Nonlinear Filtering: Interacting Particle Resolution

P. DEL MORAL†

Abstract

This paper cover stochastic particle methods for the numerical solving of the nonlinear filtering equations based upon the simulation of interacting particle systems. The main contribution of this paper is to prove the convergences of such approximations to the optimal filter, yielding what seemed to be the first convergence results for such approximations of the nonlinear filtering equations. This new treatment was influenced primarily by the development of genetic algorithms (J.H. Holland [11], R. Cerf [2]) and secondarily by the papers of H.Kunita and L.Stettner ([12], [13]). Such interacting particle resolutions encompass genetic algorithms, incidentally our models provide essential insight for the analysis of genetic algorithms with a non homogeneous fitness function with respect to time.

Introduction

The basic model for the general nonlinear filtering problem consists of a nonlinear plant X with state noise W and nonlinear observation Y with observation noise V. Let (X,Y) be the Markov process taking values in $S\times\mathbb{R}$ and defined by the system:

$$\mathcal{F}(X/Y) \qquad \begin{cases} X\sim(\nu,K) \\ Y_n = h\left(X_n\right)+V_n, \quad n\geq 1 \end{cases} \tag{1}$$

where $S=\mathbb{R}^d$, $d\geq 1$, $h:S\rightarrow\mathbb{R}$ and V_n are independent random variables having a density g_n with respect to Lebesgue measure. The signal process X that we consider is assumed to be a temporally homogeneous Markov process with transition probability kernel K and initial probability measure ν, on S. We assume the observation noise V and the state plant X are independent. For simplicity the observation process Y is real valued, the extension to vector observation is straightforward. One can study filtering problems in more general settings. We choose not to do so there, preferring to focus on the central ideas and the methods used in this paper can be extended if desired.
The filtering problem is concerned with estimating a functional of the state process using the information contained in the observation process Y. The information is encoded in the filtration defined by the sigma algebra generated by the observations $Y_1,\ldots,Y_n$. Let f be an integrable

†Laboratoire de Statistiques et Probabilités, CNRS-UMR C55830, Bat 1R1, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, electronic adress: delmoral@cict.fr
AMS 1994 subject classifications: *60F10, 60G35, 60J10, 62D05, 62E25, 62F12, 62G05, 62L20, 62M05 92D15, 93E10, 93E11*
keywords: *Nonlinear Filtering, Bayesian Estimation, Monte-Carlo Sampling, Interacting Particle Systems, Sampling-Resampling, Evolutionary Processes, Genetic Algorithms.*
I am deeply grateful to Laure Coutin and Pierre Vandekerkhove for her careful reading of the original manuscript. I also thank the anonymous referee for his comments and suggestions which greatly improved the presentation of the paper.

Borel test function from S into $\mathbb{R}$, the best estimate of X_n given the observations up to time n is the conditional expectation:

$$\pi_n(f) \stackrel{def}{=} E(f(X_n)/Y^n) \qquad Y^n \stackrel{def}{=} (Y_1, \ldots, Y_n)$$

With the notable exception of the linear-Gaussian situation general optimal filters have no finitely recursive solution [3]. This paper cover stochastic particle methods for the numerical solving of the nonlinear filtering equations based upon the simulation of interacting particle systems. Such algorithms are an extension of the Sampling/Resampling (S/R) principles introduced by Gordon, Salmon and Smith in [10] and independently by Del Moral, Noyer, Rigal and Salut in [4] and [5]. Several practical problems examples which can be solved using these methods are given in [1] and [8], including problems in Radar/Sonar signal processing and GPS/INS integration. Such particle nonlinear filters will differ from the others ([6], [7]) in the way they store and update the information that is accumulated through the resampling of the positions.
We start by giving some general notations and we recall some basic facts related to the theory introduced by Kunita and Stettner([12], [13]). In section 2 we introduce the interacting particle approximation and we design a natural stochastic basis for the study of the convergence. In section 3 we describe recursive formulas for the conditional distributions and the context we are interested in. We propose a natural framework which allows to formulate explicitly mean error bounds in terms of the likelihood functions related to the resampling/selection procedure. The hardest point of our program is contained in section 4: the study of the convergence of our algorithm requires specific development because of the difficulty to compute mean error estimates which are essential to perform convergence rates. The key idea is to introduce in the mean square error estimates a martingale with unit mean using the functions g_n. This martingale approach simplifies drastically the evaluation of convergence rates discussed in section 5. This last section contains our main result: We prove that the interacting particle filters converges to the conditional distribution when the number of particles tends to infinity. The convergence rate estimates arise quite naturally from the results and associated methodologies of section 3 and 4.

1 Non Linear Filtering Equation

1.1 General Notations

Before starting out on the description of the non linear filtering equation let us first introduce some general notations.
Let $\mathcal{C}(S)$ be the space of bounded continuous functions on S with norm $\|f\|_\infty = \sup_{x\in S}|f(x)|$.
Let $\mathcal{P}(S)$ be the space of all probability measures on S in which the weak topology is induced.

$$\lim_{n\to+\infty} \mu_n = \mu \quad \text{in} \quad \mathcal{P}(S) \Longleftrightarrow \forall f \in \mathcal{C}(S) \quad \lim_{n\to+\infty} \int f\,\mu_n = \int f\,\mu$$

Let $\mu \in \mathcal{P}(S)$, $f \in \mathcal{C}(S)$ and, let K_1 and K_2 be two Markov kernels. We will use the standard notations

$$\mu K_1(dy) = \int \mu(dx)\, K_1(x,dy) \qquad K_1K_2(x,dz) = \int K_1(x,dy)\, K_2(y,dz) \tag{2}$$

$$K_1 f(x) = \int K_1(x,dy)\, f(y) \qquad \mu f = \int \mu(dx)\, f(x) \tag{3}$$

With $m \in \mathcal{P}(S^2)$ we associate two measures $\overline{m}$ and $\underline{m} \in \mathcal{P}(S)$ as follows

$$\forall f \in \mathcal{C}(S) \qquad \overline{m}f = \int m(dx_1,dx_2)\, f(x_2) \qquad \underline{m}f = \int m(dx_1,dx_2)\, f(x_1)$$

With a Markov kernel K and a measure $\mu \in\in \mathcal{P}(S)$ we associate a measure $\mu \times K \in\in \mathcal{P}(S^2)$ by setting

$$\forall f \in \mathcal{C}(S^2) \qquad (\mu \times K)f = \int \mu(dx_1)\, K(x_1, dx_2)\, f(x_1, x_2)$$

Finally we denote $\mathcal{C}(\mathcal{P}(S))$ the space of bounded continuous functions $F : \mathcal{P}(S) \to \mathbb{R}$

1.2 Recursive Filters and Bayes' Formula

In this section we describe recursive expressions for the conditional distribution of X_n and (X_n, X_{n+1}) given the observations $Y^n = (Y_1, \ldots, Y_n)$. Let $(\Omega = \Omega_1 \times \Omega_2, F_n, P)$ be the canonical space for the signal observation pair (X, Y). Therefore P is the probability measure on Ω corresponding to the filtering model $\mathcal{F}(X/Y)$ when

- ν is the probability measure of X_0.
- the marginal of P on Ω_1 is the law of X.
- $V_n = Y_n - h(X_n)$ is a sequence of independent random variables with densities g_n.

We use $E(\bullet)$ to denote the expectations with respect to P on Ω. To clarify the presentation we will also note

$$m_{n+1} = \pi_n \times K \tag{4}$$

Using Bayes' Theorem we see that the conditional distribution of X_n given the observations up to time n is given by

$$\pi_n = \rho_n(\pi_{n-1}, Y_n) \quad n \geq 1 \qquad \pi_0 = \nu \tag{5}$$

where

$$\rho_n(\mu, y)f = \frac{\displaystyle\int f(x)\, g_n(y - h(x))\, (\mu\, K)(dx)}{\displaystyle\int g_n(y - h(x))\, (\mu\, K)(dx)} \tag{6}$$

for all $f \in \mathcal{C}(S)$, $\mu \in \mathcal{P}(S)$, $y \in \mathbb{R}$ and $n \geq 1$. Much more is true. Following Kunita and Stettner([12], [13]) the above description enables us to consider the conditional distributions π_n as a $(\sigma(Y^n), P)$-Markov process with infinite dimensional state space $\mathcal{P}(S)$ and transition probability kernel Π_n defined by

$$\Pi_n F(\mu) = \int F(\rho_n(\mu, y))\;\; g_n(y - h(z))\, (\mu\, K)(dz)\, dy$$

for any bounded continuous function $F : \mathcal{P}(S) \longrightarrow \mathbb{R}$ and $\mu \in \mathcal{P}(S)$. In other words with some obvious abusive notations

$$\begin{aligned} dp(y_n, x_n, x_{n-1}/\pi_{n-1}) &= g_n(y_n - h(x_n))\, dy_n\, \pi_{n-1}(dx_{n-1})\, K(x_{n-1}, dx_n) & (7)\\ p(y_n/\pi_{n-1}) &= \int g_n(y_n - h(x_n))\, (\pi_{n-1}\, K)(dx_n) & (8) \end{aligned}$$

Now, the construction of the recursive expression for the conditional distribution of (X_n, X_{n+1}) given the observations Y^n is a fairly immediate consequence of (4) and (5). Using the above notations one gets easily

$$m_{n+1} = \Phi_n(m_n, Y_n) \quad n \geq 1 \qquad m_0 = \nu \times K \tag{9}$$

where

$$\Phi_n(m, y)f = \int \frac{\displaystyle\int f(x_1, x_2)\, g_n(y - h(x_1))\, m(dx_0, dx_1)}{\displaystyle\int g_n(y - h(x_1))\, m(dx_0, dx_1)}\, K(x_1, dx_2) \tag{10}$$

for all $f \in \mathcal{C}(S^2)$, $\mu \in \mathcal{P}(S)$ and $y \in \mathbb{R}$.

2 Interacting Particle Systems

2.1 Description of the Algorithm

The particle system under study will be a Markov chain with state space S^N, where $N \geq 1$ is the size of the system. The N-uple of elements of S, i.e. the points of the set S^N are called systems of particles and will be mostly denoted by the letters x, y, z. Given the observations $Y = y$, we note $(\widehat{x}_n, x_{n+1})_{n\geq 0}$ the $\mathcal{P}(S^2)$-Markov process defined by the transition probabilities

$$\tilde{P}_{[y]}((\widehat{x}_0, x_1) \in d(z_0, z_1)) = \prod_{p=1}^{N} m_0(dz_0^p, dz_1^p) \tag{11}$$

$$\tilde{P}_{[y]}((\widehat{x}_n, x_{n+1}) \in d(z_0, z_1)/(\widehat{x}_{n-1}, x_n) = (x_0, x_1)) = \prod_{p=1}^{N} \Phi_n(\frac{1}{N}\sum_{i=1}^{N} \delta_{(x_0^i, x_1^i)}, y_n)(d(z_0^p, z_1^p)) \tag{12}$$

By the very definition of m_0 and Φ_n we also have the following

1. **Initial Particle System** $\tilde{P}_{[y]}\,(\widehat{x}_0 = dx) = \prod_{p=1}^{N} \nu(dx^p)$

2. **Sampling/Exploration** $\tilde{P}_{[y]}\,(x_n = dx/\widehat{x}_{n-1} = z) = \prod_{p=1}^{N} K(z^p, dx^p)$

3. **Resampling/Selection** $\tilde{P}_{[y]}\,(\widehat{x}_n = dx/x_n = z) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y_n - h(z^i))}{\sum_{j=1}^{N} g_n(y_n - h(z^j))}\,\delta_{z^i}(dx^p)$

Finally one gets a sequence of particles systems

$$\widehat{x}_{n-1} = (\widehat{x}_{n-1}^1, \ldots, \widehat{x}_{n-1}^N) \longrightarrow x_n = (x_n^1, \ldots, x_n^N) \longrightarrow \widehat{x}_n = (\widehat{x}_n^1, \ldots, \widehat{x}_n^N)$$

It is essential to remark that the particles $\widehat{x}_n^i$ are chosen randomly and independently in the population $\{x_n^1, \ldots, x_n^N\}$ from the law π_n^N defined by the likelihood functions and the present measurement Y_n. Namely

$$\pi_n^N = \sum_{i=1}^{N} \frac{g_n(y_n - h(x_n^i))}{\sum_{j=1}^{N} g_n(y_n - h(x_n^j))}\,\delta_{x_n^i} \tag{13}$$

Then it moves to x_{n+1}^i according the transition probability kernel K. In other words the S^N-valued Markov chain $\widehat{x}_n = (\widehat{x}_n^1, \ldots, \widehat{x}_n^N)$ is obtained through overlapping another chain $\widehat{x}_n = (x_n^1, \ldots, x_n^N)$ which represents the successive particles obtained by exploring the probability space with the transitions K. More precisely the motion of particles is decomposed in two stages:

$$\widehat{x}_n \xrightarrow{\textit{Sampling/Exploration}} x_n \xrightarrow{\textit{Resampling/Selection}} \widehat{x}_{n+1} \tag{14}$$

The algorithm constructed in this way will be called an interacting particle filter. The terminology **interacting** is intended to emphasize that the particles are not independent and differs from the particle resolutions introduced in [7] and [6].

Remark 1 *The formulated algorithm permits generalization in the case when the selecting procedure is used from time to time. In practical situations a simple way to do it consists in introducing a resampling schedule. For instance we may choose to resample the particles when fifty percents of the weights are lower than* $\frac{A}{N^p}$ *with a convenient choice of* $A > 0$ *and* $p \geq 2$.

To point out the connection with genetic algorithms and so emphasize the role of the likelihood functions g_n, assume further:

1. The state space S is finite.
2. The Markov transition kernels $K_l(x,z)$ are governed by a parameter l with $K_l(x,z) \to 1_x(z)$ as l grows toward infinity.
3. A noise observation V_n also governed by a parameter l with distribution

$$dP_n^V(v) = \frac{\exp\left(-V(v)\log l\right)\,dv}{\int \exp\left(-V(u)\log l\right)\,du}$$

4. Homogeneous series of observations with respect to time

$$\forall n \geq 1 \quad Y_n = y \qquad\qquad V(y - h(x)) \stackrel{def}{=} f(x)$$

The corresponding Exploration and Selection mechanisms are governed by a parameter l and take the following form:

$$\begin{aligned} Pr\left(x_n^{(l)} = dx / \widehat{x}_n^{(l)} = z\right) &= \prod_{p=1}^{N} K_l(z^p, x^p) \\ Pr\left(\widehat{x}_n^{(l)} = x / x_{n-1}^{(l)} = z\right) &= \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{l^{-f(z^i)}}{\sum_{j=1}^{N} l^{-f(z^j)}}\, 1_{z^i}(x^p) \end{aligned}$$

In this very special situation Cerf [2] gives several conditions on the rate of decrease of the perturbations $1/\log l(n)$ to ensure all particles $\widehat{x}_n^{(l(n)),i}$ visit the set of global maxima of the fitness function f in finite time when the number of particles N is greater than a critical value.
Unfortunately there is a critical lack of theoretical results for the convergence of such algorithms for the numerical solving of the nonlinear filtering equations.

The crucial question is of course whether the empirical random measure $N^{-1}\sum_{i=1}^{N} \delta_{\widehat{x}_n^i}$ converges to the conditional distribution π_n when the size of the particle system is growing. This is answered in the positive in theorem 2 section 5. We will show that for every bounded Borel test function $f : S \to \mathbb{R}$ and $n \geq 1$

$$\lim_{N \to +\infty} E\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n(f)\right|\right) = 0 \tag{15}$$

We give a new and detailed analysis of this problem. These results are largely recent although the question of the local convergence occurs in [5]. Unlike genetic algorithms it should be underline that we are not necessarily trying to recover the unknown state variables exactly. The conditional distribution gives the conditional minimum variance estimate but the error does not in general converge to zero as the time tends to infinity (see Kunita [12]).

Such particle nonlinear filters differs from those introduced in [7] and [6]. It clearly encompasses the genetic algorithms introduced by Holland [11] and recently developed by R. Cerf [2]. An advantage of this procedure is that it simultaneously explore the probability space according the *a priori* Markov kernel and update the information that is accumulated through the resampling of the positions. In introducing such adaptation/selection laws the particle system acquires certain configurations as to represent an estimate of the conditional distribution. They provide a natural

procedure for a system of particles to sense its environment according their likelihood functions and the observations.
It is clear that such particle resolutions could be naturally formulated in other context: neural networks, model parameter identification, optimal control ([9]).

2.2 The Associate Markov Process

Now we shall proceed to model such interacting particle procedure and the nonlinear filtering problem in a natural stochastic basis. In the preceding paragraph we have remarked on the fact that the particles coincide with the support of random measures which estimate the conditional distribution. A convenient tool for the analysis of the modelling of such interacting particle systems is the splitting transition kernel:

$$\mathcal{P}(S^2) \xrightarrow{C_{\mathrm{N}}} \left\{ \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i} \ : \ x^i \in S^2 \right\} \subset \mathcal{P}(S^2) \tag{16}$$

defined for every $F \in \mathcal{C}(\mathcal{P}(S^2)), \eta \in \mathcal{P}(S^2)$ and $N \geq 1$ by

$$C_{\mathrm{N}}F(\eta) = \int F(m)\, C_{\mathrm{N}}(\eta, dm) \overset{def}{=} \int_{S^{2\mathrm{N}}} F(\frac{1}{N}\sum_{i=1}^{N} \delta_{x^i})\ \eta(dx^1)\ldots\eta(dx^{\mathrm{N}}) \tag{17}$$

The interpretation of C_N is this: by starting with a measure $\eta \in \mathcal{P}(S^2)$, the next measure is the result of the sampling N independent random variables x^i with common law η:

$$\eta \xrightarrow{C_{\mathrm{N}}} m = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}$$

Given a series of observations $Y = y$ we observe that the random empirical measures

$$m_{n+1}^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{(\widehat{x}_n^i, x_{n+1}^i)}$$

are the result of sampling N independent random variables with common law

$$\Phi_n(m_n^N, y_n) = \pi_n^N \times K = \sum_{i=1}^{N} \frac{g_n(y_n - h(x_n^i))}{\sum_{j=1}^{N} g_n(y_n - h(x_n^j))}\ \delta_{x_n^i} \ \times\ K$$

The above observations enables us to consider the empirical measures m_n^N as a canonical $\mathcal{P}(S^2)$-valued Markov process $(\Omega', \beta_n, \tilde{P}_{[y]})$ by setting

$$\tilde{P}_{[y]}\left(m_1^N \in d\eta\right) = C_N\left(\Phi_n(m_0, y_n), d\eta\right) \qquad \tilde{P}_{[y]}\left(m_{n+1}^N \in d\eta / m_n^N = \mu\right) = C_N\left(\Phi_n(\mu, y_n), d\eta\right) \tag{18}$$

Now we design a stochastic basis for the convergence of our particle approximations. To capture all randomness we list all outcomes into the canonical space defined as follows:

1. Recall (Ω, F_n, P) is the canonical space for the signal observation pair (X, Y).
2. We define $\tilde{\Omega} = \Omega' \times \Omega$ and $\tilde{F}_n = \beta_n \times F_n$ and, for every $\tilde{\omega} \overset{def}{=} (\omega^1, \omega^2, \omega^3) \in \tilde{\Omega}$ we define:
$$m_n^N(\tilde{\omega}) = \omega_n^1 \quad X_n(\tilde{\omega}) = \omega_n^2 \quad Y_n(\tilde{\omega}) = \omega_n^3$$
3. For every $A \in \beta_n$ and $B \in F_n$ we define $\tilde{P}$ as follows:
$$\tilde{P}(A \times B) \overset{def}{=} \int_B \tilde{P}_{[Y(\omega)]}(A)\ dP(\omega) \tag{19}$$

As usual we use $\tilde{E}(\cdot)$ to denote expectations with respect to $\tilde{P}$ and $\tilde{E}_{[y]}(\cdot)$ to denote expectations with respect to $\tilde{P}_{[y]}$. With some obvious abusive notations we have

$$d\tilde{p}\left(m_1^N, \ldots, m_n^N, y_1, \ldots, y_n\right) = \prod_{k=1}^{n} C_N\left(\Phi_{k-1}\left(m_{k-1}^N, y_{k-1}\right), dm_k^N\right) \; dp\left(y_1, \ldots, y_n\right)$$

with the convention $\Phi_0\left(m_0^N, y_0\right) = \nu \times K$

3 General Recursive Formulas

We shall adopt in this section an unconventional model for the conditional distributions. The choice is dictated by our desire to have very simple relationships between the likelihood functions and the conditional distributions. Initially, this will require a setup quite different from that used in sections 1 and 2 but in the end of our investigations will resemble more and more the models presented in section 1. The setting is the same as in section 2.1. In particular, unless otherwise stated we assume the observation data is a fixed series of real numbers $Y = y$.
This assumption enables us to consider the conditional distribution π_n as a probability parameterized by the observation parameters $y_1, \ldots, y_n, \ldots$. Therefore, when the context is unambiguous we will often write for brevity $g_n(x)$ instead of $g_n(y_n - h(x_n))$. Using this notation an alternative notation for (5) is the recursive formula

$$\pi_n f = \frac{\pi_{n-1} K(\, f \; g_n)}{\pi_{n-1} K(g_n)} \qquad \forall f \in \mathcal{C}(S)$$

3.1 Likelihood functions analysis

The proof of the convergence (15) involves

1. the Maximum Log-likelihood functions given by

$$V_{n/p}(y) = \sup_{x_p \in S} \log \int_S \left(\frac{p(y_n, \ldots, y_{p+1}/x_{p+1})}{p(y_n, \ldots, y_{p+1}/y^p)}\right)^2 dp(x_{p+1}/x_p) \qquad p+1 \le n \qquad (20)$$

$$V_n(y) = \sum_{p=1}^{n} V_{p/p-1}(y) \qquad (21)$$

   where

   - $p(y_n, \ldots, y_{p+1}/y^p)$ denotes the density under P of the distribution of $(Y_n, \ldots, Y_{p+1})$ conditionally to $Y^p = (Y_1, \ldots, Y_p)$
   - $p(y_n, \ldots, y_{p+1}/x_{p+1})$ denotes the density under P of the distribution of $(Y_n, \ldots, Y_{p+1})$ conditionally to X_{p+1}
   - $dp(x_{p+1}/x_p) = K(x_p, dx_{p+1})$.

2. The conditional expectations given by

$$f_n^{(p)}(x_p) = E\left(f(X_n)/X_p, Y_{p+1}, \ldots, Y_n\right)(x_p, y_{p+1}, \ldots, y_n) \qquad 0 \le p \le n-1 \qquad (22)$$

The functions $V_{n/p}$ represents Log-likelihood functions on the observation process Y and they are closely related to the resampling/selection mechanism of our algorithm. The relationship is due to the fact that the selection mechanism is formulated in terms of the fitness functions and

$$p(y_n/x_n) = g_n(y_n - h(x_n)) \qquad (23)$$

$$p(y_n, \ldots, y_p/x_p) = \int_S p(y_n, \ldots, y_{p+1}/x_{p+1}) \; dp(x_{p+1}/x_p) \qquad (24)$$

Example 1 1. *Assume here that state and observation processes* (X, Y) *are given by the linear dynamics*

$$\begin{aligned} X_n &= A\, X_{n-1} + W_n && (25)\\ Y_n &= C\, X_n + V_n && (26) \end{aligned}$$

with $X_n \in \mathbb{R}$, $Y_n \in \mathbb{R}$, $Y_0 = 0$, A *and* C *are real numbers,* X_0, W_n *and* V_n *are normally distributed with means* 0 *and respective non negative covariance* Q_0, Q *and* R. *The conditional densities of* Y_n *given* X_n *and* $Y^{n-1} = (Y_1, \ldots, Y_{n-1})$ *are given by*

$$\begin{aligned} p(y_n/x_n) &= g_n(y_n - C\,x_n) = \frac{1}{\sqrt{2\pi|R|}} \exp\left(-\frac{1}{2}(y_n - C\,x_n)^2 R^{-1}\right)\\ p(y_n/y^{n-1}) &= = \frac{1}{\sqrt{2\pi|C^2\,P_{n/n-1} + R|}} \exp\left(-\frac{1}{2}(y_n - C\,A\,\widehat{X}_{n-1})^2 \left(C^2\,P_{n/n-1} + R\right)^{-1}\right) \end{aligned}$$

with the well known measurement update equations

$$\begin{aligned} \widehat{X}_n &\stackrel{def}{=} E(X_n/Y^n) = A\,\widehat{X}_{n-1} + K_n\,\left(Y_n - C\,A\,\widehat{X}_{n-1}\right)\\ K_n &\stackrel{def}{=} C\,P_{n/n-1}\,\left(C^2\,P_{n/n-1} + R\right)^{-1}\\ P_{n/n-1} &\stackrel{def}{=} E\left(\left(X_n - E(X_n/Y^{n-1})\right)^2\right) = A^2\,P_{n-1/n-1} + Q\\ P_{n-1/n-1} &\stackrel{def}{=} E\left(\left(X_{n-1} - E(X_{n-1}/Y^{n-1})\right)^2\right) = \left(P^{-1}_{n-1/n-2} + C\,R^{-1}\,C\right)^{-1} \end{aligned}$$

In this very special situation

$$V_{n/n-1}(Y) \leq \log\left[\frac{|C^2\,P_{n/n-1} + R|}{|R|} \exp\left((Y_n - C\,A\,\widehat{X}_{n-1})^2 \left(C^2\,P_{n/n-1} + R\right)^{-1}\right)\right]$$

Further manipulations yields

$$V_{n/p}(Y) \leq \log\left[\prod_{k=p}^{n} \frac{|C^2\,P_{k/k-1} + R|}{|C^2\,S_{k-p} + R|} \exp\left((Y_k - C\,A\,\widehat{X}_{k-1})^2 \left(C^2\,P_{k/k-1} + R\right)^{-1}\right)\right]$$

with $S_0 = 0$ *and for* $A \neq 1$, $E\left((X_k - E\,(X_k/X_p))^2\right) = S_{k-p} = \dfrac{1 - A^{k-p}}{1 - A}\,Q$

2. *Let* X *be a discrete time Markov process belonging to a finite discrete set* S *and assume that the observation noise* V_n *form a sequence of independent, random variables with*

$$dp^V_n(v) = \frac{\exp\left(-U_n\right)}{\displaystyle\int \exp\left(-U_n(v)\right)\,dv} \qquad U_n : S \to \mathbb{R}^+$$

In this hidden Markov model $V_{n/n-1}(Y) \leq 2\,\sup_{x\in S} U_n\,(Y_n - h(x))$

The functions (20), (22) and (24) will be used in our analysis of the convergence (15). This analysis which requires specific developments because of the difficulty to estimates mean errors. For instance:

$$\begin{aligned} \tilde{E}_{[y]}\left(\left(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}^i_n)\right)^2\right) &= \tilde{E}_{[y]}\left(\left(\underline{m}^N_{n+1} f\right)^2\right)\\ &= \frac{1}{N}\tilde{E}_{[y]}\left(\pi^N_n(f^2)\right) + (1 - \frac{1}{N})\tilde{E}_{[y]}\left(\left(\pi^N_n f\right)^2\right)\\ &= \frac{1}{N}\tilde{E}_{[y]}\left(\pi^N_n\left(f - \pi^N_n f\right)^2\right) + \tilde{E}_{[y]}\left(\left(\pi^N_n f\right)^2\right) \end{aligned}$$

Thus, Tchebitchev's inequality and convenient estimates of

$$\tilde{E}_{[y]}\left(\left(\pi_n^N f\right)^2\right) = \tilde{E}_{[y]}\left(\left(\sum_{i=1}^{N} \frac{g_n(y_n - h(x_n^i))}{\sum_{j=1}^{N} g_n(y_n - h(x_n^j))}\, f(x_n^i)\right)^2\right) \tag{27}$$

would give a complete answer to the convergence in $\mathbf{L}^0(\tilde{P}_{[y]})$. Unfortunately it is difficult to estimate (27) or even $\tilde{E}_{[y]}\left(\pi_n^N f\right)$.

Accordingly, to point out the connections between (22) and (24) and so emphasize the role of the Log-likelihood functions $V_{n/p}$ we first introduce a natural framework in which the relationship between (22) and (24) is made explicit. The recursive expressions described in the next section will be used repeatedly in the last part of the paper for the convergence (15).

3.2 Recursive Formulas

The chief purpose of this section is to introduce recursive expressions for (22) and (24). The technical approach presented here is to work with a the same given sequence of observations $Y = y$. The conditional distributions π_n and the conditional expectations $(f_n^{(p)})_{1\leq p\leq n}$ will be formulated as a probability and a sequence of functions parameterized by the observation parameters y. The recursive formulas described in this section will show how the sequence of observations scales the update of both π_n and $f_n^{(p)}$. Our constructions will be explicit and the recursions will have a simple form. First, let us give some details about the use of Bayes' formula in our setting. With some abusive but obvious notations we have the recursions for all $1 \leq p \leq n$

$$\begin{aligned} p(y_n,\ldots,y_p/x_p) &= p(y_p/x_p)p(y_n,\ldots,y_{p+1}/x_p) \\ &= p(y_p/x_p)\int p(y_n,\ldots,y_{p+1}/x_{p+1})\; dp(x_{p+1}/x_p) \end{aligned} \tag{28}$$

$$\frac{p(y_n,\ldots,y_p/x_p)}{p(y_n,\ldots,y_p/y^{p-1})} = \frac{p(y_p/x_p)}{p(y_p/y^{p-1})}\int \frac{p(y_n,\ldots,y_{p+1}/x_{p+1})}{p(y_{p+1}/y^p)}\; dp(x_{p+1}/x_p) \tag{29}$$

$$p(y_n,\ldots,y_1) = p(y_n/y^{n-1})\; p(y_{n-1}/y^{n-2})\ldots p(y_2/y_1)\; p(y_1) \tag{30}$$

To clarify the presentations we introduce the following definitions, for all $0 \leq p \leq n$

1. $g_n^{(p)}(x_p) = p(y_n,\ldots,y_p/x_p)$
2. $g_{n/p-1}(x_p) = p(y_n,\ldots,y_p/x_p)/p(y_n,\ldots,y_p/y^{p-1})$

We will slightly abuse notations and we will often write $g_n(x_n)$ instead of $g_n^{(n)}(x_n) = p(y_n/x_n)$. It is now easily checked from the above remarks that

$$g_{n/p-1}(x_p) = g_n^{(p)}(x_p)/\int \pi_p(dz_{p-1})\; K(z_{p-1},dz_p)\; g_n^{(p)}(z_p) \tag{31}$$

$$g_n^{(p)}(x_p) = g_p(x_p)\int K(x_p,dx_{p+1})\; g_n^{(p+1)}(x_{p+1}) \tag{32}$$

$$g_{n/p-1}(x_p) = g_{p/p-1}(x_p)\int K(x_p,dx_{p+1})\; g_{n/p}(x_{p+1}) \tag{33}$$

In summary then, we have the backward recursive formulas.

Proposition 1 *For every* $0 \leq p < n-1$

$$g_n^{(p)} = g_p\;(K\,g_n^{(p+1)}) \qquad g_{n/p-1} = \frac{g_n^{(p)}}{\pi_{p-1}Kg_n^{(p)}} \qquad g_{n/p-1} = g_{p/p-1}\;K(g_{n/p}) \tag{34}$$

We will adopt the conventions $\pi_{-1}K = \nu$ *and* $g_0^{(0)} = 1$.

Moreover, by the very definition of π_n and π_n^N, one gets

$$\pi_n(f) = \frac{\int f(z)\, g_n(y_n - h(z))\, \overline{m}_n(dz)}{\int g_n(y_n - h(z))\, \overline{m}_n(dz)} \qquad \pi_n^N(f) = \sum_{i=1}^{N} \frac{g_n(y_n - h(x_n^i))}{\sum_{j=1}^{N} g_n(Y_n - h(x_n^j))}\, f(x_n^i) \tag{35}$$

For brevity we will write $\pi_n(f) = \dfrac{\overline{m}_n\,(g_n\, f)}{\overline{m}_n\,(g_n)}$ $\qquad \pi_n^N(f) = \dfrac{\overline{m}_n^N\,(g_n\, f)}{\overline{m}_n^N\,(g_n)}$

Continuing in the same vein we derive the conditional expectations $f_n^{(p)}$ introduced in (22). Using Bayes' rule one has the following basic equation for all $1 \leq p \leq n$

$$\begin{aligned} dp(x_n, x_p/x_{p-1}, y_p, \ldots, y_n) &= dp(x_n/x_p, y_{p+1}, \ldots, y_n)\; dp(x_p/x_{p-1}, y_p, \ldots, y_n) \\ &= dp(x_n/x_p, y_{p+1}, \ldots, y_n)\; \frac{p(y_n, \ldots, y_p/x_p)}{p(y_n, \ldots, y_p/x_{p-1})} dp(x_p/x_{p-1}) \end{aligned}$$

By the same line of arguments, for all $1 \leq p \leq n$, one gets

$$\begin{aligned} dp(x_n, x_p, x_{p-1}/y^n) &= dp(x_n/x_p, y_{p+1}, \ldots, y_n)\; dp(x_p, x_{p-1}/y^n) \\ &= dp(x_n/x_p, y_{p+1}, \ldots, y_n)\; \frac{p(y_n, \ldots, y_p/x_p)}{p(y_n, \ldots, y_p/y^{p-1})} dp(x_p/x_{p-1}) dp(x_{p-1}/y^{p-1}) \end{aligned}$$

Thus we arrive at

$$\pi_n f = \frac{\int \pi_{p-1}(dx_{p-1})\; K(x_{p-1}, dx_p)\; f_n^{(p)}(x_p)\; g_n^{(p)}(x_p)}{\int \pi_{p-1}(dx_{p-1})\; K(x_{p-1}, dx_p)\; g_n^{(p)}(x_p)} \qquad \forall f \in \mathcal{C}(S) \quad 1 \leq p \leq n \tag{36}$$

As well, one has

$$f_n^{(p-1)}(x_{p-1}) = \frac{\int K(x_{p-1}, dx_p)\; f_n^{(p)}(x_p)\; g_n^{(p)}(x_p)}{\int K(x_{p-1}, dx_p)\; g_n^{(p)}(x_p)} \qquad \forall f \in \mathcal{C}(S) \quad 1 \leq p \leq n \tag{37}$$

In summary then, our conditional expectations $f_n^{(p)}$ can be described as follows

Proposition 2 *For every $f \in \mathcal{C}(S)$, the conditional expectations $(f_n^{(p)})_{1 \leq p \leq n}$ satisfy the recursive formula For every $f \in \mathcal{C}(S)$ and $1 \leq p \leq n$*

$$f_n^{(p-1)} \stackrel{def}{=} \frac{K(f_n^{(p)}\; g_{n/p-1})}{K(g_{n/p-1})} = \frac{K(f_n^{(p)}\; g_n^{(p)})}{K(g_n^{(p)})} \qquad \forall 1 \leq p \leq n \tag{38}$$

Moreover for every $f \in \mathcal{C}(S)$ and $1 \leq p \leq n$

$$\pi_n f = \frac{(\pi_{p-1}\, K)\left(f_n^{(p)}\; g_n^{(p)}\right)}{(\pi_{p-1}\, K)\left(g_n^{(p)}\right)} = \frac{(\pi_{p-1}\, K)\left(f_n^{(p)}\; g_{n/p-1}\right)}{(\pi_{p-1}\, K)\left(g_{n/p-1}\right)} \tag{39}$$

We will adopt the conventions $f_n^{(-1)} \stackrel{def}{=} \nu(f_n^{(0)}\; g_n^{(0)})/\nu(g_n^{(0)})$

4 Mean Square Estimates

In this section we analyze the structure of the Log-likelihood functions $V_{n/p}$ pointing out explicit bounds. Our next objective is to estimate convergence rate and mean errors. We shall do this now, beginning with some lemmas which will be use repeatedly in this section.

Lemma 1 *Let $f : S \longrightarrow \mathbb{R}$ be any integrable Borel test function, $n \geq 0$ and $N \geq 1$.*
We have $\tilde{P}$-a.e.

$$\tilde{E}_{[Y]}\left(\underline{m}_{n+1}^{N} f/\beta_n\right) = \pi_n^N f \qquad \tilde{E}_{[Y]}\left(\overline{m}_{n+1}^{N} f/\beta_n\right) = \pi_n^N K f \tag{40}$$

$$\tilde{E}_{[Y]}\left((\underline{m}_{n+1}^{N} f)^2/\beta_n\right) = \frac{1}{N}\pi_n^N(f^2) + (1-\frac{1}{N})(\pi_n^N f)^2 \tag{41}$$

$$\tilde{E}_{[Y]}\left((\overline{m}_{n+1}^{N} f)^2/\beta_n\right) = \frac{1}{N}\pi_n^N K(f^2) + (1-\frac{1}{N})(\pi_n^N K f)^2 \tag{42}$$

proof:
It suffices to note that

$$\begin{aligned}
\tilde{E}_{[Y]}\left(\underline{m}_{n+1}^{N} f/\beta_n\right) &= \int \underline{m}(f)\ C_N(\pi_n^N \times K, dm) = \pi_n^N f \\
\tilde{E}_{[Y]}\left(\underline{m}_{n+1}^{N} (f)^2/\beta_n\right) &= \int (\underline{m} f)^2\ C_N(\pi_n^N \times K, dm) = \frac{1}{N}\ \pi_n^N(f^2) + \left(1-\frac{1}{N}\right)\ (\pi_n^N f)^2 \\
\tilde{E}_{[Y]}\left(\overline{m}_{n+1}^{N} f/\beta_n\right) &= \int \overline{m}(f)\ C_N(\pi_n^N \times K, dm) = \pi_n^N K f \\
\tilde{E}_{[Y]}\left(\overline{m}_{n+1}^{N} (f)^2/\beta_n\right) &= \int (\overline{m} f)^2\ C_N(\pi_n^N \times K, dm) = \frac{1}{N}\ \pi_n^N K(f^2) + \left(1-\frac{1}{N}\right)\ (\pi_n^N K f)^2
\end{aligned}$$

■

Now we derive a technical recursion formula in p for the expressions

$$\tilde{E}_{[Y]}\left(\overline{m}_{p+1}^{N}(g_{n/p}\ f_n^{(p+1)})/\beta_p\right) \qquad 0 \leq p \leq n-1$$

Lemma 2 *Let $f : S \longrightarrow \mathbb{R}$ be any integrable Borel test function, $n \geq 0$ and $N \geq 1$. For every $0 \leq p \leq n-1$ we have the recursion*

$$\tilde{E}_{[Y]}\left(\overline{m}_{p+1}^{N}(g_{n/p}\ f_n^{(p+1)})/\beta_p\right) = \pi_p^N K\left(g_{n/p}\ f_n^{(p+1)}\right) = \frac{\overline{m}_p^N(g_{n/p-1}\ f_n^{(p)})}{\overline{m}_p^N(g_{p/p-1})} \qquad \tilde{P}-a.e. \tag{43}$$

proof:
Using the inductive definitions of $f_n^{(p)}$, $g_{n/p}$ and lemma 1 we have

$$\begin{aligned}
\tilde{E}_{[Y]}\left(\overline{m}_{p+1}^{N}\left(g_{n/p} f_n^{(p+1)}\right)/\beta_p\right) &= \pi_p^N K(g_{n/p} f_n^{(p+1)}) \\
&= \frac{\overline{m}_p^N\left(g_{p/p-1} K(g_{n/p} f_n^{(p+1)})\right)}{\overline{m}_p^N\left(g_{p/p-1}\right)} = \frac{\overline{m}_p^N\left(g_{n/p-1}\ f_n^{(p)}\right)}{\overline{m}_p^N\left(g_{p/p-1}\right)}
\end{aligned}$$

■

4.1 Martingale Approach

To estimate mean square errors the key idea is to introduce a $(\tilde{P}, \tilde{F}_n)$-martingale U^N using the functions $g_{n/n-1}$ and the random measures $\overline{m}_n^N$. More precisely we define U_n^N as follows

Definition 1 *We note U^N the stochastic process defined by*

$$U_0^N = 1 \qquad U_n^N = \overline{m}_n^N(g_{n/n-1})\, U_{n-1}^N \qquad \forall n \geq 1 \tag{44}$$

In other words $\quad U_n^N = \prod_{k=1}^{n} \left(\frac{1}{N} \sum_{i=1}^{N} g_{k/k-1}(x_k^i) \right).$

Lemma 3 *U_n^N is a $(\tilde{P}, \tilde{F}_n)$-martingale with $\tilde{E}(U_n^N) = 1$ and $\tilde{P} - a.s.$ $\tilde{E}_{[Y]}(U_n^N) = 1$.*

proof:
The first statement follows on recalling that

$$g_{n/n-1}(x) = \frac{g_n(x)}{\pi_{n-1} K g_n} = \frac{g_n(Y_n - h(x))}{p(Y_n/Y^{n-1})}$$

This gives $\tilde{P} - a.s.$

$$\begin{aligned} \tilde{E}(U_n^N/\tilde{F}_{n-1}) &= U_{n-1}^N \int \overline{m}(g_{n/n-1})\; C_N\left(\pi_{n-1}^N \times K, dm\right)\, dp(y_n/Y^{n-1}) \\ &= U_{n-1}^N \int g_n(y_n - h(z))\, (\pi_{n-1}^N K)(dz)\, dy_n = U_{n-1}^N \end{aligned}$$

and the first assertion follows. To prove $\tilde{E}_{[Y]}(U_n^N) = 1$, the above discussion go through with minor changes. Indeed $\tilde{P} - a.s.$

$$\tilde{E}_{[Y]}(U_n^N) = \tilde{E}_{[Y]}\left(\overline{m}_n^N\left(g_{n/n-1}\right)\, U_{n-1}^N\right) = \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\overline{m}_n^N(g_{n/n-1})\,/\beta_{n-1}\right) U_{n-1}^N\right) \tag{45}$$

Now, using the inductive definition of $g_{n/p}$ we obtain

$$\tilde{E}_{[Y]}\left(\overline{m}_n^N g_{n/n-1}\,/\beta_{n-1}\right) = \overline{m}_{n-1}^N(g_{n/n-2})/\overline{m}_{n-1}^N(g_{n-1/n-2})$$

Then (45) $\Longrightarrow \tilde{E}_{[Y]}\left(U_n^N\right) = \tilde{E}_{[Y]}\left(\overline{m}_{n-1}^N(g_{n/n-2})\, U_{n-2}^N\right) = \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\overline{m}_{n-1}^N(g_{n/n-2})\,/\beta_{n-2}\right) U_{n-2}^N\right)$

This procedure can be repeated. using the recursive formulas described in section 3.2 we note that

$$\begin{aligned} \tilde{E}_{[Y]}\left(U_n^N\right) &= \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\overline{m}_p^N(g_{n/p-1})\,/\beta_{p-1}\right) U_{p-1}^N\right) \\ \tilde{E}_{[Y]}\left(\overline{m}_p^N(g_{n/p-1})\,/\beta_{p-1}\right) &= \overline{m}_{p-1}^N(g_{n/p-2})/\overline{m}_{p-1}^N(g_{p-1/p-2}) \\ \Longrightarrow \tilde{E}_{[Y]}\left(U_n^N\right) &= \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\overline{m}_{p-1}^N(g_{n/p-2})\,/\beta_{p-2}\right) U_{p-2}^N\right) \end{aligned}$$

Using backward induction in p the result follows from the fact that

$$\tilde{E}_{[Y]}\left(\overline{m}_1^N(g_{n/0})\right) = \nu K(g_{n/0}) = 1$$

■

The analysis of U^N is a powerful tool to study the convergence rate of our interacting particle filter. That is, introducing the process U^N one makes possible the calculation of mean errors, then these estimates will be reinterpreted back. As a typical example we have the following lemma

Lemma 4 *For any integrable Borel test function $f : S \longrightarrow \mathbb{R}$ we have*

$$\tilde{E}_{[Y]}\left(U_n^N\ \underline{m}_{n+1}^N\,(f-\pi_n f)\right)=0 \qquad \tilde{P}-a.e. \tag{46}$$

proof:
Using lemma 3 it is equivalent to prove

$$\tilde{E}_{[Y]}\left(U_n^N\ \underline{m}_{n+1}^N f\right)=\pi_n f$$

Now, using lemma 1 and the fact that $f_n^{(n)}=f$ note that

$$\begin{aligned}
\tilde{E}_{[Y]}\left(U_n^N\ \underline{m}_{n+1}^N f\right) &= \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\underline{m}_{n+1}^N f/\beta_n\right)U_n^N\right)\\
\tilde{E}_{[Y]}\left(\underline{m}_{n+1}^N f/\beta_n\right) &= \pi_n^N f=\overline{m}_n^N(g_{n/n-1}\ f)/\overline{m}_n^N(g_{n/n-1})\\
\Longrightarrow \tilde{E}_{[Y]}\left(U_n^N\ \underline{m}_{n+1}^N f\right) &= \tilde{E}_{[Y]}\left(U_{n-1}^N\ \overline{m}_n^N(g_{n/n-1}f_n^{(n)})\right)
\end{aligned}$$

Now, using backward induction in $1\le p\le n-1$ and the recursive formulas described in section 3.2

$$\begin{aligned}
\tilde{E}_{[Y]}\left(U_p^N\ \overline{m}_{p+1}^N(g_{n/p}\ f_n^{(p+1)})\right) &= \tilde{E}_{[Y]}\left(U_p^N\ \tilde{E}_{[y]}\left(\overline{m}_{p+1}^N(g_{n/p}\ f_n^{(p+1)})/\beta_p\right)\right)\\
\tilde{E}_{[Y]}\left(\overline{m}_{p+1}^N(g_{n/p}\ f_n^{(p+1)})/\beta_p\right) &= \overline{m}_p^N(g_{n/p-1}\ f_n^{(p)})/\overline{m}_{n-1}^N(g_{p/p-1})\\
\Longrightarrow \tilde{E}_{[Y]}\left(U_n^N\ \underline{m}_{n+1}^N f\right) &= \tilde{E}_{[Y]}\left(U_{p-1}^N\ \overline{m}_p^N(g_{n/p-1}\ f_n^{(p)})\right)
\end{aligned}$$

Finally the result follows from the recursive formulas described in section 3.2 and the fact that

$$\tilde{E}_{[Y]}\left(\overline{m}_1^N(g_{n/0}\ f_n^{(1)})\right)=\nu K(g_{n/0}\ f_n^{(1)})=\frac{\nu K(g_{n/0}\ f_n^{(1)})}{\nu K(g_{n/0})}=\pi_n f$$

■

4.2 Mean Square Estimates

To prove the convergence (15) it clearly suffices suffices to prove that

$$\lim_{N\to+\infty}\tilde{E}_{[Y]}\left(\left(U_n^N-1\right)^2\right)=0=\lim_{N\to+\infty}\tilde{E}_{[Y]}\left(\left(\left(\frac{1}{N}\sum_{i=1}^N f(\hat{x}_n^i)-\pi_n f\right)U_n^N\right)^2\right) \tag{47}$$

The chief purpose of this section is to provide a way for estimate the rate of convergence of (47) in terms of the log-likelihood functions $V_{n/p}$. Such computations are at the heart of the development. They will simplify drastically the evaluation of the convergence rates discussed in the last section. We quote first the following result.

Proposition 3 *For every series of observations $y\in\mathbb{R}^{\mathbb{N}}$,*

1. $V_{n/p}(y)\le V_{n/n-1}(y)+V_{n-1/n-2}(y)+\ldots+V_{p+1/p}(y)\le V_n(y) \qquad \forall 0\le p<n$
2. *For every $0\le k\le n$ and $0\le p_0<\ldots<p_{k-1}<p_k=n$ we have* $\sum_{l=1}^{k} V_{p_l/p_{l-1}}(y)\le V_n(y)$

proof:
The proof is a straightforward computation using the implication

$$g_{n/p} = K(g_{n/p+1})\ g_{p+1/p} \Longrightarrow V_{n/p}(y) = \log \|K(g^2_{n/p})\|_\infty \leq V_{n/p+1}(y) + V_{p+1/p}(y)$$

■

Now we want to develop explicit formulas for expressing the effect of the population size and the observation likelihood functions on the mean square errors (47). The following propositions will be used repeatedly in the last section.

Proposition 4 *For every $N \geq 1$ and $n \geq 0$ we have $\tilde{P}$-a.e.*

1. $\tilde{E}_{[Y]}\left((U^N_n - 1)^2\right) \leq \sum_{k=1}^n \frac{1}{N^k}\left(1 - \frac{1}{N}\right)^{n-k} \sum_{0\leq p_0<\ldots<p_k=n} \exp\left(\sum_{l=1}^k V_{p_l/p_{l-1}}(Y)\right)$
2. $\tilde{E}_{[Y]}\left((U^N_n - 1)^2\right) \leq \left(1 - \left(1 - \frac{1}{N}\right)^n\right)\ e^{V_n(Y)}$

proof:

2) is a clear consequence of proposition 3 and 1). Let us prove 1). Using lemma 3

$$\tilde{E}_{[Y]}\left((U^N_n - 1)^2\right) = \tilde{E}_{[Y]}\left((U^N_n)^2\right) - 1$$

Hence, with the standard convention $\sum_\emptyset = 0$, it is sufficient to prove that

$$\tilde{E}_{[Y]}\left((U^N_n)^2\right) \leq \sum_{k=0}^n \frac{1}{N^k}\left(1 - \frac{1}{N}\right)^{n-k} \sum_{0\leq p_0<\ldots<p_k=n} \exp\left(\sum_{l=1}^k V_{p_l/p_{l-1}}(Y)\right) \tag{48}$$

The proof of (48) is based on backward and forward induction in n and maximization techniques. Using lemma 1 and the recursion in lemma 2 we have

$$\begin{aligned}
\tilde{E}_{[Y]}\left((U^N_n)^2\right) &= \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\overline{m}^N_n(g_{n/n-1})^2/\beta_{n-1}\right)(U^N_{n-1})^2\right) \\
\tilde{E}_{[Y]}\left(\overline{m}^N_n(g_{n/n-1})^2/\beta_{n-1}\right) &\leq \frac{1}{N}\ \pi^N_{n-1}K(g^2_{n/n-1}) + \left(1 - \frac{1}{N}\right)\ (\pi^N_{n-1}Kg_{n/n-1})^2 \\
&\leq \frac{1}{N}\ e^{V_{n/n-1}(Y)} + \left(1 - \frac{1}{N}\right)\ \left(\frac{\overline{m}^N_{n-1}(g_{n/n-2})}{\overline{m}^N_{n-1}(g_{n-1/n-2})}\right)^2
\end{aligned}$$

Thus

$$\tilde{E}_{[Y]}\left((U^N_n)^2\right) \leq \frac{1}{N}\ e^{V_{n/n-1}(Y)}\tilde{E}_{[Y]}\left((U^N_{n-1})^2\right) + \left(1 - \frac{1}{N}\right)\ \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\overline{m}^N_{n-1}(g_{n/n-2})^2/\beta_{n-2}\right)(U^N_{n-2})^2\right)$$

By the same line of arguments for every $1 \leq p \leq n-1$

$$\begin{aligned}
\tilde{E}_{[Y]}\left(\overline{m}^N_{p+1}(g_{n/p})^2/\beta_p\right) &\leq \frac{1}{N}\ e^{V_{n/p}(Y)} + \left(1 - \frac{1}{N}\right)\ \left(\frac{\overline{m}^N_p(g_{n/p-1})}{\overline{m}^N_p(g_{p/p-1})}\right)^2 \\
\tilde{E}_{[Y]}\left(\overline{m}^N_1(g_{n/0})^2/\beta_0\right) &\leq \frac{1}{N}\ e^{V_{n/0}(Y)} + \left(1 - \frac{1}{N}\right)\ \left(\nu K g_{n/0}\right)^2
\end{aligned}$$

Cascading in the above expression

$$\tilde{E}_{[Y]}\left((U^N_n)^2\right) \leq \left(1 - \frac{1}{N}\right)^n + \frac{1}{N}\sum_{q=0}^{n-1}\left(1 - \frac{1}{N}\right)^{n-1-q} e^{V_{n/q}(Y)}\tilde{E}_{[Y]}\left((U^N_q)^2\right)$$

Suppose the inequalities (48) have been proved for every $q \leq n-1$, that is

$$\tilde{E}_{[Y]}\left((U_q^N)^2\right) \leq \sum_{k=0}^{q} \frac{1}{N^k}\left(1-\frac{1}{N}\right)^{q-k} \sum_{0\leq p_0<\ldots<p_k=q} \exp\left(\sum_{l=1}^{k} V_{p_l/p_{l-1}}(Y)\right)$$

then

$$\tilde{E}_{[Y]}\left((U_n^N)^2\right) \leq \left(1-\frac{1}{N}\right)^n + \sum_{k=0}^{n-1}\left(1-\frac{1}{N}\right)^{n-(1+k)} \frac{1}{N^{k+1}} \sum_{0\leq p_0<\ldots<p_k<n} \exp\left(V_{n/p_k}(Y)+\sum_{l=1}^{k} V_{p_l/p_{l-1}}(Y)\right)$$

and the result follows.

■

The same techniques then establish the following result

Proposition 5 *For any bounded Borel test function $f : S \to \mathbb{R}$, $N \geq 1$ and $n \geq 0$ we have $\tilde{P}$-a.e.*

$$\tilde{E}_{[Y]}\left(((\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f)U_n^N)^2\right) \leq 4\|f\|_\infty^2 \sum_{k=1}^{n+1} \frac{1}{N^k}\left(1-\frac{1}{N}\right)^{(n+1)-k} \sum_{0\leq p_0<\ldots<p_{k-1}\leq p_k=n} \exp\left(\sum_{l=1}^{k} V_{p_l/p_{l-1}}(Y)\right) \tag{49}$$

Therefore

$$\tilde{E}_{[Y]}\left(((\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f)U_n^N)^2\right) \leq 4\|f\|_\infty^2 \left(1-\left(1-\frac{1}{N}\right)^{n+1}\right) \exp\left(V_n(Y)\right) \tag{50}$$

proof:

Let us prove 1). For any bounded Borel test function $f : S \to \mathbb{R}$, using the recursive formulas described in section (3.2) we have

$$\begin{aligned}
\tilde{E}_{[Y]}\left((\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i)U_n^N)^2\right) &= \tilde{E}_{[Y]}\left((\underline{m}_{n+1}^N(f)U_n^N)^2\right) = \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left((\underline{m}_{n+1}^N f)^2/\beta_n\right)(U_n^N)^2\right) \\
\tilde{E}_{[Y]}\left((\underline{m}_{n+1}^N f)^2/\beta_n\right) &\leq \frac{1}{N}\|f\|_\infty^2 + \left(1-\frac{1}{N}\right)\left(\frac{\overline{m}_n^N(g_{n/n-1}f)}{\overline{m}_n^N(g_{n/n-1})}\right)^2 \\
\Longrightarrow \tilde{E}_{[Y]}\left(((\underline{m}_{n+1}^N f)U_n^N)^2\right) &\leq \frac{1}{N}\|f\|_\infty^2 \tilde{E}_{[Y]}\left((U_n^N)^2\right) + \left(1-\frac{1}{N}\right)\tilde{E}_{[Y]}\left(\left(\overline{m}_n^N(g_{n/n-1}f)U_{n-1}^N\right)^2\right)
\end{aligned} \tag{51}$$

Arguing as above for every $1 \leq p \leq n-1$ one gets

$$\begin{aligned}
\tilde{E}_{[Y]}\left(\left(\overline{m}_{p+1}^N(g_{n/p}f_n^{(p+1)})U_p^N\right)^2\right) &= \tilde{E}_{[Y]}\left(\tilde{E}_{[Y]}\left(\overline{m}_{p+1}^N\left(g_{n/p}f_n^{(p+1)}\right)^2/\beta_p\right)(U_p^N)^2\right) \\
\tilde{E}_{[Y]}\left(\overline{m}_{p+1}^N\left(g_{n/p}f_n^{(p+1)}\right)^2/\beta_p\right) &\leq \frac{1}{N}\|f\|_\infty^2 e^{V_{n/p}(Y)} + \left(1-\frac{1}{N}\right)\left(\frac{\overline{m}_p^N\left(g_{n/p-1}\, f_n^{(p)}\right)}{\overline{m}_p^N\left(g_{p/p-1}\right)}\right)^2 \\
\tilde{E}_{[Y]}\left(\overline{m}_1^N\left(g_{n/0}f_n^{(1)}\right)^2/\beta_0\right) &\leq \frac{1}{N}\|f\|_\infty^2 e^{V_{n/0}(Y)} + \left(1-\frac{1}{N}\right)(\pi_n f)^2
\end{aligned}$$

Cascading these inequalities with the expression (51) we conclude that

$$\tilde{E}_{[Y]}\left(\left(\underline{m}_{n+1}^N(f)\,U_n^N\right)^2\right) \leq \frac{1}{N}\|f\|_\infty^2 \sum_{k=0}^{n}\left(1-\frac{1}{N}\right)^{n-k} e^{V_{n/k}(Y)}\tilde{E}_{[Y]}\left((U_k^N)^2\right) + \left(1-\frac{1}{N}\right)^{n+1}(\pi_n f)^2$$

with the convention $V_{n/n}=0$. Then, using formula (48) one obtain easily

$$\begin{aligned}
\tilde{E}_{[Y]}\left(\left(\underline{m}^N_{n+1}(f-\pi_n f)\; U^N_n\right)^2\right) &\leq \frac{4\|f\|^2_\infty}{N}\sum_{l=0}^{n}\frac{1}{N^l}\left(1-\frac{1}{N}\right)^{n-l}\sum_{0\leq p_0<\ldots<p_l\leq p_{l+1}=n}\exp\left(\sum_{s=1}^{l+1} V_{p_l/p_{l-1}}(Y)\right)\\
&\leq 4\|f\|^2_\infty\sum_{l=1}^{n+1}\frac{1}{N^l}\left(1-\frac{1}{N}\right)^{(n+1)-l}\sum_{0\leq p_0<\ldots<p_{l-1}\leq p_l=n}\exp\left(\sum_{s=1}^{l} V_{p_l/p_{l-1}}(Y)\right)
\end{aligned}$$

The second statement follows from 1) proposition 3 and $\binom{n}{l}\leq\binom{n+1}{l}$.

■

5 Convergence Theorems

We are now ready to prove the convergence of the interacting particle approximation described in section 2. The following theorem is our main result, it states the relevant consequences of the mean square error estimates stated in propositions 4 and 5.

Theorem 1 *For every $N\geq 1$, $n\geq 0$ and any bounded Borel test function $f:S\to\mathbb{R}$, $T>0$, and $0<\epsilon<1/2$ we have*

$$\sup_{n\in[0,T]}\tilde{P}_{[Y]}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}^i_n)-\pi_n f\right|>\epsilon\right)\leq\frac{A(T,f,Y)}{\epsilon^4}\left(1-\left(1-\frac{1}{N}\right)^{T+1}\right)\qquad \tilde{P}-a.e. \tag{52}$$

with $A(T,f,Y)=2\sup(4\|f\|^2_\infty,1)\;e^{V_T(Y)}$. Moreover

$$\sup_{n\in[0,T]}\tilde{E}_{[Y]}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}^i_n)-\pi_n f\right|\right)\leq B(T,f,Y)\left(1-\left(1-\frac{1}{N}\right)^{T+1}\right)^{1/2}\qquad \tilde{P}-a.e. \tag{53}$$

with $B(T,f,Y)=4\|f\|_\infty\; e^{V_T(Y)/2}$

proof:
Using propositions 4 and 5 and Cauchy-Schwartz inequality we have

$$\begin{aligned}
\tilde{E}_{[Y]}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}^i_n)-\pi_n f\right|\right) &= \tilde{E}_{[Y]}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}^i_n)-\pi_n f\right| U^N_n\right)+\tilde{E}_{[Y]}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}^i_n)-\pi_n f\right|(1-U^N_n)\right)\\
&\leq B(T,f,Y)\left(1-\left(1-\frac{1}{N}\right)^{n+1}\right)^{1/2}
\end{aligned}$$

and the inequality (53) follows. To prove (52) write

$$A=\{U^N_n\geq 1-\epsilon\}\qquad B=\{|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}^i_n)-\pi_n f|>\epsilon\}$$

Using the implication

$$\forall 0<\epsilon<1\quad |U^N_n-1|\leq\epsilon\Longrightarrow U^N_n\geq 1-\epsilon$$

and proposition 4 we have

$$\tilde{P}_{[Y]}\left(U_n^N \geq 1-\epsilon\right) \geq \tilde{P}_{[Y]}\left(|U_n^N - 1| \leq \epsilon\right) \geq 1-\frac{1}{\epsilon^2}\tilde{E}_{[Y]}\left((U_n^N-1)^2\right) \geq 1-\frac{1}{\epsilon^2}\left(1-\left(1-\frac{1}{N}\right)^n\right)\ e^{V_n(Y)}$$

Therefore

$$\tilde{P}_{[Y]}\left(\overline{A}\right) \leq \frac{1}{\epsilon^2}\ \left(1-\left(1-\frac{1}{N}\right)^{n+1}\right)\ e^{V_n(Y)}$$

On the other hand 5 gives the inequalities

$$\begin{aligned}\tilde{P}_{[Y]}(B\cap A) &\leq \frac{1}{\epsilon^2}\tilde{E}_{[Y]}\left(|\frac{1}{N}\sum_{i=1}^N f(\widehat{x}_n^i) - \pi_n f|^2\ 1_A\right) \leq \frac{1}{((1-\epsilon)\epsilon)^2}\tilde{E}_{[Y]}\left(\left((\frac{1}{N}\sum_{i=1}^N f(\widehat{x}_n^i) - \pi_n f)\ U_n^N\right)^2\right) \\ &\leq \frac{4\,\|f\|_\infty^2}{((1-\epsilon)\epsilon)^2}\left(1-\left(1-\frac{1}{N}\right)^{n+1}\right)\ e^{V_n(Y)}\end{aligned}$$

Finally the inequality

$$\tilde{P}_{[Y]}(B) \leq \tilde{P}_{[Y]}(B\cap A) + \tilde{P}_{[Y]}(\overline{A}) \tag{54}$$

and the implication $0 < \epsilon < 1/2 \Rightarrow \epsilon < 1-\epsilon$ end the proof.

■

Corollary 1 *For every $N \geq 2$, $n \geq 0$ and for any integrable Borel test function $f: S \to \mathbb{R}$, $T > 0$, and $0 < \epsilon < 1/2$ we have*

$$\sup_{n\in[0,T]} \tilde{P}_{[Y]}\left(\left|\frac{1}{N}\sum_{i=1}^N f(\widehat{x}_n^i) - \pi_n f\right| > \epsilon\right) \leq \frac{A(\epsilon,T,f,Y)}{N} \qquad \tilde{P}-a.e. \tag{55}$$

where $A(\epsilon,T,f,Y) = 4\,T\ A(T,f,Y)/\epsilon^4$.

proof:
From the inequality $e^y \geq 1+y$ it follows that

$$1-\left(1-\frac{1}{N}\right)^{n+1} \leq -(n+1)\log\left(1-\frac{1}{N}\right) = (n+1)\log\left(\frac{N}{N-1}\right)$$

and $\log x \leq x-1$. Thus we conclude

$$1-\left(1-\frac{1}{N}\right)^{n+1} \leq \frac{n+1}{N-1} \leq \frac{2(n+1)}{N} \leq \frac{4\,n}{N} \qquad \forall N \geq 2 \quad \forall n \geq 1$$

This completes the proof.

■

Theorem 2 *Let $Y = y \in \mathbb{R}^{\mathbb{N}}$ be a series of observations such that $V_n(y) < +\infty$. For every bounded Borel test function $f: S \to \mathbb{R}$ $N \geq 1$ and $n \geq 0$, we have*

$$\forall p > 0 \qquad \sup_{n\in[0,T]} \tilde{E}_{[y]}\left(\left(\frac{1}{N}\sum_{i=1}^N f(\widehat{x}_n^i) - \pi_n f\right)^p\right) \xrightarrow[N\to+\infty]{} 0 \tag{56}$$

proof:
Using the inequality

$$\forall \epsilon > 0 \qquad \tilde{E}_{[y]}\left(\left(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\right)^p\right) \leq \epsilon^p + (2\,\|f\|_\infty)^p\,\tilde{P}_{[y]}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\right| > \epsilon\right)$$

and Corollary 1 the result follows.

∎

Consider the following assumption on the observation process:

$$(\mathcal{V}) \qquad \forall n \geq 0 \quad \tilde{E}\left(V_n(Y)\right) \stackrel{def}{=} \alpha(n) < +\infty \tag{57}$$

This assumption enable us to estimate the convergence rate of our approximations in spaces $L^p(\tilde{P})$ with $p > 0$

Theorem 3 *Assume $\mathcal{V}$ is satisfied. For every bounded Borel test function $f : S \to \mathbb{R}$ $N \geq 1$, $n \geq 0$, $0 < \epsilon < 1/2$ and $M > 0$ we have*

$$\sup_{n\in[0,T]} \tilde{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\right| > \epsilon\right) \leq \frac{1}{M} + \frac{A(T,M,f,\epsilon)}{N} \tag{58}$$

with $A(T,M,f,\epsilon) = 8\sup(4\|f\|_\infty, 1)\; T\; e^{M\,\alpha(T)}/\epsilon^4$

proof:
Using theorem 1

$$\tilde{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\right| \leq \epsilon\right) \geq \tilde{P}\left(V_n(Y) \leq M\alpha\,(n)\right) - \frac{A(f)\; e^{M\alpha(n)}}{\epsilon^4}\left(1 - \left(1 - \frac{1}{N}\right)^{n+1}\right)$$

with $A(f) = 2\sup(4\,\|f\|_\infty, 1)$, then

$$\tilde{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\right| \leq \epsilon\right) \geq 1 - \left(\frac{1}{M} + \frac{A(f)\; e^{M\alpha(n)}}{\epsilon^4}\left(1 - \left(1 - \frac{1}{N}\right)^{n+1}\right)\right)$$

and the arguments used in the proof of Corollary 1 end the proof.

∎

Finally by the same line of arguments as for Theorem 2

Corollary 2 *Assume $\mathcal{V}$ is satisfied. For every bounded Borel test function $f : S \to \mathbb{R}$, $N \geq 1$ and $n \geq 0$, we have*

$$\forall p > 0 \qquad \sup_{n\in[0,T]} \tilde{E}\left(\left(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\right)^p\right) \xrightarrow[N \to +\infty]{} 0 \tag{59}$$

References

[1] H. Carvalho, P. Del Moral, A. Monin, G. Salut, Optimal non-linear filtering in GPS/INS Integration, IEEE trans. on Aerospace and electronic systems 1995 (to appear).

[2] R. Cerf, *Une théorie asymptotique des Algorithmes Génétiques*, Université Montpellier II, Sciences et techniques du Languedoc, 1994.

[3] M. Chaleyat-Maurel, D. Michel, *Des résultats de non existence de filtres de dimension finie*, C.R.Acad.Sci. Paris, Serie I, t. 296, 1983.

[4] P. Del Moral, G. Rigal, J.C. Noyer, G. Salut, *Traitement non-linéaire du signal par reseau particulaire : Application Radar*, quatorzième colloque GRETSI, Juan les Pins, 13-16 Septembre 1993.

[5] P. Del Moral, *Résolution particulaire des problèmes d'estimation et d'optimisation non-linéaires*, Thèse de l'Université Paul Sabatier, Rapport LAAS n^o94269, Juin 1994.

[6] P. Del Moral, G. Salut, *Filtrage non-linéaire : résolution particulaire à la Monte-Carlo*, C.R. Acad. Sci. Paris, Serie I, t. 320, 1995.

[7] P. Del Moral, *Non linear filtering using random particles* Probability Theory an Applications, Volume 40, No 4, 1995.

[8] P. Del Moral, J.C. Noyer, G. Salut, Résolution particulaire et traitement non-linéaire du signal : Application Radar/Sonar, Traitement du signal, Septembre 1995.

[9] P. Del Moral, G. Salut, Particle interpretation of nonlinear filtering and optimization Russian Journal of Mathematical Physics, in press, 1996.

[10] N.J. Gordon, D.J. Salmon, A.F.M. Smith, *Novel Approach to non-linear/non-Gaussian Bayesian state estimation*, IEE 1993

[11] J.H. Holland *Adaptation in natural and artificial systems* University of Michigan Press, Ann Arbor, 1975

[12] H. Kunita *Asymptotic Behavior of Nonlinear Filtering Errors of Markov Processes*, Journal of Multivariate Analysis, Vol. 1, *No* 4, p. 365-393, Dec. 1971

[13] L. Stettner *On Invariant Measures of Filtering Processes*, Stochastic Differential Systems, Lecture Notes in Control and Information Sciences, 126, Springer Verlag, 1989.