k-1}) \; , \qquad \text{set} \qquad \mu^N_k = \frac{1}{N} \sum_{i=1}^N \omega^i_k \, \delta_{\xi^i_{k|k-1}} \; ,$$

and for any point x in the support of the particle system, compute

$$r^N_{k|k-1}(x) = \frac{1}{|I^N_{k|k-1}(x)|} \sum_{i \in I^N_{k|k-1}(x)} \rho^i_{k|k-1} \; ,$$

where $I^N_{k|k-1}(x) = \{i = 1, \cdots, N \; : \; \xi^i_{k|k-1} = x\}$,

update : with $\gamma_k \simeq k^{-2/3}$, set

$$\widehat{\theta}^N_k = \widehat{\theta}^N_{k-1} + \gamma_k \, [\sum_{i=1}^N \rho^i_{k|k-1} \, \omega^i_k \,] \; ,$$

and

$$\overline{\theta}^N_k = \overline{\theta}^N_{k-1} + \frac{1}{k} \, (\widehat{\theta}^N_k - \overline{\theta}^N_{k-1}) \; ,$$

mutation : independently for any $i = 1, \cdots, N$, simulate

$$\xi^i_k \sim \mu^N_k(dx) \; ,$$

and

$$(\xi^i_{k+1|k}, \Xi^i_{k+1}) \sim K^{\widehat{\theta}^N_k}(\xi^i_k, dx', ds') \; ,$$

and set

$$\rho^i_{k+1|k} = r^N_{k|k-1}(\xi^i_k) - c_k + \Xi^i_{k+1} \; ,$$

with normalization

$$c_k = \frac{1}{N} \sum_{j=1}^N r^N_{k|k-1}(\xi^j_k) \; ,$$

iteration : $k \longleftarrow k+1$, and return to the selection step.

Table 1. Particle implementation of the recursive MLE.