An Analysis of Regularized Interacting Particle Methods for Nonlinear Filtering[1]

François LeGland
IRISA / INRIA
Campus de Beaulieu
35042 RENNES Cédex, France
legland@irisa.fr

Christian Musso + Nadia Oudjane
ONERA / DTIM–MCT
29 avenue de la division Leclerc
92322 CHATILLON Cédex, France
{musso,oudjane}@onera.fr

Abstract

Interacting particle methods have been recently proposed for the approximation of nonlinear filters. These are efficient recursive Monte Carlo methods, which in principle could be implemented in high dimensional problems — i.e. which could beat the curse of dimensionality — and where the particles automatically concentrate in regions of interest of the state space. In this paper we show that it is sometimes necessary to add a regularization step, and we analyze the approximation error for the resulting regularized interacting particle methods.

1. Introduction

We consider the following model, where the unobserved state process $\{X_t\,,\,t\,\geq\,0\}$ satisfies the stochastic differential equation (SDE) on $\mathbf{R}^m$

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t\ , \qquad X_0 \sim \mu_0\ , \quad (1)$$

with standard Wiener process $\{W_t\,,\,t\,\geq\,0\}$, and where d–dimensional observations $\{z_n\,,\,n\,\geq\,1\}$ are available at discrete time instants $0 < t_1 < \cdots < t_n < \cdots$

$$z_n = h(X_{t_n}) + v_n\ ,$$

in additional white noise sequence, i.e. $\{v_n\,,\,n\,\geq\,1\}$ is an i.i.d. sequence (not necessarily Gaussian) with absolutely continuous probability distribution $g(v)\,dv$. We further assume that the initial state X_0, the Wiener process $\{W_t\,,\,t\,\geq\,0\}$, and the white noise sequence $\{v_n\,,\,n\,\geq\,1\}$ are mutually independent.

With the SDE (1) is associated the following second order partial differential operator

$$L = \sum_{i=1}^{m} b^i(\cdot)\,\frac{\partial}{\partial x_i} + \tfrac{1}{2}\,\sum_{i,j=1}^{m} a^{i,j}(\cdot)\,\frac{\partial^2}{\partial x_i \partial x_j}\ ,$$

with $a = (a^{i,j}) = \sigma\,\sigma^*$, and let $\{P_t\,,\,t\,\geq\,0\}$ be the corresponding Markov semigroup, i.e.

$$P_t\,\phi(x) = \mathbf{E}[\phi(X_{t+s}) \mid X_s = x]\ ,$$

for any test function ϕ defined on $\mathbf{R}^m$, and any time instant $s \geq 0$. Let $t_0 = 0$, and for any $n \geq 1$, let $Q_n = P_{t_n - t_{n-1}}$ denote the nonhomogeneous probability transition kernel for the Markov chain $\{X_{t_n}\,,\,n\,\geq\,0\}$, i.e.

$$Q_n\,\phi(x) = \mathbf{E}[\phi(X_{t_n}) \mid X_{t_{n-1}} = x]\ ,$$

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

Notice that

$$\mathbf{P}[z_n \in dz \mid X_{t_n} = x] = g(z - h(x))\,dz\ ,$$

hence the likelihood function for the estimation of the unknown state X_{t_n} based on the observation z_n alone, is given by

$$\Psi_n(x) = g(z_n - h(x))\ ,$$

for any $x \in \mathbf{R}^m$.

The purpose of nonlinear filtering is to estimate the unknown state at each time instant, from past observations. Optimal estimate in mean square sense is given by the conditional probability distribution μ_n of the state X_{t_n} given past observations, i.e. given $\mathcal{Z}_n = \sigma(z_1, \cdots, z_n)$. By definition

$$\mu_n(dx) = \mathbf{P}[X_{t_n} \in dx \mid \mathcal{Z}_n]\ .$$

We introduce also the conditional probability distribution $\mu_{n|n-1}$ of the state X_{t_n} given $\mathcal{Z}_{n-1}$, i.e.

$$\mu_{n|n-1}(dx) = \mathbf{P}[X_{t_n} \in dx \mid \mathcal{Z}_{n-1}]\ .$$

The sequence $\{\mu_n\,,\,n\,\geq\,0\}$ takes values in the space $\mathcal{P} = \mathcal{P}(\mathbf{R}^m)$ of probability distributions on $\mathbf{R}^m$, equipped with its Borel σ–field. The transition

[1]This work was partially supported by the CNRS programme *Modélisation et Simulation Numérique*.

from μ_{n-1} to μ_n, is conveniently described by the following two steps :

$$\mu_{n-1} \xrightarrow{\text{prediction}} \mu_{n|n-1} = Q_n^* \, \mu_{n-1}$$

$$\xrightarrow{\text{correction}} \mu_n = \Psi_n \cdot \mu_{n|n-1} \ ,$$

where $\cdot$ denotes the projective product on $\mathcal{P}$.

In the prediction step, $\mu_{n|n-1}$ is the value taken at time t_n by the solution (in weak sense) of the Fokker–Planck equation

$$\frac{\partial \mu_t^n}{\partial t} = L^* \, \mu_t^n \ , \qquad t \geq t_{n-1} \tag{2}$$

starting from μ_{n-1} at time t_{n-1}. Notice that

$$\langle \mu_{n|n-1}, \phi \rangle = \langle \mu_{n-1}, Q_n \, \phi \rangle$$

$$= \int_{\mathbf{R}^m} \mathbf{E}[\phi(X_{t_n}) \mid X_{t_{n-1}} = x] \, \mu_{n-1}(dx) \ ,$$

for any test function ϕ defined on $\mathbf{R}^m$. In the correction step, μ_n is simply given by the Bayes rule

$$\mu_n = \Psi_n \cdot \mu_{n|n-1} = \frac{\Psi_n \ \mu_{n|n-1}}{\langle \mu_{n|n-1}, \Psi_n \rangle} \ .$$

Let $W^{r,p}$ denote the Sobolev space of measurable functions ϕ defined on $\mathbf{R}^m$, whose derivatives up to order r (in the sense of distributions) belong to L^p. The norm $\| \cdot \|_{r,p}$ and the seminorm $| \cdot |_{r,p}$ are defined by

$$\|\phi\|_{r,p} = \Big\{ \sum_{|\alpha| \leq r} \int_{\mathbf{R}^m} |D^\alpha \, \phi(x)|^p \, dx \Big\}^{1/p} \ ,$$

and

$$|\phi|_{r,p} = \Big\{ \sum_{|\alpha| = r} \int_{\mathbf{R}^m} |D^\alpha \, \phi(x)|^p \, dx \Big\}^{1/p} \ ,$$

respectively. Throughout the paper, the notation $\| \cdot \|$ denotes the total variation norm on $\mathcal{P}$, or the supremum norm on the space of bounded measurable functions defined on $\mathbf{R}^m$, depending on the context.

The following *stability* result holds.

Lemma 1.1 *If μ_0 is absolutely continuous, with density $\frac{d\mu_0}{dx} \in W^{2,1}$, then under suitable regularity assumptions, μ_n and $\mu_{n|n-1}$ are absolutely continuous for any $n \geq 1$, with density $\frac{d\mu_n}{dx} \in W^{2,1}$ and $\frac{d\mu_{n|n-1}}{dx} \in W^{2,1}$ respectively. In addition*

$$|\frac{d\mu_{n|n-1}}{dx}|_{2,1} \leq H(t_n) \ \|\frac{d\mu_0}{dx}\|_{2,1} \ ,$$

where the mapping H is nondecreasing.

For any $\mu, \mu' \in \mathcal{P}$, and any bounded measurable function ϕ defined on $\mathbf{R}^m$, the following decomposition holds

$$\langle \Psi_n \cdot \mu' - \Psi_n \cdot \mu, \phi \rangle$$

$$= \frac{\langle \mu' - \mu, \Psi_n \, [\phi - \langle \Psi_n \cdot \mu, \phi \rangle \,] \rangle}{\langle \mu', \Psi_n \rangle} \ , \tag{3}$$

hence the following Lipschitz (rough) estimate.

Lemma 1.2 *If the likelihood function Ψ_n is bounded, and bounded away from zero, then for any $\mu, \mu' \in \mathcal{P}$*

$$\|\Psi_n \cdot \mu' - \Psi_n \cdot \mu\| \leq 2 \, \delta_n \, \|\mu' - \mu\| \ .$$

where $\delta_n = \dfrac{\max\limits_{x \in \mathbf{R}^m} \Psi_n(x)}{\min\limits_{x \in \mathbf{R}^m} \Psi_n(x)}$ is finite.

Remark 1.3 In principle, it should be possible to adapt the Lemmas A.1 and A.2 in LeGland and Mevel [1], and to get rid of the factor 2 in the above estimate.

Obviously, the most time consuming part is the prediction step, which requires the solution of the linear parabolic partial differential equation (2). Even though efficient numerical methods have been recently proposed, see e.g. Cai, LeGland and Zhang [2] which are based on implicit Euler time discretization scheme, and multigrid method with adaptive locally refined grid, these methods are limited to low–dimensional examples. To overcome the curse of dimensionality, interacting particle methods have been recently proposed in Gordon, Salmond and Smith [3] — see also Kitagawa [4] — to approximate nonlinear filtering equations, and have been thoroughly studied for discrete–time models, in Del Moral [5, 6] and in Del Moral and Guionnet [7, 8].

2. Interacting particle methods

For any probability distribution $\mu \in \mathcal{P}$, let $\mathcal{S}^N(\mu)$ denote the empirical distribution of an N–sample $\{\xi^i \, , \, i = 1, \cdots, N\}$ of i.i.d. random variables with common probability distribution μ, i.e.

$$\mathcal{S}^N(\mu) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i} \ .$$

The following classical estimate holds, where $\mathbf{E}_{\mathrm{MC}}$ denotes the expectation w.r.t. the simulated random variables only.

Lemma 2.1 *For any $\mu \in \mathcal{P}$, and any bounded measurable function ϕ defined on $\mathbf{R}^m$*

$$\mathbf{E}_{\mathrm{MC}} | \langle \mathcal{S}^N(\mu) - \mu, \phi \rangle | \leq \frac{1}{\sqrt{N}} \|\phi\| \ .$$

The approximate sequence $\{\mu_n^N \, , \, n \geq 0\}$ is related with the empirical distribution of an interacting particle system. The initial condition is $\mu_0^N = \mu_0$, and the transition from μ_{n-1}^N to μ_n^N is described by the following two steps :

$$\mu_{n-1}^N \xrightarrow[\text{prediction}]{\text{sampled}} \mu_{n|n-1}^N = \mathcal{S}^N(Q_n^* \, \mu_{n-1}^N)$$

$$\xrightarrow{\text{correction}} \mu_n^N = \Psi_n \cdot \mu_{n|n-1}^N \ .$$

In the sampled prediction step, one has to produce an N–sample $\{\xi_{n|n-1}^i \, , \, i = 1, \cdots, N\}$ of i.i.d. random variables with common probability distribution $Q_n^* \, \mu_{n-1}^N$. This can be achieved by the following algorithm :

(i) produce an N–sample $\{\xi_{n-1}^i \, , \, i = 1, \cdots, N\}$ of i.i.d. random variables with (discrete) common probability distribution μ_{n-1}^N,

(ii) independently for each $i = 1, \cdots, N$, generate a random variable $\xi_{n|n-1}^i$ with probability distribution $Q_n(\xi_{n-1}^i, \cdot)$.

The second step (ii) can be realized by solving independent copies of the SDE (1), i.e. $\xi_{n|n-1}^i$ is the value taken at time t_n by the solution of

$$dX_t^i = b(X_t^i) \, dt + \sigma(X_t^i) \, dW_t^i \ , \qquad t \geq t_{n-1} \quad (4)$$

starting from ξ_{n-1}^i at time t_{n-1}. In the correction step, μ_n^N is given by the Bayes rule, and since

$$\mu_{n|n-1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{n|n-1}^i}$$

is a discrete probability distribution, this reduces to

$$\mu_n^N = \sum_{i=1}^N \omega_{n|n-1}^i \, \delta_{\xi_{n|n-1}^i} \ ,$$

where the weights $\{\omega_{n|n-1}^i \, , \, i = 1, \cdots, N\}$ are proportional to the likelihood function evaluated at the particle locations, i.e.

$$\omega_{n|n-1}^i = \frac{\Psi_n(\xi_{n|n-1}^i)}{\sum_{j=1}^N \Psi_n(\xi_{n|n-1}^j)} \ ,$$

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

Remark 2.2 The effect of the resampling step (i) is to select the more likely particles, i.e. those which are associated with larger values of the likelihood function. Even though several copies of these particles could be present in the particle system $\{\xi_{n-1}^i \, , \, i = 1, \cdots, N\}$, one could rely on the independent noises in equations (4) to produce a sufficiently diverse particle system $\{\xi_{n|n-1}^i \, , \, i = 1, \cdots, N\}$. This raises however the question of noise–free state models, where the number of different particle locations can only decrease, and ultimately reduces to one. This motivates the introduction of regularized interacting particle methods, see Section 4 below, as a way to prevent the *degeneracy of particle locations* to occur.

Remark 2.3 If one would skip the resampling step (i), then the method would reduce to the weighted particle method first proposed by Davis [9], where the sequence of observations is used to update the weights associated with a non–interacting particle system. It was observed by LeGland [10] that this method is very unefficient, in the sense that after a few time steps, most of the particles get a negligible weight, and only a few (ultimately only one) particles significantly contribute to the approximation. Resampling the particle system (at each time step, or from time to time), is sufficient in most cases to prevent the *degeneracy of particle weights* to occur. This raises however the question of systems with small observation noise, where the degeneracy of particle weights can occur already at the first step, and where the resampling mechanism presented above is not sufficient. This motivates the introduction of alternate regularized interacting particle methods, see Section 5 below.

Remark 2.4 A more efficient implementation of the resampling mechanism has been proposed by Crişan [11] — see also Beadle and Djurić [12] — where the number of offsprings of each particle is proportional to its weight. In general, this number is not an integer, and there is an additional sample of a Bernoulli random variable, so as to decide between the two nearest integers, which results in a (slightly) varying number of particles. On the other hand, this branching mechanism runs much faster and provides minimum variance approximation. Alternative branching mechanisms can be found in Crişan, Del Moral and Lyons [13].

3. Time discretization scheme

Obviously, equations (4) cannot be solved explicitely, and one has to consider numerical approximations. For any $\Delta > 0$, any $x \in \mathbf{R}^m$, and any $w \in \mathbf{R}^d$, let

$$F_\Delta(x, w) = x + b(x) \, \Delta + \sigma(x) \, \sqrt{\Delta} \, w \ ,$$

and let $\Gamma_\Delta(x,\cdot)$ denote the Gaussian probability distribution with mean vector $m_\Delta(x) = x + b(x)\,\Delta$, and covariance matrix $\Sigma_\Delta(x) = a(x)\,\Delta$. In principle, the inter–observation time $t_n - t_{n-1}$ may be too large to be used directly as a time discretization step, and we introduce a subdivision $t_{n-1} = \tau_{n,0} < \tau_{n,1} < \cdots < \tau_{n,l_n} = t_n$ of the time interval $[t_{n-1}, t_n]$, with time steps $\Delta_{n,l} = \tau_{n,l} - \tau_{n,l-1}$ for $l = 1, \cdots, l_n$. The second step (ii) is now realized approximately, by running independent copies of the Euler scheme for equation (1), i.e. $\bar\xi^i_{n|n-1}$ is the value taken at step $l = l_n$ by the solution of the recursion equation

$$\bar X^i_{n,l} = F_{\Delta_{n,l}}(\bar X^i_{n,l-1}, \bar w^i_{n,l}) \ , \qquad l = 1, \cdots, l_n$$

starting from $\bar\xi^i_{n-1}$ at step $l = 0$, where $(\bar w^i_{n,1}, \cdots, \bar w^i_{n,l_n})$ are i.i.d. Gaussian random variables with zero mean and identity covariance matrix. Notice that for any $x \in \mathbf{R}^m$

$$\mathbf{P}[\bar X^i_{n,l} \in dx' \mid \bar X^i_{n,l-1} = x] = \Gamma_{\Delta_{n,l}}(x, dx') \ ,$$

for any $l = 1, \cdots, l_n$, hence

$$\mathbf{P}[\bar\xi^i_{n|n-1} \in dx' \mid \bar\xi^i_{n-1} = x] = \bar Q_n(x, dx') \ ,$$

where

$$\bar Q_n = \Gamma_{\Delta_{n,1}} \cdots \Gamma_{\Delta_{n,l_n}}$$

denotes the transition probability kernel associated with the Euler scheme. Under suitable assumptions, the probability distribution of the Euler scheme has bounded moments of any order.

For any $\mu, \mu' \in \mathcal{P}$, and any bounded measurable function ϕ defined on $\mathbf{R}^m$, the following decomposition holds

$$\begin{aligned} &\langle \bar Q^*_n\,\mu' - Q^*_n\,\mu, \phi\rangle \\ &\qquad = \langle \bar Q^*_n\,\mu - Q^*_n\,\mu, \phi\rangle + \langle \mu' - \mu, \bar Q_n\,\phi\rangle \ , \end{aligned}$$

hence

$$\begin{aligned} |\,\langle \bar Q^*_n\,\mu' - Q^*_n\,\mu, \phi\rangle\,| &\le \|\bar Q^*_n\,\mu - Q^*_n\,\mu\|\ \|\phi\| \\ &\quad + |\,\langle \mu' - \mu, \bar Q_n\,\phi\rangle\,| \ . \end{aligned} \tag{5}$$

The following error estimate is a direct consequence of the estimate proved in Bally and Talay [14, Theorem 3.1].

Lemma 3.1 *Under suitable regularity assumptions, for any $\mu \in \mathcal{P}$ and any $n \ge 1$*

$$\|\bar Q^*_n\,\mu - Q^*_n\,\mu\| \le M(t_n - t_{n-1})\ \Delta_n \ ,$$

where $\Delta_n = \max_{l=1,\cdots,l_n} \Delta_{n,l}$, and where the mapping M is nondecreasing.

Let $\{\bar\mu^N_n\,,\ n \ge 0\}$ denote the resulting fully discretized approximate sequence. The initial condition is $\bar\mu^N_0 = \mu_0$, and the transition from $\bar\mu^N_{n-1}$ to $\bar\mu^N_n$ is decribed by the following steps, which can be implemented in a straightforward manner :

$$\begin{aligned} \bar\mu^N_{n-1} &\xrightarrow[\text{discretized}\\ \text{prediction}]{\text{sampled}} \bar\mu^N_{n|n-1} = \mathcal{S}^N(\bar Q^*_n\ \bar\mu^N_{n-1}) \\ &\xrightarrow{\text{correction}} \bar\mu^N_n = \Psi_n \cdot \bar\mu^N_{n|n-1} \ . \end{aligned}$$

4. Regularized interacting particle methods

Let K be a symmetric probability density function (kernel) on $\mathbf{R}^m$, such that

$$\int_{\mathbf{R}^m} K(x)\,dx = 1 \ , \qquad \int_{\mathbf{R}^m} x\,K(x)\,dx = 0 \ ,$$

and

$$\int_{\mathbf{R}^m} |x|^2\,K(x)\,dx < \infty \ ,$$

and for any bandwidth $h > 0$, let the mollifier (approximation of identity) K_h be defined by the following scaling

$$K_h(x) = \frac{1}{h^m}\ K(\frac{x}{h}) \ ,$$

for any $x \in \mathbf{R}^m$. For any $\mu \in \mathcal{P}$, let $K_h * \mu$ be the absolutely continuous probability distribution on $\mathbf{R}^m$, with density

$$\frac{d(K_h * \mu)}{dx}(x) = \int_{\mathbf{R}^m} K_h(x - x')\,\mu(dx') \ ,$$

i.e. such that

$$\begin{aligned} &\langle K_h * \mu, \phi\rangle \\ &\quad = \int_{\mathbf{R}^m} \Big\{ \int_{\mathbf{R}^m} K_h(x - x')\,\mu(dx') \Big\}\ \phi(x)\,dx \\ &\quad = \int_{\mathbf{R}^m} \Big\{ \int_{\mathbf{R}^m} K_h(x' - x)\,\phi(x)\,dx \Big\}\ \mu(dx') \\ &\quad = \langle \mu, K_h * \phi\rangle \ , \end{aligned}$$

for any test function ϕ defined on $\mathbf{R}^m$. Notice that for any $\mu, \mu' \in \mathcal{P}$, and any bounded measurable function ϕ defined on $\mathbf{R}^m$

$$\begin{aligned} |\,\langle K_h * \mu - K_h * \mu', \phi\rangle\,| &= |\,\langle \mu - \mu', K_h * \phi\rangle\,| \\ &\le \|\mu - \mu'\|\ \|K_h * \phi\| \le \|\mu - \mu'\|\ \|\phi\| \ , \end{aligned}$$

hence

$$\|K_h * \mu - K_h * \mu'\| \le \|\mu - \mu'\| \ .$$

The following bias estimate is proved in Raviart [15, Lemma 4.4] — see also Holmström and Klemelä [16, Proposition 4].

Lemma 4.1 *If $\mu \in \mathcal{P}$ is absolutely continuous, with density $\frac{d\mu}{dx} \in W^{2,1}$, then for any bounded measurable function ϕ defined on $\mathbf{R}^m$*

$$| \langle K_h * \mu - \mu, \phi \rangle | \leq C\ h^2\ |\frac{d\mu}{dx}|_{2,1}\ \|\phi\| \ ,$$

hence

$$\|K_h * \mu - \mu\| \leq C\ h^2\ |\frac{d\mu}{dx}|_{2,1} \ .$$

For any $\mu \in \mathcal{P}$, let $J(\mu)$ denote the $(m+1)$–th (noncentered) moment, i.e.

$$J(\mu) = \int_{\mathbf{R}^m} |x|^{m+1}\ \mu(dx) \ ,$$

which could be finite or infinite. Assuming that the kernel K satisfies

$$V = \int_{\mathbf{R}^m} (1 + |x|^{m+1})\ K^2(x)\, dx < \infty \ ,$$

the following strong estimate holds, which improves on the variation estimate of Holmström and Klemelä [16, Proposition 8]. The proof is omitted.

Lemma 4.2 *For any $\mu \in \mathcal{P}$ such that $J(\mu)$ is finite*

$$\mathbf{E}_{\mathrm{MC}} \| K_h * \mathcal{S}^N(\mu) - K_h * \mu \|$$
$$\leq \frac{1}{\sqrt{N\ h^m}}\ C'\ (1 + \sqrt{J(\mu)}) \ .$$

The approximate sequence $\{\bar{\mu}_n^{N,h}\,,\ n \geq 0\}$ is obtained by regularizing before the prediction step. The initial condition is $\bar{\mu}_0^{N,h} = \mu_0$, and the transition from $\bar{\mu}_{n-1}^{N,h}$ to $\bar{\mu}_n^{N,h}$ is decribed by the following steps, which can be implemented in a straightforward manner :

$$\bar{\mu}_{n-1}^{N,h} \xrightarrow{\text{regularization}} \bar{\nu}_{n-1}^{N,h} = K_h * \bar{\mu}_{n-1}^{N,h}$$

$$\xrightarrow[\text{discretized}]{\text{sampled}}_{\text{prediction}} \bar{\mu}_{n|n-1}^{N,h} = \mathcal{S}^N(\bar{Q}_n^*\ \bar{\nu}_{n-1}^{N,h})$$

$$\xrightarrow{\text{correction}} \bar{\mu}_n^{N,h} = \Psi_n \cdot \bar{\mu}_{n|n-1}^{N,h} \ .$$

Considering instead the absolutely continuous approximate sequence $\{\bar{\nu}_n^{N,h}\,,\ n \geq 0\}$, the initial condition is $\bar{\nu}_0^{N,h} = K_h * \mu_0$, and the transition from $\bar{\nu}_{n-1}^{N,h}$ to $\bar{\nu}_n^{N,h}$ is decribed by the following steps :

$$\bar{\nu}_{n-1}^{N,h} \xrightarrow[\text{discretized}]{\text{sampled}}_{\text{prediction}} \bar{\mu}_{n|n-1}^{N,h} = \mathcal{S}^N(\bar{Q}_n^*\ \bar{\nu}_{n-1}^{N,h})$$

$$\xrightarrow{\text{correction}} \bar{\mu}_n^{N,h} = \Psi_n \cdot \bar{\mu}_{n|n-1}^{N,h}$$

$$\xrightarrow{\text{regularization}} \bar{\nu}_n^{N,h} = K_h * \bar{\mu}_n^{N,h} \ .$$

In the sampled (discretized) prediction step, one has to produce an N–sample $\{\bar{\xi}_{n|n-1}^i\,,\ i = 1, \cdots, N\}$ of i.i.d. random variables with common probability distribution $\bar{Q}_n^*\ \bar{\nu}_{n-1}^{N,h}$. This can be achieved by the following algorithm :

(i) produce an N–sample $\{\bar{\xi}_{n-1}^i\,,\ i = 1, \cdots, N\}$ of i.i.d. random variables with common (discrete) probability distribution $\bar{\mu}_{n-1}^{N,h}$,

(ii) independently for each $i = 1, \cdots, N$, generate a random variable ε^i with probability distribution K, and set $\bar{\eta}_{n-1}^i = \bar{\xi}_{n-1}^i + h\,\varepsilon^i$,

(iii) independently for each $i = 1, \cdots, N$, generate a random variable $\bar{\xi}_{n|n-1}^i$ with probability distribution $\bar{Q}_n(\bar{\eta}_{n-1}^i, \cdot)$.

In the correction step, $\bar{\mu}_n^{N,h}$ is given by the Bayes rule, and since

$$\bar{\mu}_{n|n-1}^{N,h} = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{\xi}_{n|n-1}^i}$$

is a discrete probability distribution, this reduces to

$$\bar{\mu}_n^{N,h} = \sum_{i=1}^N \bar{\omega}_{n|n-1}^i\ \delta_{\bar{\xi}_{n|n-1}^i} \ ,$$

with weights

$$\bar{\omega}_{n|n-1}^i = \frac{\Psi_n(\bar{\xi}_{n|n-1}^i)}{\sum_{j=1}^N \Psi_n(\bar{\xi}_{n|n-1}^j)} \ ,$$

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

$$\frac{d\bar{\nu}_n^{N,h}}{dx}(x) = \sum_{i=1}^N \bar{\omega}_{n|n-1}^i\ K_h(x - \bar{\xi}_{n|n-1}^i) \ .$$

5. Alternative regularization scheme

Alternatively, one can consider the absolutely continuous approximate sequence $\{\bar{\mu}_n^{N,h}\,,\ n \geq 0\}$, obtained by regularizing before the correction step.

The initial condition is $\bar{\mu}_0^{N,h} = \mu_0$, and the transition from $\bar{\mu}_{n-1}^{N,h}$ to $\bar{\mu}_n^{N,h}$ is decribed by the following steps :

$$\bar{\mu}_{n-1}^{N,h} \xrightarrow[\text{discretized prediction}]{\text{sampled}} \bar{\mu}_{n|n-1}^{N,h} = \mathcal{S}^N(\bar{Q}_n^* \, \bar{\mu}_{n-1}^{N,h})$$

$$\xrightarrow{\text{regularization}} \bar{\nu}_{n|n-1}^{N,h} = K_h * \bar{\mu}_{n|n-1}^{N,h}$$

$$\xrightarrow{\text{correction}} \bar{\mu}_n^{N,h} = \Psi_n \cdot \bar{\nu}_{n|n-1}^{N,h} \; .$$

In the sampled (discretized) prediction step, one has to produce an N–sample $\{\bar{\xi}_{n|n-1}^i \, , \, i = 1, \cdots, N\}$ of i.i.d. random variables with common probability distribution $\bar{Q}_n^* \; \bar{\mu}_{n-1}^{N,h}$. This can be achieved along the following steps :

(i) produce an N–sample $\{\bar{\xi}_{n-1}^i \, , \, i = 1, \cdots, N\}$ of i.i.d. random variables with common probability distribution $\bar{\mu}_{n-1}^{N,h}$,

(ii) independently for each $i = 1, \cdots, N$, generate a random variable $\bar{\xi}_{n|n-1}^i$ with probability distribution $\bar{Q}_n(\bar{\xi}_{n-1}^i, \cdot)$.

In the correction step, $\bar{\mu}_n^{N,h}$ is given by the Bayes rule, and since $\bar{\nu}_{n|n-1}^{N,h}$ is an absolutely continuous probability distribution with density

$$\frac{d\bar{\nu}_{n|n-1}^{N,h}}{dx}(x) = \frac{1}{N} \sum_{i=1}^{N} K_h(x - \bar{\xi}_{n|n-1}^i) \; ,$$

this reduces to

$$\frac{d\bar{\mu}_n^{N,h}}{dx}(x) = \frac{\Psi_n(x) \sum_{i=1}^{N} K_h(x - \bar{\xi}_{n|n-1}^i)}{\sum_{i=1}^{N} K_h * \Psi_n(\bar{\xi}_{n|n-1}^i)} \; .$$

In step (i) of the algorithm, one has to generate random variables ξ with absolutely continuous probability distribution $\bar{\mu}_n^{N,h}$. This can be achieved by the following *rejection* algorithm, see Devroye [17, page 47] :

(i-1) pick $j \in \{1, \cdots, N\}$ with uniform probability, generate a random variable ε with probability distribution K, and set $\xi = \bar{\xi}_{n|n-1}^j + h\,\varepsilon$,

(i-2) generate a uniform random variable U on the interval $[0, 1]$,

(i-3) return to step (i-1) unless $\Psi_n(\xi) \geq U \; \Psi_n^{\max}$,

where $\Psi_n^{\max} = \max_{x \in \mathbf{R}^m} \Psi_n(x) = \max_{x \in \mathbf{R}^m} g_n(z_h - h(x))$.

6. Error estimates (weak sense)

The following decomposition holds

$$\langle K^h * \mathcal{S}^N(\mu') - \mu, \phi \rangle = \langle \mathcal{S}^N(\mu') - \mu', K^h * \phi \rangle$$

$$+ \langle \mu' - \mu, K^h * \phi \rangle + \langle K^h * \mu - \mu, \phi \rangle \; ,$$

and combining the estimates of Lemmas 2.1 and 4.1 yields the following (weak sense) estimate.

Lemma 6.1 *For any $\mu, \mu' \in \mathcal{P}$, such that μ is absolutely continuous, with density $\frac{d\mu}{dx} \in W^{2,1}$, and for any bounded measurable function ϕ defined on $\mathbf{R}^m$*

$$\mathbf{E}_{\mathrm{MC}} | \, \langle K^h * \mathcal{S}^N(\mu') - \mu, \phi \rangle \, | \leq | \, \langle \mu' - \mu, K^h * \phi \rangle \, |$$

$$+ [\, \frac{1}{\sqrt{N}} + C \; h^2 \, |\frac{d\mu}{dx}|_{2,1} \,] \, \|\phi\| \; .$$

For the approximation defined in Section 5, the following error estimate holds.

Theorem 6.2 *Under suitable regularity assumptions*

$$\sup_{\|\phi\|=1} \mathbf{E}_{\mathrm{MC}} | \, \langle \bar{\mu}_n^{N,h} - \mu_n, \phi \rangle \, |$$

$$\leq 2^n \; \delta_n \cdots \delta_1 \, [M_n^{\max} \max_{k=1,\cdots,n} \Delta_k + \frac{1}{\sqrt{N}}$$

$$+ H_n^{\max} \, C \, h^2 \, \|\frac{d\mu_0}{dx}\|_{2,1} \,] \; ,$$

where $M_n^{\max} = \max_{k=1,\cdots,n} M(t_k - t_{k-1})$, and $H_n^{\max} = \max_{k=1,\cdots,n} H(t_k) = H(t_n)$.

Proof: Let $\varepsilon_n = \sup_{\|\phi\|=1} \mathbf{E}_{\mathrm{MC}} | \, \langle \bar{\mu}_n^{N,h} - \mu_n, \phi \rangle \, |$ for any $n \geq 0$.

For the approximation defined in Section 5, the decomposition (3) yields

$$\langle \bar{\mu}_n^{N,h} - \mu_n, \phi \rangle$$

$$= \frac{\langle K_h * \mathcal{S}^N(\bar{Q}_n^* \, \bar{\mu}_{n-1}^{N,h}) - Q_n^* \, \mu_{n-1}, u_n \rangle}{\langle \bar{\mu}_n^{N,h}, \Psi_n \rangle} \; ,$$

with $u_n = \Psi_n \, [\phi - \langle \mu_n, \phi \rangle]$, hence

$$| \, \langle \bar{\mu}_n^{N,h} - \mu_n, \phi \rangle \, |$$

$$\leq \frac{1}{\Psi_n^{\min}} \, | \, \langle K_h * \mathcal{S}^N(\bar{Q}_n^* \, \bar{\mu}_{n-1}^{N,h}) - Q_n^* \, \mu_{n-1}, u_n \rangle \, | \; .$$

Notice that u_n does not depend on any of the simulated random variables, and

$$\|\bar{Q}_n \, (K_h * u_n) \, \| \leq \|K_h * u_n\| \leq \|u_n\| \leq 2 \, \Psi_n^{\max} \, \|\phi\| \; .$$

Taking expectation w.r.t. simulated random variables used in the transition from $\bar{\mu}_{n-1}^{N,h}$ to $\bar{\mu}_{n|n-1}^{N,h}$ only, i.e. conditioning w.r.t. the σ–field $\mathcal{F}_{n-1}$ generated by the simulated random variables used in the transition from $\bar{\mu}_0^{N,h}$ to $\bar{\mu}_{n-1}^{N,h}$, and using the estimate of Lemma 6.1, yields

$$\mathbf{E}_{\mathrm{MC}}[\,|\,\langle \bar{\mu}_n^{N,h} - \mu_n, \phi\rangle\,|\,|\,\mathcal{F}_{n-1}\,]$$

$$\leq \frac{1}{\Psi_n^{\min}}\,|\,\langle \bar{Q}_n^*\,\bar{\mu}_{n-1}^{N,h} - Q_n^*\,\mu_{n-1}, K_h * u_n\,\rangle\,|$$

$$+ \frac{1}{\Psi_n^{\min}}\,[\,\frac{1}{\sqrt{N}} + C\,h^2\,|\frac{d\mu_{n|n-1}}{dx}|_{2,1}\,]\,\|u_n\|\,.$$

Using the estimate (5) yields

$$|\,\langle \bar{Q}_n^*\,\bar{\mu}_{n-1}^{N,h} - Q_n^*\,\mu_{n-1}, K_h * u_n\rangle\,|$$

$$\leq \|\bar{Q}_n^*\,\mu_{n-1} - Q_n^*\,\mu_{n-1}\|\,\|K_h * u_n\|$$

$$+\,|\,\langle \bar{\mu}_{n-1}^{N,h} - \mu_{n-1}, \bar{Q}_n\,(K_h * u_n)\,\rangle\,|\,.$$

Combining these estimates and the estimates in Lemmas 1.1 and 3.1 yields

$$\mathbf{E}_{\mathrm{MC}}[\,|\,\langle \bar{\mu}_n^{N,h} - \mu_n, \phi\rangle\,|\,|\,\mathcal{F}_{n-1}\,]$$

$$\leq \frac{1}{\Psi_n^{\min}}\,|\,\langle \bar{\mu}_{n-1}^{N,h} - \mu_{n-1}, \bar{Q}_n\,(K_h * u_n)\,\rangle\,|$$

$$+ 2\,\delta_n\,[\,M(t_n - t_{n-1})\,\Delta_n + \frac{1}{\sqrt{N}}$$

$$+ H(t_n)\,C\,h^2\,\|\frac{d\mu_0}{dx}\|_{2,1}\,]\,\|\phi\|\,.$$

Notice that

$$\mathbf{E}_{\mathrm{MC}}|\,\langle \bar{\mu}_{n-1}^{N,h} - \mu_{n-1}, \bar{Q}_n\,(K_h * u_n)\,\rangle\,|$$

$$\leq \varepsilon_{n-1}\,\|\bar{Q}_n\,(K_h * u_n)\,\| \leq \varepsilon_{n-1}\,2\,\Psi_n^{\max}\,\|\phi\|\,.$$

Taking expectation w.r.t. simulated random variables yields

$$\mathbf{E}_{\mathrm{MC}}|\,\langle \bar{\mu}_n^{N,h} - \mu_n, \phi\rangle\,|$$

$$\leq 2\,\delta_n\,[\,\varepsilon_{n-1} + M(t_n - t_{n-1})\,\Delta_n + \frac{1}{\sqrt{N}}$$

$$+ H(t_n)\,C\,h^2\,\|\frac{d\mu_0}{dx}\|_{2,1}\,]\,\|\phi\|\,,$$

hence the sequence $\{\varepsilon_n\,,\,n \geq 0\}$ satisfies

$$\varepsilon_n \leq 2\,\delta_n\,[\,\varepsilon_{n-1} + M(t_n - t_{n-1})\,\Delta_n + \frac{1}{\sqrt{N}}$$

$$+ H(t_n)\,C\,h^2\,\|\frac{d\mu_0}{dx}\|_{2,1}\,]\,,$$

and the result follows from the discrete Gronwall lemma. ■

Remark 6.3 When $h = 0$, one recovers the result obtained by Del Moral, Jacod and Protter [18]. Our result says that the error estimate for regularized schemes — which have qualitatively more stable behaviour, for reasons explained at the end of Section 2 — exhibits an additional term which can be controlled in terms of the bandwidth parameter h.

Remark 6.4 As already mentioned in [18], the constants in the estimate of Theorem 6.2 depend exponentially on time. In principle, it should be possible to obtain similar estimates, with uniformly bounded constants over an infinite time interval, provided some stability property holds for the filtering process, such as exponential forgetting of its initial condition. Some results in this direction have been obtained recently by Del Moral and Guionnet [19] for models which unfortunately do not include the model considered here.

7. Error estimates (strong sense)

The following decomposition holds

$$K_h * \mathcal{S}^N(\mu') - \mu = (K_h * \mathcal{S}^N(\mu') - K_h * \mu')$$

$$+ K_h * (\mu' - \mu) + (K_h * \mu - \mu)$$

and combining the estimates of Lemmas 4.2 and 4.1 yields the following (strong sense) estimate.

Lemma 7.1 *For any $\mu \in \mathcal{P}$, such that μ is absolutely continuous, with density $\frac{d\mu}{dx} \in W^{2,1}$, for any $\mu' \in \mathcal{P}$, such that $J(\mu')$ is finite, and for any bounded measurable function ϕ defined on $\mathbf{R}^m$*

$$\mathbf{E}_{\mathrm{MC}}\|\,K_h * \mathcal{S}^N(\mu') - \mu\,\| \leq \|\mu' - \mu\|$$

$$+ [\,\frac{1}{\sqrt{N\,h^m}}\,C'\,(1 + \sqrt{J(\mu')}) + C\,h^2\,|\frac{d\mu}{dx}|_{2,1}\,]\,.$$

For the approximation defined in Section 5, the following error estimate holds. The proof is omitted.

Theorem 7.2 *Under suitable regularity assumptions, the probability distribution of the Euler scheme has bounded $(m+1)$–th moments, i.e. $J_n^{\max} = \max_{k=1,\cdots,n} \mathbf{E}_{\mathrm{MC}}[J(\bar{Q}_k^*\,\bar{\mu}_{k-1}^{N,h})\,]$ is finite, and*

$$\mathbf{E}_{\mathrm{MC}}\|\,\bar{\mu}_n^{N,h} - \mu_n\,\|$$

$$\leq 2^n\,\delta_n \cdots \delta_1\,[M_n^{\max} \max_{k=1,\cdots,n} \Delta_k$$

$$+ \frac{1}{\sqrt{N\,h^m}}\,C'\,(1 + \sqrt{J_n^{\max}})$$

$$+ H_n^{\max}\,C\,h^2\,\|\frac{d\mu_0}{dx}\|_{2,1}\,]\,,$$

where $M_n^{\max} = \max_{k=1,\cdots,n} M(t_k - t_{k-1})$, *and* $H_n^{\max} = \max_{k=1,\cdots,n} H(t_k) = H(t_n)$.

Remark 7.3 Obviously one cannot take $h = 0$ in the estimate of Theorem 7.2. In any case, when $h = 0$, the probability distributions μ_n and $\bar{\mu}_n^{N,h}$ are respectively absolutely continuous and discrete, and it does not make sense to study the total variation distance $\|\bar{\mu}_n^{N,h} - \mu_n\|$. Instead, one should consider weak sense estimates, as in Theorem 6.2.

Remark 7.4 The question of bandwidth selection arises there. The optimal bandwidth could not be computed, since it would depend on some unknown quantities. On the other hand, to make the bias, variation, and time discretization contribute equally to the global error, the bandwidth should be taken asymptotically as $h \sim N^{-1/(m+4)}$.

References

[1] F. LE GLAND and L. MEVEL, "Basic properties of the projective product, with application to products of column–allowable nonnegative matrices." (submitted).

[2] CAI Zhiqiang, F. LE GLAND, and ZHANG Huilong, "An adaptive local grid refinement method for nonlinear filtering," Rapport de Recherche 2679, INRIA, Oct. 1995.

[3] N. GORDON, D. SALMOND, and A. SMITH, "Novel approach to nonlinear/non–Gaussian Bayesian state estimation," *IEE Proceedings, Part F*, vol. 140, pp. 107–113, Apr. 1993.

[4] G. KITAGAWA, "Monte Carlo filter and smoother for non–Gaussian nonlinear state space models," *Journal on Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[5] P. DEL MORAL, "Measure valued processes and interacting particle systems. Application to nonlinear filtering problems," Publication du Laboratoire de Statistique et Probabilités 15–96, Université Paul Sabatier, Toulouse, 1996.

[6] P. DEL MORAL, "Nonlinear filtering : interacting particle solution," *Markov Processes and Related Fields*, vol. 2, no. 4, pp. 555–580, 1996.

[7] P. DEL MORAL and A. GUIONNET, "Large deviations for interacting particle systems. Applications to nonlinear filtering problems," Publication du Laboratoire de Statistique et Probabilités 05–97, Université Paul Sabatier, Toulouse, 1997.

[8] P. DEL MORAL and A. GUIONNET, "A central limit theorem for nonlinear filtering using interacting particle systems," Publication du Laboratoire de Statistique et Probabilités 11–97, Université Paul Sabatier, Toulouse, 1997.

[9] M. DAVIS, "New approach to filtering for nonlinear systems," *IEE Proceedings, Part D*, vol. 128, no. 5, pp. 166–172, 1981.

[10] F. LE GLAND, "Monte–Carlo methods in nonlinear filtering," in *Proceedings of the 23rd IEEE Conference on Decision and Control, Las Vegas 1984*, pp. 31–32, IEEE–CSS, Dec. 1984.

[11] D. CRIŞAN, *The Problem of Nonlinear Filtering.* Ph.D. Thesis, Department of Mathematics, University of Edinburgh, Edinburgh, July 1996.

[12] E. BEADLE and P. DJURIĆ, "A fast weighted Bayesian bootstrap filter for nonlinear model state estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES–33, pp. 338–343, Jan. 1997.

[13] D. CRIŞAN, P. DEL MORAL, and T. LYONS, "Discrete filtering using branching and interacting particle systems," Publication du Laboratoire de Statistique et Probabilités 1–98, Université Paul Sabatier, Toulouse, 1998.

[14] V. BALLY and D. TALAY, "The law of the Euler scheme for stochastic differential equations. I : Convergence rate of the distribution function," *Probability Theory and Related Fields*, vol. 104, no. 1, pp. 43–60, 1996.

[15] P. RAVIART, "An analysis of particle methods," in *Numerical Methods in Fluid Dynamics, Como 1983* (F. Brezzi, ed.), vol. 1127 of *Lecture Notes in Mathematics*, pp. 243–324, Berlin: Springer Verlag, 1985.

[16] L. HOLMSTRÖM and J. KLEMELÄ, "Asymptotic bounds for the expected L_1 error of a multivariate kernel density estimator," *Journal of Multivariate Analysis*, vol. 42, no. 2, pp. 245–266, 1992.

[17] L. DEVROYE, *Non–Uniform Random Variate Generation.* New York: Springer Verlag, 1986.

[18] P. DEL MORAL, J. JACOD, and P. PROTTER, "The Monte–Carlo method for filtering with discrete–time observations." (preprint).

[19] P. DEL MORAL and A. GUIONNET, "On the stability of measure valued processes. Applications to nonlinear filtering and interacting particle systems," Publication du Laboratoire de Statistique et Probabilités 3–98, Université Paul Sabatier, Toulouse, 1998.