Nonlinear Filtering: Interacting Particle Resolution

P. Del Moral
Laboratoire de Statistiques et Probabilités, CNRS–UMR C55830, Bat 1R1, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France

Markov Processes and Related Fields, Volume 2 Number 4, 555–580 (1996).
Received April 2, 1996, revised October 3, 1996

Abstract. This paper covers stochastic particle methods for the numerical solution of the nonlinear filtering equations based on the simulation of interacting particle systems. The main contribution of this paper is to prove convergence of such approximations to the optimal filter, thus yielding what seemed to be the first convergence results for such approximations of the nonlinear filtering equations. This new treatment has been 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.

KEYWORDS: nonlinear filtering, Bayesian estimation, Monte-Carlo sampling, interacting particle systems, sampling-resampling, evolutionary processes, genetic algorithms

AMS SUBJECT CLASSIFICATION: 60F10, 60G35, 60J10, 62D05, 62E25, 62F12, 62G05, 62L20, 62M05, 92D15, 93E10, 93E11

1. 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 $S \times \mathbf{R}$ and defined by the system:

$$\mathcal{F}(X \mid Y) \qquad \begin{cases} X & \sim & (\nu, K) \\ Y_n & = & h(X_n) + V_n, \quad n \geq 1, \end{cases} \tag{1.1}$$

where $S = \mathbf{R}^d$, $d \geq 1$, $h : S \to \mathbf{R}$ and V_n are independent random variables having a density g_n with respect to the Lebesgue measure. The signal process X

that we consider, is assumed to be a temporally homogeneous Markov process on S with transition probability kernel K and initial probability measure ν. We assume that the observation noise V and the state plant X are independent. For simplicity, the observation process Y is taken real valued, the extension to vector observation processes is straightforward. One can study filtering problems in more general settings. We choose not to do so here, and prefer to focus on the central ideas. 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 Borel test function from S into $\mathbf{R}$. The best estimate of X_n given the observations up to time n, is the conditional expectation

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

With the notable exception of the linear-Gaussian situation, general optimal filters have no finitely recursive solution [3]. This paper covers stochastic particle methods for the numerical solution of the nonlinear filtering equations based on 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 examples of practical problems that 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 notation and we recall some basic facts related to the theory introduced by Kunita and Stettner [12, 13]. In Section 3 we introduce the interacting particle approximation and we design a natural stochastic basis for the convergence study. In Section 4 we describe recursive formulas for the conditional distributions and the context that we are interested in. We propose a natural framework which allows to explicitly formulate 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 5: the convergence study of our algorithm requires a 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 the convergence rates is discussed in Section 6. This last section contains our main result: we prove that the interacting particle filters converge to the conditional distribution, as the number of particles tends to infinity. The convergence rate estimates arise quite naturally from the results and associated methodologies of Sections 5 and 6.

2. Non linear filtering equation

2.1. General notations

Before starting the description of the nonlinear filtering equation, let us first introduce some general notation.

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:

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

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 notation

$$\begin{aligned} \mu K_1(dy) &= \int \mu(dx)\,K_1(x,dy) & K_1K_2(x,dz) &= \int K_1(x,dy)\,K_2(y,dz) \\ K_1 f(x) &= \int K_1(x,dy)\,f(y) & \mu f &= \int \mu(dx)\,f(x). \end{aligned}$$

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

$$\text{for all } f \in \mathcal{C}(S): \quad \overline{m}f = \int m(dx_1,dx_2)\,f(x_2), \quad \underline{m}f = \int m(dx_1,dx_2)\,f(x_1).$$

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

$$\text{for all } f \in \mathcal{C}(S^2): \quad (\mu \times K)f = \int \mu(dx_1)\,K(x_1,dx_2)\,f(x_1,x_2).$$

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

2.2. 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 $(\Omega = \Omega_1 \times \Omega_2, F_n, \mathsf{P})$ be the canonical space for the signal observation pair (X, Y). Therefore P is the probability measure on Ω corresponding to the filtering model $\mathcal{F}(X \mid Y)$, 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 $\mathsf{E}(\cdot)$ to denote expectation with respect to P on Ω. To clarify the presentation, we will also write

$$m_{n+1} = \pi_n \times K. \tag{2.1}$$

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

$$\pi_n = \rho_n(\pi_{n-1}, Y_n) \quad n \geq 1 \quad \pi_0 = \nu, \tag{2.2}$$

where

$$\rho_n(\mu, y)f = \frac{\int f(x)\, g_n(y - h(x))\, (\mu\, K)(dx)}{\int g_n(y - h(x))\, (\mu\, K)(dx)} \tag{2.3}$$

for all $f \in \mathcal{C}(S)$, $\mu \in \mathcal{P}(S)$, $y \in \mathbf{R}$ 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 $(\sigma(Y^n), \mathsf{P})$-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 $F : \mathcal{P}(S) \longrightarrow \mathbf{R}$ and $\mu \in \mathcal{P}(S)$. In other words, with some obvious abuse of notation

$$\begin{aligned} dp(y_n, x_n, x_{n-1} \mid \pi_{n-1}) & \\ &= g_n(y_n - h(x_n))\, dy_n\, \pi_{n-1}(dx_{n-1})\, K(x_{n-1}, dx_n) && (2.4) \\ p(y_n/\pi_{n-1}) &= \int g_n(y_n - h(x_n))\, (\pi_{n-1}\, K)(dx_n) && (2.5) \end{aligned}$$

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 (2.1) and (2.2). Using the above notation one easily gets

$$m_{n+1} = \Phi_n(m_n, Y_n) \quad n \geq 1 \quad m_0 = \nu \times K, \tag{2.6}$$

where

$$\Phi_n(m, y)f = \frac{\int f(x_1, x_2)\, g_n(y - h(x_1))\, m(dx_0, dx_1)\, K(x_1, dx_2)}{\int g_n(y - h(z_1))\, m(dz_0, dz_1)} \tag{2.7}$$

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

3. Interacting particle systems

3.1. 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-tuple of elements of S, i.e. the points of the set S^N, are called systems of particles and they will be mostly denoted by the letters x, y, z. Given the observations $Y = y$, we denote by $(\widehat{x}_n, x_{n+1})_{n\geq 0}$ the $\mathcal{P}(S^2)$-Markov process defined by the transition probabilities

$$\begin{aligned}
&\tilde{\mathsf{P}}_{[y]}\big\{(\widehat{x}_0, x_1) \in d(z_0, z_1)\big\} = \prod_{p=1}^{N} m_0(dz_0^p, dz_1^p) \\
&\tilde{\mathsf{P}}_{[y]}\big\{(\widehat{x}_n, x_{n+1}) \in d(z_0, z_1) \;\big|\; (\widehat{x}_{n-1}, x_n) = (x_0, x_1)\big\} \\
&\quad = \prod_{p=1}^{N} \Phi_n\Big(\frac{1}{N}\sum_{i=1}^{N} \delta_{(x_0^i, x_1^i)}, y_n\Big)(d(z_0^p, z_1^p)). \qquad (3.1)
\end{aligned}$$

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

Initial Particle System

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

Sampling/Exploration

$$\tilde{\mathsf{P}}_{[y]}\{x_n = dx \mid \widehat{x}_{n-1} = z\} = \prod_{p=1}^{N} K(z^p, dx^p),$$

Resampling/Selection

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

Finally one gets a sequence of particle 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 from the population $\{x_n^1, \ldots, x_n^N\}$ by the law π_n^N defined by the likelihood functions and the present measurement Y_n. Namely

$$\pi_n^N = \sum_{i=1}^{N} \frac{g_n(y_n - h(x_n^i))}{\sum_{j=1}^{N} g_n(y_n - h(x_n^j))}\, \delta_{x_n^i} \qquad (3.2)$$

Then it moves to x_{n+1}^i using 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 $\widehat{x}_n = (x_n^1, \ldots, x_n^N)$, representing the successive particles obtained by exploring the probability space with the transitions K. More precisely, the motion of particles is decomposed into two stages:

$$\widehat{x}_n \xrightarrow{\text{Sampling/Exploration}} x_n \xrightarrow{\text{Resampling/Selection}} \widehat{x}_{n+1}. \tag{3.3}$$

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 it differs from the particle resolutions introduced in [7] and [6].

Remark 3.1. The formulated algorithm permits generalisation in the case that the selecting procedure is used from time to time. In practical situations, a simple way to do this consists of introducing a resampling schedule. For instance, we may choose to resample the particles, when fifty percent of the weights is lower than AN^{-p} for a convenient choice of $A > 0$ and $p \geq 2$.

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

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 tends to infinity;
3. a noise observation V_n, also governed by a parameter l, with distribution
$$d\mathsf{P}_n^V(v) = \frac{\exp\left(-V(v)\log l\right) dv}{\int \exp\left(-V(u)\log l\right) du};$$
4. homogeneous series of observations with respect to time
$$\text{for all } n \geq 1 \quad Y_n = y \qquad V(y - h(x)) \stackrel{\text{def}}{=} f(x).$$

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

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

For this very special situation, Cerf [2] gives several conditions on the rate of decrease of the perturbations $1/\log l(n)$, to ensure that 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 on the convergence of such algorithms for numerically solving the nonlinear filtering equations.

The crucial question is of course, whether the empirical random measure

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{\widehat{x}_n^i}$$

converges to the conditional distribution π_n, when the size of the particle system is growing. This is positively answered in Theorem 6.2 Section 6. We will show that for every bounded Borel test function $f : S \to \mathbf{R}$ and for $n \geq 1$

$$\lim_{N\to+\infty} \mathsf{E}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n(f)\right|\right) = 0. \tag{3.4}$$

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

Such particle nonlinear filters differ 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 explores the probability space using the *a priori* Markov kernel and updates the information that is accumulated through resampling of the positions. In introducing such adaptation/selection laws, the particle system acquires certain configurations, so as to represent an estimate of the conditional distribution. They provide a natural procedure for a system of particles to sense its environment through their likelihood functions and the observations.

It is clear that such particle resolutions could be formulated in an natural way in other contexts as neural networks, model parameter identification, optimal control ([9]).

3.2. The associated Markov process

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

$$\mathcal{P}(S^2) \xrightarrow{C_{\mathrm{N}}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \delta_{x^i} \ : \ x^i \in S^2 \right\} \subset \mathcal{P}(S^2), \tag{3.5}$$

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

$$C_{\mathrm{N}} F(\eta) = \int F(m)\ C_{\mathrm{N}}(\eta, dm) \stackrel{\text{def}}{=} \int\limits_{S^{2\mathrm{N}}} F\Big(\frac{1}{N} \sum_{i=1}^{N} \delta_{x^i} \Big)\, \eta(dx^1) \ldots \eta(dx^{\mathrm{N}}). \tag{3.6}$$

The interpretation of C_N is the following: by starting with a measure $\eta \in \mathcal{P}(S^2)$, the next measure is the result of 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 observation enables us to consider the empirical measures m_n^N as a canonical $\mathcal{P}(S^2)$-valued Markov process $(\Omega', \beta_n, \tilde{\mathsf{P}}_{[y]})$, by setting

$$\begin{aligned} \tilde{\mathsf{P}}_{[y]}\{m_1^N \in d\eta\} \quad &= \quad C_N(\Phi_n(m_0, y_n), d\eta) \\ \tilde{\mathsf{P}}_{[y]}\{m_{n+1}^N \in d\eta \mid m_n^N = \mu\} \quad &= \quad C_N(\Phi_n(\mu, y_n), d\eta). \end{aligned} \tag{3.7}$$

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 $(\Omega, F_n, \mathsf{P})$ 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 $\tilde{\omega} \stackrel{\text{def}}{=} (\omega^1, \omega^2, \omega^3) \in \tilde{\Omega}$, 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 $\tilde{\mathsf{P}}$ as follows:

$$\tilde{\mathsf{P}}\{A \times B\} \stackrel{\text{def}}{=} \int_B \tilde{\mathsf{P}}_{[Y(\omega)]}(A)\, d\mathsf{P}(\omega). \tag{3.8}$$

As usual we use $\tilde{\mathsf{E}}\,(\cdot)$ to denote the expectation with respect to $\tilde{\mathsf{P}}$, and $\tilde{\mathsf{E}}_{[y]}\,(\cdot)$ to denote the expectation with respect to $\tilde{\mathsf{P}}_{[y]}$. With some obvious abuse of notation we have

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

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

4. General recursive formulas

In this section we shall adopt 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 quite different setup from the one used in Sections 2 and 3, but in the end of our investigations will resemble more and more the models presented in Section 2.

The setting is the same as in Section 3.1. In particular, unless otherwise stated, we assume the observation data to be a fixed series of real numbers $Y = y$.

This assumption enables us to consider the conditional distribution π_n as a probability parametrised 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 (2.2) is the recursive formula

$$\pi_n f = \frac{\pi_{n-1} K(\, f\, g_n)}{\pi_{n-1} K(g_n)}, \qquad \text{for all } f \in \mathcal{C}(S).$$

4.1. Likelihood functions analysis

The proof of the convergence (3.4) involves

1. the Maximum Log-likelihood functions given by

$$\begin{aligned} V_{n/p}(y) &= \sup_{x_p \in S} \log \int_S \left(\frac{p(y_n, \ldots, y_{p+1} \mid x_{p+1})}{p(y_n, \ldots, y_{p+1} \mid y^p)} \right)^2 dp(x_{p+1} \mid x_p), \quad (4.1) \\ & \qquad\qquad p+1 \leq n \\ V_n(y) &= \sum_{p=1}^{n} V_{p/p-1}(y), \quad (4.2) \end{aligned}$$

where

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

2. the conditional expectations for $0 \leq p \leq n-1$, given by

$$f_n^{(p)}(x_p) = \mathsf{E}\left(f(X_n) \mid X_p, Y_{p+1}, \ldots, Y_n\right)(x_p, y_{p+1}, \ldots, y_n). \tag{4.3}$$

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

$$p(y_n \mid x_n) = g_n(y_n - h(x_n)) \tag{4.4}$$

$$p(y_n, \ldots, y_p \mid x_p) = \int_S p(y_n, \ldots, y_{p+1} \mid x_{p+1})\, dp(x_{p+1} \mid x_p). \tag{4.5}$$

Example 4.1.
1. Assume here that state and observation processes (X, Y) are given by the linear dynamics

$$X_n = A\, X_{n-1} + W_n \tag{4.6}$$

$$Y_n = C\, X_n + V_n, \tag{4.7}$$

where $X_n \in \mathbf{R}$, $Y_n \in \mathbf{R}$, $Y_0 = 0$, A and C are real numbers and X_0, W_n and V_n are normally distributed with means 0 and respective non-negative covariances 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

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

with the well known measurement update equations

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

In this very special situation

$$V_{n/n-1}(Y) \leq \log\Big[\frac{|C^2 P_{n/n-1}+R|}{|R|}\exp\big((Y_n - CA\widehat{X}_{n-1})^2(C^2P_{n/n-1}+R)^{-1}\big)\Big].$$

Further manipulations yield

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

with $S_0 = 0$ and, for $A \neq 1$,

$$\mathsf{E}\Big((X_k - \mathsf{E}(X_k/X_p))^2\Big) = S_{k-p} = \frac{1-A^{k-p}}{1-A}\,Q.$$

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

$$dp_n^V(v) = \frac{\exp(-U_n)}{\int \exp(-U_n(v))\,dv} \qquad U_n : S \to \mathbf{R}^+.$$

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

The functions (4.1), (4.3) and (4.5) will be used in our analysis of the convergence (3.4). This analysis requires a specific development, because of the difficulty of estimating mean errors. For instance we have

$$\begin{aligned}
\tilde{\mathsf{E}}_{[y]}\Big(\Big(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i)\Big)^2\Big) &= \tilde{\mathsf{E}}_{[y]}\big((\underline{m}_{n+1}^N f)^2\big)\\
&= \frac{1}{N}\tilde{\mathsf{E}}_{[y]}(\pi_n^N(f^2)) + \big(1-\frac{1}{N}\big)\tilde{\mathsf{E}}_{[y]}\big((\pi_n^N f)^2\big)\\
&= \frac{1}{N}\tilde{\mathsf{E}}_{[y]}\big(\pi_n^N(f-\pi_n^N f)^2\big) + \tilde{\mathsf{E}}_{[y]}\big((\pi_n^N f)^2\big).
\end{aligned}$$

Thus, Chebyshev's inequality and convenient estimates of

$$\tilde{\mathsf{E}}_{[y]}\big((\pi_n^N f)^2\big) = \tilde{\mathsf{E}}_{[y]}\Big(\Big(\sum_{i=1}^{N}\frac{g_n(y_n - h(x_n^i))}{\sum_{j=1}^N g_n(y_n - h(x_n^j))}\,f(x_n^i)\Big)^2\Big) \tag{4.8}$$

would give a complete answer concerning convergence in $\mathbf{L}^0(\tilde{\mathsf{P}}_{[y]})$. Unfortunately, it is difficult to estimate (4.8) or even $\tilde{\mathsf{E}}_{[y]}\left(\pi_n^N f\right)$.

Accordingly, in order to point out the connections between (4.3) and (4.5) and to so emphasize the role of the Log-likelihood functions $V_{n/p}$, we first introduce a natural framework in which the relationship between (4.3) and (4.5) is made explicit. The recursive expressions described in the next section, will be used repeatedly for the convergence (3.4) in the last part of the paper.

4.2. Recursive formulas

The main purpose of this section is to introduce recursive expressions for (4.3) and (4.5). The technical approach presented here, is to work with 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 parametrised by the observation parameters y. The recursive formulas described in this section, will show how the sequence of observations scales the updates 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 on the use of Bayes' formula in our setting. With some obvious abuse of notation, we have for all $1 \leq p \leq n$ the recursion

$$\begin{aligned}
p(y_n, \ldots, y_p \mid x_p) &= p(y_p \mid x_p) p(y_n, \ldots, y_{p+1} \mid x_p) \\
&= p(y_p \mid x_p) \int p(y_n, \ldots, y_{p+1} \mid x_{p+1})\, dp(x_{p+1} \mid x_p) \\
\frac{p(y_n, \ldots, y_p \mid x_p)}{p(y_n, \ldots, y_p \mid y^{p-1})} &= \frac{p(y_p \mid x_p)}{p(y_p \mid y^{p-1})} \int \frac{p(y_n, \ldots, y_{p+1} \mid x_{p+1})}{p(y_{p+1} \mid y^p)}\, dp(x_{p+1} \mid x_p) \\
p(y_n, \ldots, y_1) &= p(y_n \mid y^{n-1})\, p(y_{n-1} \mid y^{n-2}) \cdots p(y_2 \mid y_1)\, p(y_1).
\end{aligned}$$

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

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

With a slight abuse of notation we will often write $g_n(x_n)$ instead of $g_n^{(n)}(x_n) = p(y_n \mid x_n)$. It is now easily checked from the above remarks, that

$$g_{n/p-1}(x_p) = g_n^{(p)}(x_p) \int \pi_p(dz_{p-1})\, K(z_{p-1}, dz_p)\, g_n^{(p)}(z_p) \tag{4.9}$$

$$g_n^{(p)}(x_p) = g_p(x_p) \int K(x_p, dx_{p+1})\, g_n^{(p+1)}(x_{p+1}) \tag{4.10}$$

$$g_{n/p-1}(x_p) = g_{p/p-1}(x_p) \int K(x_p, dx_{p+1})\, g_{n/p}(x_{p+1}). \tag{4.11}$$

Summarising, we have the backward recursive formulas.

Proposition 4.1. *For every* $0 \leq p < n-1$

$$\begin{aligned}
g_n^{(p)} &= g_p\, (K\, g_n^{(p+1)}) \\
g_{n/p-1} &= \frac{g_n^{(p)}}{\pi_{p-1} K g_n^{(p)}} \\
g_{n/p-1} &= g_{p/p-1}\, K(g_{n/p}).
\end{aligned} \tag{4.12}$$

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

$$\begin{aligned} \pi_n(f) &= \frac{\int f(z)\, g_n(y_n - h(z))\, \overline{m}_n(dz)}{\int g_n(y_n - h(z))\, \overline{m}_n(dz)} \\ \pi_n^N(f) &= \sum_{i=1}^N \frac{g_n(y_n - h(x_n^i))}{\sum_{j=1}^N g_n(Y_n - h(x_n^j))}\, f(x_n^i). \end{aligned} \tag{4.13}$$

For brevity we will write

$$\pi_n(f) = \frac{\overline{m}_n\,(g_n\, f)}{\overline{m}_n\,(g_n)}, \qquad \pi_n^N(f) = \frac{\overline{m}_n^N\,(g_n\, f)}{\overline{m}_n^N\,(g_n)}.$$

Continuing in the same vein, we derive the conditional expectations $f_n^{(p)}$, introduced in (4.3).
Using Bayes' rule we have for all $1 \leq p \leq n$ the following basic equation

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

By the same line of arguments, for all $1 \leq p \leq n$, we get

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

Thus we arrive at

$$\pi_n f = \frac{\int \pi_{p-1}(dx_{p-1})\, K(x_{p-1}, dx_p)\, f_n^{(p)}(x_p)\, g_n^{(p)}(x_p)}{\int \pi_{p-1}(dx_{p-1})\, K(x_{p-1}, dx_p)\, g_n^{(p)}(x_p)}, \tag{4.14}$$

for all $f \in \mathcal{C}(S)$, $1 \leq p \leq n$. Additionally we have

$$f_n^{(p-1)}(x_{p-1}) = \frac{\int K(x_{p-1}, dx_p)\, f_n^{(p)}(x_p)\, g_n^{(p)}(x_p)}{\int K(x_{p-1}, dx_p)\, g_n^{(p)}(x_p)}, \tag{4.15}$$

for all $f \in \mathcal{C}(S)$, $1 \leq p \leq n$. Summarising, our conditional expectations $f_n^{(p)}$ can be described as follows.

Proposition 4.2. *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 $1 \leq p \leq n$*

$$f_n^{(p-1)} \stackrel{\text{def}}{=} \frac{K(f_n^{(p)}\ g_{n/p-1})}{K(g_{n/p-1})} = \frac{K(f_n^{(p)}\ g_n^{(p)})}{K(g_n^{(p)})}, \qquad \textit{for all}\ \ 1 \leq p \leq n. \tag{4.16}$$

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

$$\pi_n f = \frac{(\pi_{p-1}\,K)\big(f_n^{(p)}\ g_n^{(p)}\big)}{(\pi_{p-1}\,K)\big(g_n^{(p)}\big)} = \frac{(\pi_{p-1}\,K)\big(f_n^{(p)}\ g_{n/p-1}\big)}{(\pi_{p-1}\,K)\big(g_{n/p-1}\big)}. \tag{4.17}$$

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

5. Mean square estimates

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

Lemma 5.1. *Let $f : S \longrightarrow \mathbf{R}$ be any integrable Borel test function, and let $n \geq 0$ and $N \geq 1$.*
We have $\tilde{\mathsf{P}}$-a.e.

$$\tilde{\mathsf{E}}_{[Y]}\left(\underline{m}_{n+1}^N f \mid \beta_n\right) = \pi_n^N f \qquad \tilde{\mathsf{E}}_{[Y]}\left(\overline{m}_{n+1}^N f \mid \beta_n\right) = \pi_n^N K f \tag{5.1}$$

$$\tilde{\mathsf{E}}_{[Y]}\left((\underline{m}_{n+1}^N f)^2 \mid \beta_n\right) = \frac{1}{N}\pi_n^N(f^2) + \big(1-\frac{1}{N}\big)(\pi_n^N f)^2 \tag{5.2}$$

$$\tilde{\mathsf{E}}_{[Y]}\left((\overline{m}_{n+1}^N f)^2 \mid \beta_n\right) = \frac{1}{N}\pi_n^N K(f^2) + \big(1-\frac{1}{N}\big)(\pi_n^N K f)^2. \tag{5.3}$$

Proof. It suffices to note that

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

□

Next we derive a technical recursive formula in p, for the expressions

$$\tilde{\mathsf{E}}_{[Y]}\big(\overline{m}^N_{p+1}(g_{n/p}\ f^{(p+1)}_n) \mid \beta_p\big) \qquad 0 \leq p \leq n-1.$$

Lemma 5.2. *Let $f : S \longrightarrow \mathbf{R}$ be any integrable Borel test function and let $n \geq 0$ and $N \geq 1$. For every $0 \leq p \leq n-1$ we have the recursion*

$$\begin{aligned}\tilde{\mathsf{E}}_{[Y]}\big(\overline{m}^N_{p+1}(g_{n/p}\ f^{(p+1)}_n) \mid \beta_p\big) &= \pi^N_p K\big(g_{n/p}\ f^{(p+1)}_n\big) \\ &= \frac{\overline{m}^N_p\big(g_{n/p-1}\ f^{(p)}_n\big)}{\overline{m}^N_p\big(g_{p/p-1}\big)}, \qquad \tilde{\mathsf{P}}\text{-a.e.}\end{aligned} \tag{5.4}$$

Proof. Using the inductive definition of $f^{(p)}_n$, $g_{n/p}$ and Lemma 5.1, we have

$$\begin{aligned}\tilde{\mathsf{E}}_{[Y]}\big(\overline{m}^N_{p+1}(g_{n/p}f^{(p+1)}_n) \mid \beta_p\big) &= \pi^N_p K(g_{n/p}f^{(p+1)}_n) \\ &= \frac{\overline{m}^N_p\big(g_{p/p-1}K(g_{n/p}f^{(p+1)}_n)\big)}{\overline{m}^N_p\big(g_{p/p-1}\big)} \\ &= \frac{\overline{m}^N_p\big(g_{n/p-1}\ f^{(p)}_n\big)}{\overline{m}^N_p\big(g_{p/p-1}\big)}.\end{aligned}$$

□

5.1. Martingale approach

For estimating mean square errors, the key idea is to introduce a $(\tilde{\mathsf{P}}, \tilde{F}_n)$-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 5.1. We denote by U^N the stochastic process defined by

$$U^N_0 = 1, \qquad U^N_n = \overline{m}^N_n(g_{n/n-1})\, U^N_{n-1}, \qquad \text{for all } \ n \geq 1. \tag{5.5}$$

In other words,

$$U^N_n = \prod_{k=1}^{n} \Big(\frac{1}{N}\sum_{i=1}^{N} g_{k/k-1}(x^i_k)\Big).$$

Lemma 5.3. *U^N_n is a $(\tilde{\mathsf{P}}, \tilde{F}_n)$-martingale with $\tilde{\mathsf{E}}(U^N_n) = 1$ and $\tilde{\mathsf{P}}$-a.s. $\tilde{\mathsf{E}}_{[Y]}(U^N_n) = 1$.*

Proof. The first statement follows by recalling that

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

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

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

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

$$\begin{aligned}\tilde{\mathsf{E}}_{[Y]}(U_n^N) &= \tilde{\mathsf{E}}_{[Y]}\big(\overline{m}_n^N(g_{n/n-1})U_{n-1}^N\big) \\ &= \tilde{\mathsf{E}}_{[Y]}\big(\tilde{\mathsf{E}}_{[Y]}(\overline{m}_n^N(g_{n/n-1}) \mid \beta_{n-1})U_{n-1}^N\big). \end{aligned} \tag{5.6}$$

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

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

Then (5.6) gives

$$\begin{aligned}\tilde{\mathsf{E}}_{[Y]}(U_n^N) &= \tilde{\mathsf{E}}_{[Y]}(\overline{m}_{n-1}^N(g_{n/n-2})\ U_{n-2}^N) \\ &= \tilde{\mathsf{E}}_{[Y]}\big(\tilde{\mathsf{E}}_{[Y]}(\overline{m}_{n-1}^N(g_{n/n-2}) \mid \beta_{n-2})U_{n-2}^N\big).\end{aligned}$$

This procedure can be repeated. Using the recursive formulas described in Section 4.2, we note that

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

Then

$$\tilde{\mathsf{E}}_{[Y]}(U_n^N) = \tilde{\mathsf{E}}_{[Y]}\big(\tilde{\mathsf{E}}_{[Y]}(\overline{m}_{p-1}^N(g_{n/p-2}) \mid \beta_{p-2})U_{p-2}^N\big).$$

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

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

□

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 makes the calculation of mean errors possible, then these estimates will be reinterpreted back. As a typical example we have the following lemma.

Lemma 5.4. *For any integrable Borel test function $f : S \longrightarrow \mathbf{R}$, we have*

$$\tilde{\mathsf{E}}_{[Y]}\left(U_n^N\ \underline{m}_{n+1}^N(f - \pi_n f)\right) = 0, \qquad \tilde{\mathsf{P}}\text{-a.e.} \tag{5.7}$$

Proof. Using Lemma 5.3, this is equivalent to proving

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

Now, using Lemma 5.1 and the fact that $f_n^{(n)} = f$, note that

$$\begin{aligned}
\tilde{\mathsf{E}}_{[Y]}(U_n^N\ \underline{m}_{n+1}^N f) &= \tilde{\mathsf{E}}_{[Y]}\big(\tilde{\mathsf{E}}_{[Y]}(\underline{m}_{n+1}^N f \mid \beta_n) U_n^N\big)\\
\tilde{\mathsf{E}}_{[Y]}(\underline{m}_{n+1}^N f \mid \beta_n) &= \pi_n^N f = \overline{m}_n^N(g_{n/n-1}\ f)/\overline{m}_n^N(g_{n/n-1}).
\end{aligned}$$

Then

$$\tilde{\mathsf{E}}_{[Y]}(U_n^N\ \underline{m}_{n+1}^N f) = \tilde{\mathsf{E}}_{[Y]}\big(U_{n-1}^N\ \overline{m}_n^N(g_{n/n-1} f_n^{(n)})\big).$$

Now, using backward induction in $1 \leq p \leq n-1$ and the recursive formulas described in Section 4.2, we have

$$\begin{aligned}
\tilde{\mathsf{E}}_{[Y]}\big(U_p^N\ \overline{m}_{p+1}^N(g_{n/p}\ f_n^{(p+1)})\big) &= \tilde{\mathsf{E}}_{[Y]}\big(U_p^N\ \tilde{\mathsf{E}}_{[y]}\big(\overline{m}_{p+1}^N(g_{n/p}\ f_n^{(p+1)}) \mid \beta_p\big)\big)\\
\tilde{\mathsf{E}}_{[Y]}\big(\overline{m}_{p+1}^N(g_{n/p}\ f_n^{(p+1)}) \mid \beta_p\big) &= \overline{m}_p^N(g_{n/p-1}\ f_n^{(p)})/\overline{m}_{n-1}^N(g_{p/p-1}).
\end{aligned}$$

Hence

$$\tilde{\mathsf{E}}_{[Y]}\big(U_n^N\ \underline{m}_{n+1}^N f\big) = \tilde{\mathsf{E}}_{[Y]}\big(U_{p-1}^N\ \overline{m}_p^N(g_{n/p-1}\ f_n^{(p)})\big).$$

The result finally follows from the recursive formulas described in Section 4.2 and the fact that

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

□

5.2. Mean square estimates

To prove the convergence (3.4), it clearly suffices to prove that

$$\lim_{N\to+\infty} \tilde{\mathsf{E}}_{[Y]}\big((U_n^N - 1)^2\big) = 0 = \lim_{N\to+\infty} \tilde{\mathsf{E}}_{[Y]}\Big(\Big(\Big(\frac{1}{N}\sum_{i=1}^N f(\widehat{x}_n^i) - \pi_n f\Big)U_n^N\Big)^2\Big). \tag{5.8}$$

The main purpose of this section is to provide a way for estimating the rate of convergence of (5.8) in terms of the log-likelihood functions $V_{n/p}$. Such computations are at the heart of the development. They will drastically simplify the evaluation of the convergence rates discussed in the last section.
We first quote the following result.

Proposition 5.1. *For every series of observations $y \in \mathbf{R}^{\mathrm{N}}$,*

1. *for all* $0 \leq p < n$,

$$V_{n/p}(y) \leq V_{n/n-1}(y) + V_{n-1/n-2}(y) + \cdots + V_{p+1/p}(y) \leq V_n(y);$$

2. *for every* $0 \leq k \leq n$ *and* $0 \leq p_0 < \cdots < p_{k-1} < p_k = n$*, we have*

$$\sum_{l=1}^{k} V_{p_l/p_{l-1}}(y) \leq V_n(y).$$

Proof. The proof is a straightforward computation using that

$$g_{n/p} = K(g_{n/p+1}) \ g_{p+1/p}$$

implies

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

□

Next we want to develop explicit formulas for expressing the effect of the population size and the observation likelihood functions on the mean square errors (5.8). The following propositions will be used repeatedly in the last section.

Proposition 5.2. *For every* $N \geq 1$ *and* $n \geq 0$ *we have* $\tilde{\mathsf{P}}$*-a.e.*

$$\begin{aligned} 1. \quad \tilde{\mathsf{E}}_{[Y]}\big((U_n^N - 1)^2\big) \quad &\leq \quad \sum_{k=1}^{n} \frac{1}{N^k}\Big(1 - \frac{1}{N}\Big)^{n-k} \\ &\quad\quad \times \sum_{0 \leq p_0 < \cdots < p_k = n} \exp\Big(\sum_{l=1}^{k} V_{p_l/p_{l-1}}(Y)\Big), \\ 2. \quad \tilde{\mathsf{E}}_{[Y]}\big((U_n^N - 1)^2\big) \quad &\leq \quad \Big(1 - \Big(1 - \frac{1}{N}\Big)^n\Big) \, e^{V_n(Y)}. \end{aligned}$$

Proof. Inequality 2 is clearly a consequence of inequality 1 and Proposition 5.1. Let us prove inequality 1. Using Lemma 5.3

$$\tilde{\mathsf{E}}_{[Y]}\big((U_n^N - 1)^2\big) = \tilde{\mathsf{E}}_{[Y]}\big((U_n^N)^2\big) - 1.$$

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

$$\tilde{\mathsf{E}}_{[Y]}\big((U_n^N)^2\big) \leq \sum_{k=0}^{n} \frac{1}{N^k}\Big(1 - \frac{1}{N}\Big)^{n-k} \sum_{0 \leq p_0 < \cdots < p_k = n} \exp\Big(\sum_{l=1}^{k} V_{p_l/p_{l-1}}(Y)\Big). \quad (5.9)$$

The proof of (5.9) is based on backward and forward induction in n and maximisation techniques. Using Lemma 5.1 and the recursion in Lemma 5.2, we have

$$\begin{aligned}
\tilde{\mathsf{E}}_{[Y]}\big((U_n^N)^2\big) &= \tilde{\mathsf{E}}_{[Y]}\big(\tilde{\mathsf{E}}_{[Y]}\left(\overline{m}_n^N(g_{n/n-1})^2 \mid \beta_{n-1}\right)(U_{n-1}^N)^2\big)\\
\tilde{\mathsf{E}}_{[Y]}(\overline{m}_n^N(g_{n/n-1})^2 \mid \beta_{n-1}) &\leq \frac{1}{N}\pi_{n-1}^N K(g_{n/n-1}^2) + \Big(1-\frac{1}{N}\Big)(\pi_{n-1}^N K g_{n/n-1})^2\\
&\leq \frac{1}{N}e^{V_{n/n-1}(Y)} + \Big(1-\frac{1}{N}\Big)\Big(\frac{\overline{m}_{n-1}^N(g_{n/n-2})}{\overline{m}_{n-1}^N(g_{n-1/n-2})}\Big)^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\tilde{\mathsf{E}}_{[Y]}\left((U_n^N)^2\right) &\leq \frac{1}{N}\, e^{V_{n/n-1}(Y)}\tilde{\mathsf{E}}_{[Y]}\left((U_{n-1}^N)^2\right)\\
&+\Big(1-\frac{1}{N}\Big)\tilde{\mathsf{E}}_{[Y]}\left(\tilde{\mathsf{E}}_{[Y]}\left(\overline{m}_{n-1}^N(g_{n/n-2})^2 \mid \beta_{n-2}\right)(U_{n-2}^N)^2\right).
\end{aligned}$$

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

$$\begin{aligned}
\tilde{\mathsf{E}}_{[Y]}\left(\overline{m}_{p+1}^N(g_{n/p})^2 \mid \beta_p\right) &\leq \frac{1}{N}\, e^{V_{n/p}(Y)} + \Big(1-\frac{1}{N}\Big)\Big(\frac{\overline{m}_p^N(g_{n/p-1})}{\overline{m}_p^N(g_{p/p-1})}\Big)^2\\
\tilde{\mathsf{E}}_{[Y]}\left(\overline{m}_1^N(g_{n/0})^2 \mid \beta_0\right) &\leq \frac{1}{N}\, e^{V_{n/0}(Y)} + \Big(1-\frac{1}{N}\Big)\left(\nu K g_{n/0}\right)^2.
\end{aligned}$$

Cascading in the above expression

$$\tilde{\mathsf{E}}_{[Y]}\left((U_n^N)^2\right) \leq \Big(1-\frac{1}{N}\Big)^n + \frac{1}{N}\sum_{q=0}^{n-1}\Big(1-\frac{1}{N}\Big)^{n-1-q} e^{V_{n/q}(Y)}\tilde{\mathsf{E}}_{[Y]}\left((U_q^N)^2\right).$$

Suppose the inequalities (5.9) have been proved for every $q \leq n-1$, that is

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

then

$$\begin{aligned}
\tilde{\mathsf{E}}_{[Y]}\left((U_n^N)^2\right) \;&\leq\; \Big(1-\frac{1}{N}\Big)^n + \sum_{k=0}^{n-1}\Big(1-\frac{1}{N}\Big)^{n-(1+k)}\\
&\times\frac{1}{N^{k+1}}\sum_{0\leq p_0<\cdots<p_k<n}\exp\Big(V_{n/p_k}(Y) + \sum_{l=1}^{k}V_{p_l/p_{l-1}}(Y)\Big),
\end{aligned}$$

and the result follows. □

The same techniques then establish the following result.

Proposition 5.3. *For any bounded Borel test function $f : S \to \mathbf{R}$, $N \geq 1$ and $n \geq 0$, we have $\tilde{\mathsf{P}}$-a.e.*

$$\tilde{\mathsf{E}}_{[Y]}\Big(\Big(\Big(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big)U_n^N\Big)^2\Big) \leq 4\|f\|_\infty^2 \sum_{k=1}^{n+1} \frac{1}{N^k}\Big(1-\frac{1}{N}\Big)^{(n+1)-k} \times \sum_{0\leq p_0 < \cdots < p_{k-1} \leq p_k = n} \exp\Big(\sum_{l=1}^{k} V_{p_l/p_{l-1}}(Y)\Big). \quad (5.10)$$

Therefore

$$\tilde{\mathsf{E}}_{[Y]}\Big(\Big(\Big(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big)U_n^N\Big)^2\Big) \leq 4\|f\|_\infty^2\Big(1-\Big(1-\frac{1}{N}\Big)^{n+1}\Big)\ \exp\left(V_n(Y)\right). \quad (5.11)$$

Proof. Let us prove the first statement. For any bounded Borel test function $f : S \to \mathbf{R}$, using the recursive formulas described in Section 4.2, we have

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

Then

$$\tilde{\mathsf{E}}_{[Y]}\left(((\underline{m}_{n+1}^N f)U_n^N)^2\right) \leq \frac{1}{N}\|f\|_\infty^2\,\tilde{\mathsf{E}}_{[Y]}\left((U_n^N)^2\right) + \Big(1-\frac{1}{N}\Big)\ \tilde{\mathsf{E}}_{[Y]}\left((\overline{m}_n^N(g_{n/n-1}f)U_{n-1}^N)^2\right). \quad (5.12)$$

Arguing as above, for every $1 \leq p \leq n-1$ we get

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

Using the above inequalities with expression (5.12), we conclude that

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

with the convention $V_{n/n}=0$. Then, using formula (5.9) we easily obtain

$$
\begin{aligned}
&\tilde{\mathsf{E}}_{[Y]}\big((\underline{m}_{n+1}^N(f-\pi_n f)\; U_n^N)^2\big)\\
&\quad\leq\;\; \frac{4\|f\|_\infty^2}{N}\sum_{l=0}^{n}\frac{1}{N^l}\Big(1-\frac{1}{N}\Big)^{n-l}\sum_{0\leq p_0<\cdots<p_l\leq p_{l+1}=n}\exp\Big(\sum_{s=1}^{l+1}V_{p_l/p_{l-1}}(Y)\Big)\\
&\quad\leq\;\; 4\|f\|_\infty^2\sum_{l=1}^{n+1}\frac{1}{N^l}\Big(1-\frac{1}{N}\Big)^{(n+1)-l}\sum_{0\leq p_0<\cdots<p_{l-1}\leq p_l=n}\exp\Big(\sum_{s=1}^{l}V_{p_l/p_{l-1}}(Y)\Big).
\end{aligned}
$$

The second statement follows from the first inequality in Proposition 5.1 and the fact that $\binom{n}{l}\leq\binom{n+1}{l}$. $\square$

6. Convergence theorems

We are now ready to prove the convergence of the interacting particle approximation described in Section 3. The following theorem is our main result, and it states the relevant consequences of the mean square error estimates stated in Propositions 5.2 and 5.3.

Theorem 6.1. *For every $N\geq 1$, $n\geq 0$, any bounded Borel test function $f:S\to\mathbf{R}$, $T>0$, and $0<\varepsilon<1/2$, we have*

$$
\begin{aligned}
&\sup_{n\in[0,T]}\tilde{\mathsf{P}}_{[Y]}\Big\{\Big|\frac{1}{N}\sum_{i=1}^{N}f(\widehat{x}_n^i)-\pi_n f\Big|>\varepsilon\Big\}\\
&\qquad\qquad\leq\;\; \frac{A(T,f,Y)}{\varepsilon^4}\left(1-\Big(1-\frac{1}{N}\Big)^{T+1}\right)\qquad \tilde{\mathsf{P}}\text{-a.e.,}
\end{aligned}
\tag{6.1}
$$

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

$$
\begin{aligned}
&\sup_{n\in[0,T]}\tilde{\mathsf{E}}_{[Y]}\Big(\Big|\frac{1}{N}\sum_{i=1}^{N}f(\widehat{x}_n^i)-\pi_n f\Big|\Big)\\
&\qquad\qquad\leq\;\; B(T,f,Y)\Big(1-\Big(1-\frac{1}{N}\Big)^{T+1}\Big)^{1/2}\qquad \tilde{\mathsf{P}}\text{-a.e.,}
\end{aligned}
\tag{6.2}
$$

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

Proof. Using Propositions 5.2 and 5.3 and the Cauchy−Schwarz inequality, we have

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

and the inequality (6.2) follows. To prove (6.1), write

$$A = \{U_n^N \geq 1-\varepsilon\} \qquad B = \Big\{\Big|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big| > \varepsilon\Big\}.$$

Using Proposition 5.2 and the fact that

$$|U_n^N - 1| \leq \varepsilon, \qquad \text{for all } 0<\varepsilon<1$$

implies

$$U_n^N \geq 1-\varepsilon,$$

we have

$$\begin{aligned}
\tilde{\mathsf{P}}_{[Y]}\{U_n^N \geq 1-\varepsilon\} &\geq \tilde{\mathsf{P}}_{[Y]}\{|U_n^N - 1| \leq \varepsilon\} \\
&\geq 1 - \frac{1}{\varepsilon^2}\tilde{\mathsf{E}}_{[Y]}\left((U_n^N-1)^2\right) \\
&\geq 1 - \frac{1}{\varepsilon^2}\Big(1-\Big(1-\frac{1}{N}\Big)^{n}\Big)\, e^{V_n(Y)}.
\end{aligned}$$

Therefore,

$$\tilde{\mathsf{P}}_{[Y]}\{\overline{A}\} \leq \frac{1}{\varepsilon^2}\Big(1-\Big(1-\frac{1}{N}\Big)^{n+1}\Big)\, e^{V_n(Y)}.$$

On the other hand, proposition 5.3 gives the inequalities

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

Finally, the inequality

$$\tilde{\mathsf{P}}_{[Y]}\{B\} \le \tilde{\mathsf{P}}_{[Y]}\{B \cap A\} + \tilde{\mathsf{P}}_{[Y]}\{\overline{A}\} \tag{6.3}$$

and the fact that $0 < \varepsilon < 1/2$ implies $\varepsilon < 1 - \varepsilon$, complete the proof. □

Corollary 6.1. *For every $N \ge 2$, $n \ge 0$, for any integrable Borel test function $f : S \to \mathbf{R}$, $T > 0$, and $0 < \varepsilon < 1/2$, we have*

$$\sup_{n\in[0,T]} \tilde{\mathsf{P}}_{[Y]}\Big\{\Big|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big| > \varepsilon\Big\} \le \frac{A(\varepsilon, T, f, Y)}{N}, \qquad \tilde{\mathsf{P}}\text{-a.e.}, \tag{6.4}$$

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

Proof. From the inequality $e^y \ge 1 + y$, it follows that

$$1 - \Big(1 - \frac{1}{N}\Big)^{n+1} \le -(n+1)\log\Big(1 - \frac{1}{N}\Big) = (n+1)\log\Big(\frac{N}{N-1}\Big)$$

and that $\log x \le x - 1$. Thus, we conclude that

$$1 - \Big(1 - \frac{1}{N}\Big)^{n+1} \le \frac{n+1}{N-1} \le \frac{2(n+1)}{N} \le \frac{4\,n}{N}$$

for all $N \ge 2$ and for all $n \ge 1$. This completes the proof. □

Theorem 6.2. *Let $Y = y \in \mathbf{R}^{\mathrm{N}}$ be a series of observations such that $V_n(y) < +\infty$. For every bounded Borel test function $f : S \to \mathbf{R}$ $N \ge 1$ and $n \ge 0$, we have*

$$\text{for all } p > 0, \qquad \sup_{n\in[0,T]} \tilde{\mathsf{E}}_{[y]}\Big(\Big(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big)^p\Big) \underset{N\to+\infty}{\longrightarrow} 0. \tag{6.5}$$

Proof. Using the inequality

$$\begin{aligned}\text{for all } \varepsilon > 0, \qquad & \tilde{\mathsf{E}}_{[y]}\Big(\Big(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big)^p\Big) \\ & \le \varepsilon^p + (2\,\|f\|_\infty)^p\, \tilde{\mathsf{P}}_{[y]}\Big\{\Big|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big| > \varepsilon\Big\}\end{aligned}$$

and Corollary 6.1, the result follows. □

Consider the following assumption on the observation process:

$$(\mathcal{V}) \qquad \text{for all } n \ge 0 \quad \tilde{\mathsf{E}}\,(V_n(Y)) \stackrel{\text{def}}{=} \alpha(n) < +\infty. \tag{6.6}$$

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

Theorem 6.3. *Assume $(\mathcal{V})$ is satisfied. For every bounded Borel test function $f : S \to \mathbf{R}$ $N \geq 1$, $n \geq 0$, $0 < \varepsilon < 1/2$ and $M > 0$, we have*

$$\sup_{n\in[0,T]} \tilde{\mathsf{P}}\Big\{\Big|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big| > \varepsilon\Big\} \leq \frac{1}{M} + \frac{A(T,M,f,\varepsilon)}{N}, \tag{6.7}$$

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

Proof. Using Theorem 6.1

$$\begin{aligned}&\tilde{\mathsf{P}}\Big\{\Big|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big| \leq \varepsilon\Big\}\\ &\quad\geq\ \tilde{\mathsf{P}}\big(V_n(Y) \leq M\alpha(n)\big) - \frac{A(f)\ e^{M\alpha(n)}}{\varepsilon^4}\Big(1 - \Big(1 - \frac{1}{N}\Big)^{n+1}\Big),\end{aligned}$$

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

$$\begin{aligned}&\tilde{\mathsf{P}}\Big\{\Big|\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big| \leq \varepsilon\Big\}\\ &\quad\geq\ 1 - \Big(\frac{1}{M} + \frac{A(f)\ e^{M\alpha(n)}}{\varepsilon^4}\ \Big(1 - \Big(1 - \frac{1}{N}\Big)^{n+1}\Big)\Big).\end{aligned}$$

The arguments used in the proof of Corollary 6.1 complete the proof. □

Finally, the following corollary is derived by the same reasoning as in Theorem 6.2.

Corollary 6.2. *Assume $\mathcal{V}$ is satisfied. For every bounded Borel test function $f : S \to \mathbf{R}$, $N \geq 1$ and $n \geq 0$, we have*

$$\text{for all } p > 0, \qquad \sup_{n\in[0,T]} \tilde{\mathsf{E}}\Big(\Big(\frac{1}{N}\sum_{i=1}^{N} f(\widehat{x}_n^i) - \pi_n f\Big)^p\Big) \underset{N\to+\infty}{\longrightarrow} 0. \tag{6.8}$$

Acknowledgements

I am deeply grateful to Laure Coutin and Pierre Vandekerkhove for their careful reading of the original manuscript. I also thank the anonymous referee for his comments and suggestions that greatly improved the presentation of the paper.

References

[1] H. Carvalho, P. Del Moral, A. Monin and G. Salut (1996) Optimal nonlinear filtering in GPS/INS integration. To appear in *IEEE Trans. Aerospace and Electron. Syst.*

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

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

[4] P. Del Moral, G. Rigal, J.C. Noyer and G. Salut (1993) Traitement non-linéaire du signal par reseau particulaire: application radar. Quatorzième colloque GRETSI, Juan les Pins, 13–16 Sept. 1993.

[5] P. Del Moral (1994) Résolution particulaire des problèmes d'estimation et d'optimisation non-linéaires. Thèse de l'Université Paul Sabatier, Rapport LAAS N 94269, Juin 1994.

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

[7] P. Del Moral (1995) Non linear filtering using random particles. *Probab. Theory Appl.* **40** (4).

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

[9] P. Del Moral and G. Salut (1996) Particle interpretation of nonlinear filtering and optimization. *Russian Journal of Mathematical Physics.* In press.

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

[11] J.H. Holland (1975) *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor.

[12] H. Kunita (1971) Asymptotic behavior of nonlinear filtering errors of Markov processes. *Journal of Multivariate Analysis* **1** (4), 365–393.

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