Interacting Particle Systems Approximations of the Kushner Stratonovitch Equation

D. Crişan*, P. Del Moral†, T.J. Lyons*

Abstract

In this paper we consider the continuous time filtering problem and we estimate the order of convergence of an interacting particle system scheme presented by the authors in previous works. We will discuss how the discrete time approximating model of the Kushner-Stratonovitch equation and the genetic type interacting particle system approximation combine. We present quenched error bounds as well as mean order converge results.

Keywords: Non Linear Filtering, Filter approximation, error bounds, interacting particle systems, genetic algorithms.

Code A.M.S: 60G35,93E11, 62L20.

1 Introduction

1.1 Background and Motivations

The aim of this work is the design of an interacting particle system approach for the numerical solving of the continuous time non linear filtering problem.
Recently intense interest has been aroused in the non linear filtering theory community concerning the connections between non linear estimation, measure valued processes and branching and interacting particle systems. The evolution of this material may be seen quite directly through the list of referenced papers.

Let us briefly survey some different approaches and motivate our work.

*Imperial College, Department of Mathematics, Huxley Building, 180 Queen's Gate, London SW7 2BZ, Work partly supported by CEC Contract No ERB-FMRX-CT96-0075

†LSP-CNRS UMR C55830, Bat.1R1, Univ. Paul Sabatier, 31062 Toulouse, delmoral@cict.fr..

In [17, 18], [19] and [23] the authors proposed a preliminary time discretization scheme of the optimal filter evolution. Then they solved the discrete time approximating model by using Monte-Carlo simulations or spatial quantizations of the signal. In these works special attention was paid to study the rate of convergence of the time discretization scheme to the continuous time original model. The Monte-Carlo particle approach described therein consists of independent particles weighted by exponentials. Unfortunately the above Monte Carlo approximation is not efficient mainly because the particles are independent of each other and the growth of the exponential weights are difficult to control as time goes on.
In [5, 8] a way was proposed to regularize these exponential weights and a natural ergodic assumption was introduced on the signal semigroup under which the particle approximation converges in law to the optimal filter uniformly with respect to time.
All the above Monte Carlo approximations are the crudest of the random particle system approaches.

It has recently been emphasized that a more efficient approach is to use interacting and branching particle systems to numerically solve the filtering problem.
In [1, 2] a branching particle system approximation for the filtering problem of diffusions is studied. Although these works give some insight into the connections between branching particle systems and non linear filtering, they entirely rely on two assumptions:
First it is assumed that the continuous semigroup of the signal is explicitly known in the sense that we can exactly simulate random transitions according to this semigroup. Further it is assumed that the stochastic integrals arising in the Girsanov exponentials are readily accessible without further work.
It can be argued that in practice the signal semigroup and the above stochastic integrals are not known exactly and another level of approximation is therefore needed.
In [6, 7, 9, 10]and [14] different interacting particle systems approximations for the discrete time filtering problem are studied. Here again it is assumed that we can exactly simulate random transitions according to the semigroup of the signal and the Girsanov exponentials are exactly known. Even if these particle methods provide essential insight for solving quite general discrete time non linear filtering problem they do not applied directly to the continuous time case.

The aim of this work is to extend the genetic type interacting particle system approach introduced in [6, 7, 9] and [14] to handle continuous time filtering problems. A crucial practical advantage of our approximating model is that it does not involves stochastic integrals and the transition probability kernel which govern our algorithm is chosen so that we can exactly simulate its random transitions. Our construction is also explicit, the recursions have a simple form and they can easily be simulated on a computer.

In contrast to [1, 2] we use a preliminary discrete time and measure valued approximating model for the Kushner-Stratonovitch equation. This time discretization scheme was introduced by Picard in [23]. The study of time discretization approximating models is still a active research area. The most accurate work in this subject seems to be [18]. In this work the authors extend their study to the time discretization problem of non linear filtering of signals driven by not necessarily white signals. As pointed out in [18] and [23] the resulting discrete time and measure valued model also characterize the evolution in time of the optimal filter for a suitably defined discrete time filtering problem. It follows that the branching and interacting particle system approaches presented in [6, 7, 9] and [3] can be applied to this discrete time approximating model.

We study how the time discretization and the particle system approximating models combinate and provide explicit quenched and mean error bounds for the global approximating scheme.

The paper has the following structure:
In Section 2 we introduce the continuous time non linear filtering model and the discrete time approximating model under study.
In Section 3 we introduce the interacting particle system approximating model and we study the order of convergence to the solution of the so-called Kushner-Stratonovitch equation. Firstly we present two different approaches to obtain quenched error bounds. Then we apply these strategies to study mean error bounds and the resulting order of convergence for the Fortet-Mourier distance (see for instance [13]).

1.2 Description of the Model and Statement of some Results

To clarify the presentation all the processes considered in this work are indexed on the compact interval $[0,1] \subset \mathbb{R}_+$.
The basic model for the continuous time non linear filtering problem consists of a time homogeneous Markov Process $\{(X_t, Y_t) \,:\, t \in [0,1]\}$ taking values in $\mathbb{R}^p \times \mathbb{R}^q$, $p, q \geq 1$. We assume that $Y_0 = 0$ and X_0 is a random variable with law ν.
The classical filtering problem can be summarized as to find the conditional distribution of the signal X at time t with respect to the observations Y up to time t. Namely

$$\pi_t f = E(f(X_t)/\mathcal{Y}_0^t) \qquad \forall f \in \mathcal{C}_b(\mathbb{R}^p) \tag{1}$$

where $\mathcal{Y}_0^t$ is the filtration generated by the observations Y up to time t.

Here we will only consider the case where the signal X is a diffusion and the observation $\{Y_t \,:\, t \in [0,1]\}$ is solution of the Ito-type differential equation

$$dY_t = h(X_t)\, dt + dV_t \tag{2}$$

where $h : \mathbb{R}^p \to \mathbb{R}^q$ is a bounded continuous function and $\{V_t \,:\, t \in [0,1]\}$ is a standard q-dimensional Brownian motion independent of $\{X_t \,:\, t \in [0,1]\}$.

As mentioned in the introduction our approximating model is obtained by first introducing a time discretization scheme of the basic model. To this end we introduce a sequence a meshes $\{(t_0, \ldots, t_M) \,:\, M \geq 1\}$ given by

$$t_n = \frac{n}{M} \qquad n \in \{0, \ldots, M\}.$$

Then, to obtain a computationally feasible solution we will use the following natural assumptions:

(H1) For any $M \geq 1$ there exists a transition probability kernel $P^{(\mathrm{M})}$ such that

$$\sup_{t\in[0,1]} E\left(|X_{t_n} - X_{t_n}^{(\mathrm{M})}|^2\right) \leq \frac{K}{M} \qquad K < \infty$$

where $\{X_{t_n}^{(\mathrm{M})} \,:\, n = 0, \ldots, M\}$ is the time homogeneous Markov chain with transition probability kernel $P^{(\mathrm{M})}$ and such that $X_0^{(\mathrm{M})} = X_0$.

(H2) We can exactly simulate random variables according to the law $P^{(\mathrm{M})}(x, \bullet)$ for any $x \in \mathbb{R}^p$.

The interacting particle system will by a Markov chain $\{\xi_{t_n} \ : \ n = 0, \ldots, M\}$ with product state space $(\mathbb{R}^p)^N$, where $N \geq 1$ is the size of the system. The points of the set $(\mathbb{R}^p)^N$ are called particle systems and when there is no confusion they will be denoted by the letters x and z.
Our approximating model is then described by

$$\begin{aligned} P_Y(\xi_{t_0} \in dx) &= \prod_{p=1}^{N} \nu(dx^p) \\ P_Y(\xi_{t_n} \in dx / \xi_{t,n-1} = z) &= \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_{t_n}^M(\Delta Y_{t_n}, z^i)}{\sum_{j=1}^N g_{t_n}^M(\Delta Y_{t_n}, z^j)} \, P^{(\mathrm{M})}(z^i, dx^p) \end{aligned}$$

where

- $dx = dx^1 \times \ldots \times dx^N$ is an infinitesimal neighbourhood of the point $x = (x^1, \ldots, x^N) \in (\mathbb{R}^p)^N$
- $\Delta Y_{t_n} = Y_{t_n} - Y_{t_{n-1}}$
- The weight functions $g_{t_n}^M$ are given by

$$g_{t_n}^M(y, x) = \exp\left(-\frac{|h(z)|^2}{2M} + h(z)^\star y \right) \qquad \forall (y, x)\, \mathbb{R}^q \times \mathbb{R}^p.$$

Let us denote by $\{\pi_t^{M,N} \ : \ t \in [0,1]\}$ the empirical measures associated with $\{\xi_{t_n} \ : \ n = 0, \ldots, M\}$ and given by

$$\pi_t^{M,N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{t_n}^i} \qquad \forall t \in [t_n, t_{n+1}) \qquad \forall n \in \{0, \ldots, M-1\}. \tag{3}$$

Under some natural regularity conditions the particle density profiles (3) will converge to the optimal filter. The aim of the paper is to discuss the order of convergence when the time discretization parameter $M = M(N)$ depends on the size N of the system.
It is now convenient to introduce some notation. We will denote by $\mathbf{M}_1(\mathbb{R}^p)$ the set of all probability measures on $\mathbb{R}^p$ furnished with the weak topology. We recall that the weak topology is generated by bounded continuous functions and we write $\mathcal{C}_b(\mathbb{R}^p)$ for the space of these functions with the supremum norm

$$\|f\| = \sup_{x \in \mathbb{R}^p} |f(x)|.$$

The main purpose of the paper is to prove the following theorem.

Theorem 1.1 *Under some regularity conditions for any bounded test function f we have*

$$M(N) \leq \sqrt{N} \Longrightarrow \sup_{t\in[0,1)} E(|\pi_t f - \pi_t^{M(N),N} f|) \leq \frac{\mathrm{Cte(f)}}{\mathrm{N}^{1/4}}. \tag{4}$$

When the test function f is sufficiently regular and if $p = q = 1$ then we have

$$M(N) \leq N^{1/3} \Longrightarrow \sup_{t\in[0,1)} E(|\pi_t f - \pi_t^{M(N),N} f|) \leq \frac{\mathrm{Cte(f)}}{\mathrm{N}^{1/3}}. \tag{5}$$

2 Discrete time approximating model for the Kushner-Stratonovitch equation

2.1 Formulation of the Non Linear Filtering Problem

The basic model for the continuous time filtering problem consists of an $\mathbb{R}^p \times \mathbb{R}^q$-valued Markov process $\{(X_t, Y_t) \; : \; t \in [0,1]\}$ strong solution on a probability space (Ω, F, P) of the Ito-type stochastic differential equations

$$\left\{ \begin{array}{lcl} dX_t & = & a(X_t)dt + b(X_t)d\beta_t \\ dY_t & = & h(X_t)dt + dV_t \end{array} \right.$$

where

1. $a : \mathbb{R}^p \to \mathbb{R}^p$, $b : \mathbb{R}^p \to \mathbb{R}^{p\times m}$ and $h : \mathbb{R}^p \to \mathbb{R}^q$ are bounded and Lipschitz continuous functions.
2. $\{(\beta_t, V_t) \; : \; t \in [0,1]\}$ is a $(\mathbb{R}^m \times \mathbb{R}^q)$-valued standard Brownian motion.
3. $Y_0 = 0$ and X_0 is a random variable independent of $\{(\beta_t, V_t) \; : \; t \in [0,1]\}$ with law ν so that $E(|X_0|^2) < \infty$.

The problem of assessing the conditional expectations is of course related to that of recursively computing the random distributions $\{\pi_t \; : \; t \in [0,1]\}$ which provide all statistical information about the state variables $\{X_t \; : \; t \in [0,1]\}$ obtainable from the observations. The first step in this direction consists of obtaining a more tractable description of the conditional expectations .
Introducing Z_t such that

$$\log Z_t = \int_0^t h^\star(X_s)\, dY_s - \frac{1}{2}\int_0^t |h(X_s)|^2\, ds \qquad t \in [0,1],$$

it is well known that the orignal probability measure P is equivalent to a so called *reference probability measure* P_0 given by

$$P = Z_1\, P_0.$$

In addition, under P_0, $\{(\beta_t, Y_t) \ : \ t \in [0,1]\}$ is a $(I\!R^m \times I\!R^q)$-valued standard Brownian motion and, X_0 is a random variable with law ν, independent of (β, Y). The following well known result gives a functional integral representation for the conditional expectations (2.1), which is known as the Kallianpur-Striebel formula:

$$\pi_t f = \frac{\rho_t(f)}{\rho_t(1)} \qquad \forall f \in \mathcal{C}_b(I\!R^p) \quad \forall t \in [0,1] \tag{6}$$

with $\rho_t(f) = E_0(f(X_t)\ Z_t/\mathcal{Y}_0^t) = E_0^Y(f(X_t)\ Z_t)$.
We use $E_0^Y(\centerdot)$ to denote the integration of the Brownian paths $\{\beta_t \ : \ t \in [0,1]\}$ and the variable X_0.
Let us recall the connections between continuous time non linear filtering problems and measure valued Markov processes.
The so-called un-normalized $\{\rho_t \ : \ t \in [0,1]\}$ can also be defined as the measure valued solution of the Zakai equation

$$\rho_t(f) = \nu(f) + \int_0^t \rho_s(Lf)\ ds + \int_0^t \rho_s(h^\star f)\ dY_s$$

for all f in the domain $D(L)$ of the infinitesimal Kolmogorov differential operator L associated with the diffusion $\{X_t \ : \ t \in [0,1]\}$.

Using the above, one can prove that $\{\pi_t \ : \ t \in [0,1]\}$ is therefore a $(\widetilde{Y}, P)$-Markov process taking values in the space of all probability measures on $I\!R^p$ and solution of the Kushner-Stratonovitch equation

$$\pi_t(f) = \nu(f) + \int_0^t \pi_s(Lf)\ ds + \int_0^t (\pi_s(h^\star f) - \pi_s(h^\star)\pi_s(f))\ (dY_s - \pi_s(h)ds)$$

for all $f \in D(L)$ and $t \in [0,1]$.

2.2 Time Discretization Scheme

As pointed out in the introduction our approximating model is obtained by first approximating the original model by a discrete time and measure valued process. The treatment that follows is standard and it is essentially contained in [17, 18] and [22].
In view of (6) the optimal filters $\{\pi_{t_n} \; : \; n = 0, \ldots, M\}$ can be written as

$$\pi_{t_n} f = \frac{E_0^Y (Z_{t_{n-1}} \; (H_{t_n} f)(X_{t_{n-1}}))}{E_0^Y (Z_{t_{n-1}} \; (H_{t_n} 1)(X_{t_{n-1}}))}$$

where H_{t_n} is the finite transition measure on $\mathbb{R}^p$ given by

$$\begin{aligned} H_{t_n} f(x) &\stackrel{\text{def}}{=} \int H_{t_n}(x, dz) \; f(z) = E_0^Y (f(X_{t_n}) \; g_{t_n}(X, Y) / X_{t_{n-1}} = x) \\ \log g_{t_n}(X, Y) &= \int_{t_{n-1}}^{t_n} h^\star(X_s) \, dY_s - \frac{1}{2} \int_{t_{n-1}}^{t_n} |h(X_s)|^2 \, ds. \end{aligned} \tag{7}$$

If, for any transition measure H and any probability measure π on $\mathbb{R}^p$ we denote by πH the finite measure so that for any bounded continuous function $f \in \mathcal{C}_b(\mathbb{R}^p)$,

$$\pi H(f) = \int \pi(dx) \; (Hf)(x),$$

then, given the observations, the dynamics structure of the conditional distributions $\{\pi_{t_n} \; : \; n = 0, \ldots, M\}$ is defined by the recursion

$$\pi_{t_n}(f) = \frac{\pi_{t_{n-1}} H_{t_n}(f)}{\pi_{t_{n-1}} H_{t_n}(1)} \qquad n = 1, \ldots, M \qquad \text{with} \quad \pi_{t_0} = \nu.$$

Solving the above dynamical system is in general an enormous task as it is non linear, involves integrations over the whole state space $\mathbb{R}^p$ and stochastic integrations.

2.2.1 Time discretization of the stochastic integral

To approximate the stochastic integrals (7) it is convenient to note that, in a sense to be given,

$$\log g_{t_n}(X, Y) \underset{\Delta t_n \sim 0}{\sim} h^\star(x) \, \Delta Y_{t_n} - \frac{1}{2} |h(x)|^2 \, \Delta t_n \quad \text{with} \quad \Delta t_n = t_n - t_{n-1}.$$

In this connection, a first step to obtain a computationally feasible solution consists in replacing H_{t_n} by the approximating multiplication operator

$$(H_{t_n}^M f)(x) = g_{t_n}^M(Y, x) \; (P_{\frac{1}{M}} f)(x)$$

where $\{P_t \ : \ t \in [0,1]\}$ is the continuous semigroup of the signal X and for any observation $g_{t_n}^M(\Delta Y_{t_n}, \bullet) : I\!R^p \to I\!R_+$ is the positive and continuous function given by

$$g_{t_n}^M(\Delta Y_{t_n}, x) = \exp\left(h^\star(x)\,\Delta Y_{t_n} - \frac{1}{2M}|h(x)|^2\right) \quad \text{with} \quad \Delta Y_{t_n} = Y_{t_n} - Y_{t_{n-1}}.$$

Remark 2.1:
The choice of the approximating function $g_{t_n}^M$ given above is not unique. We can also use the functions $\widetilde{g}_{t_n}^M$ given by

$$\widetilde{g}_{t_n}^M(\Delta Y_{t_n}, x) = 1 + h^\star(x)\Delta Y_{t_n} + \frac{1}{2}|h(x)|^2\left(|\Delta Y_{t_n}|^2 - \frac{1}{M}\right).$$

The function h being bounded we can choose M large enought so that

$$\|h\| < \sqrt{M} \qquad \text{and} \qquad g_{t_n}^M(\Delta Y_{t_n}, x) > 0 \qquad \forall x \in I\!R^p.$$

From now on we denote by $\{\widetilde{\pi}_{t_n} \ : \ n = 0, \ldots, M\}$ the solution of the resulting approximating discrete time model

$$\left\{\begin{array}{lcll} \widetilde{\pi}_{t_n} & = & \Psi_{t_n}^M(\Delta Y_{t_n}, \widetilde{\pi}_{t_{n-1}})P_{\frac{1}{M}} & ; n = 1, \ldots, M \\ \widetilde{\pi}_0 & = & \nu & \end{array}\right. \tag{8}$$

where for any probability measure π on $I\!R^p$ and for any $f \in \mathcal{C}_b(I\!R^p)$

$$\Psi_{t_n}^M(\Delta Y_{t_n}, \pi)f = \frac{\displaystyle\int f(x)\, g_{t_n}^M(\Delta Y_{t_n}, x)\, \pi(dx)}{\displaystyle\int g_{t_n}^M(\Delta Y_{t_n}, x)\, \pi(dx)}.$$

Elementary manipulations show that the solution of the latter system is given by the formula

$$\widetilde{\pi}_{t_n} f = \frac{\displaystyle\int f(x_n) \prod_{k=1}^n g_{t_k}^M(\Delta Y_{t_n}, x_{k-1}) \prod_{k=1}^n P_{\frac{1}{M}}(x_{k-1}, dx_k)\, \nu(dx_0)}{\displaystyle\int \prod_{k=1}^n g_{t_k}^M(\Delta Y_{t_n}, x_{k-1}) \prod_{k=1}^n P_{\frac{1}{M}}(x_{k-1}, dx_k)\, \nu(dx_0)}. \tag{9}$$

Remark 2.2:
It appears from the above that an equivalent way to compute recursively in time the multiple integrals arising in (12) consists of solving the discrete time and measure

valued dynamical system (8).

Let us also remark that each transition $\widetilde{\pi}_{t_{n-1}} \leadsto \widetilde{\pi}_{t_n}$ is decomposed into two separate mechanisms

$$\widetilde{\pi}_{t_{n-1}} \xrightarrow{(1)} \Psi^M_{t_n}(\Delta Y_{t_n}, \widetilde{\pi}_{t_{n-1}}) \xrightarrow{(2)} \Psi^M_{t_n}(\Delta Y_{t_n}, \widetilde{\pi}_{t_{n-1}})P_{\frac{1}{M}}.$$

The first one is usually called the "updating" transition. The transformation $\Psi^M_{t_n}(Y, \bullet)$ is a Bayes' rule which depends on the current observation data ΔY_{t_n}.
The second one is called the "prediction" transition because it does not depend on the observations but on the semigroup of the signal.
As we shall see in the forthcoming development our interacting particle approximating scheme is also decomposed into the same kind of transitions.

A crucial practical advantage of the approximating model (8) is that it does not involves stochastic integrations. We recall that in practical situations the observational data are often sampled with a constant period of a given length $\frac{1}{M}$. In our approximating model the exponential weights $\{g^M_{t_n}(\Delta Y_{t_n}, \bullet) : n = 0, \ldots, M\}$ have an explicit and an analytic expression so that then can readily be computed when the observations $\{Y_{t_n} : n = 0, \ldots, M\}$ are received.
Of course the difficulty in solving (8) lies in computing, recursively in time, a series of integrations over the whole state space $\mathbb{R}^p$. It is therefore tempting to use the interacting particle system approximations introduced in [3] and [7, 9]. Unfortunately we do not know how to simulate random transitions according to the semigroup of the signal and another kind of approximation is therefore needed.

2.2.2 Approximation of the signal semigroup

Under the assumptions (H1) and (H2) we now introduce a final discrete time and measure valued approximating model.
Firstly it is worth noting that an example of an approximating Markov chain $\{X^{(M)}_{t_n} : n = 0, \ldots, M\}$ satisfying these assumptions is given by the classical Euler scheme

$$X^{(M)}_{t_n} = X^{(M)}_{t_{n-1}} + a(X^{(M)}_{t_{n-1}})(t_n - t_{n-1}) + c(X^{(M)}_{t_{n-1}})\ (\beta_{t_n} - \beta_{t_{n-1}}) \qquad n = 1, \ldots, M \quad (10)$$

with $X^{(M)}_0 = X_0$.
This scheme is the crudest of the discretization scheme that can be used in our settings. Other time discretization schemes for diffusive signals are described in full detail in [21] and [25].
Once again we fix the observations and we note finally that $\{\pi^M_{t_n} : n = 0, \ldots, M\}$

the solution of the discrete time and measure valued dynamical system

$$\begin{cases} \pi_{t_n}^M &= \phi_{t_n}^M(\Delta Y_{t_n}, \pi_{t_{n-1}}^M) \qquad ; n = 1, \ldots, M \\ \pi_0^M &= \nu \end{cases} \tag{11}$$

where $\phi_{t_n}^M(\Delta Y_{t_n}, \bullet)$ is defined by the formula

$$\phi_{t_n}^M(\Delta Y_{t_n}, \pi) \stackrel{\text{def}}{=} \Psi_{t_n}^M(\Delta Y_{t_n}, \pi)P^{(M)}$$

for any probability measure π on $I\!R^p$.
Arguing as before, we see that the solution of the above system is given by

$$\pi_{t_n}^M f = \frac{\displaystyle\int f(x_n) \prod_{k=1}^n g_{t_k}^M(\Delta Y_{t_k}, x_{k-1}) \prod_{k=1}^n P^{(M)}(x_{k-1}, dx_k)\ \nu(dx_0)}{\displaystyle\int \prod_{k=1}^n g_{t_k}^M(\Delta Y_{t_k}, x_{k-1}) \prod_{k=1}^n P^{(M)}(x_{k-1}, dx_k)\ \nu(dx_0)}. \tag{12}$$

The error bound caused by the discretization of the time interval $[0, 1]$ and the approximation of the signal semigroup is well understood (see for instance Proposition 5.2 pp. 31 [17], Theorem 2 in [22], Theorem 4.2 in [18] and also Theorem 4.1 in [4]). More precisely we have the well known result.

Theorem 2.3 ([18]) *Let f be a bounded test function on $I\!R^p$ satisfying the Lipschitz condition*

$$|f(x) - f(z)| \leq k(f)\ |x - z|.$$

Then

$$\sup_{t\in[0,1)} E\left(|\pi_t f - \pi_t^M f|\right) \leq \frac{\mathrm{C}}{\sqrt{\mathrm{M}}}\left(\|f\| + k(f)\right) \tag{13}$$

where C is some finite constant.

3 Interacting Particle System Model

3.1 Description of the genetic type particle algorithm

Even if it looks innocent, the dynamical system (11) involves at each step several integrations over the whole state space $I\!R^p$. Therefore it is necessary to find a new strategy to solve the former integrals recursively in time.
Now we introduce a particle system approximating model for the numerical solving of (11). The genetic-type algorithm presented in the forthcoming development was introduced in [6, 9] in discrete time settings and further developed in [7, 10, 11].

Namely, the interacting particle system associated to (11) consists of a Markov chain $\{\xi_{t_n} \ : \ n = 0, \ldots, M\}$ with product state space $(I\!R^p)^N$, where $N \geq 1$ is the size of the system and defined by

$$\begin{aligned} P_Y(\xi_{t_0} \in dx) &= \prod_{p=1}^{N} \nu(dx^p) \\ P_Y(\xi_{t_n} \in dx/\xi_{t_{n-1}} = z) &= \prod_{p=1}^{N} \phi_{t_n}^M(\Delta Y_{t_n}, \frac{1}{N}\sum_{i=1}^{N} \delta_{z^i})(dx^p) \end{aligned}$$

where $dx = dx^1 \times \ldots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in (I\!R^p)^N$.
Let us remark that

$$\phi_{t_n}^M(Y, \frac{1}{N}\sum_{i=1}^{N} \delta_{z^i}) = \left(\sum_{i=1}^{N} \frac{g_{t_n}^M(\Delta Y_{t_n}, z^i)}{\sum_{j=1}^{N} g_{t_n}^M(\Delta Y_{t_n}, z^j)} \delta_{z^i} \right) P^{(\mathrm{M})}$$

and therefore the transitions of the systems may be written

$$P_Y(\xi_{t_n} \in dx/\xi_{t_{n-1}} = z) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_{t_n}^M(\Delta Y_{t_n}, z^i)}{\sum_{j=1}^{N} g_{t_n}^M(\Delta Y_{t_n}, z^j)} \ P^{(\mathrm{M})}(z^i, dx^p).$$

Using the above observations we see that the particles evolve according two separate mechanisms

$$\xi_{t_{n-1}} \xrightarrow{\text{Updating}} \widehat{\xi}_{t_{n-1}} \xrightarrow{\text{Prediction}} \xi_{t_n}.$$

More precisely an equivalent formulation is the following

- Initial Particle System

$$P_Y(\xi_{t_0} \in dx) = \prod_{p=1}^{N} \nu(dx^p)$$

- Updating

$$P_Y(\widehat{\xi}_{t_{n-1}} \in dx/\xi_{t_{n-1}} = z) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_{t_n}^M(\Delta Y_{t_n}, z^i)}{\sum_{j=1}^{N} g_{t_n}^M(\Delta Y_{t_n}, z^j)} \delta_{z^i}(dx^p)$$

- Prediction

$$P_Y(\xi_{t_n} \in dx/\widehat{\xi}_{t_{n-1}} = z) = \prod_{p=1}^{N} P^{(\mathrm{M})}(z^p, dx^p).$$

As pointed out in [6] we see that this particle approximation belongs to the class of algorithms called genetic algorithms which guide natural evolution: the exploration/Mutation is the prediction step and the Selection transition correspond to the updating procedure. These algorithms were introduced by Holland in 1975 [15] to handle global optimization problems on a finite state space.
The situation is different when dealing with non linear filtering problems mainly because there is no temperature parameter which tends to zero and the state space is not finite.
A crucial practical advantage of the genetic type approximating introduced above is that it leads to an empirical measure valued approximating model

$$\pi_{t_{n-1}}^{M,N} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_{t_{n-1}}^i} \xrightarrow{\text{Updating}} \frac{1}{N}\sum_{i=1}^{N} \delta_{\widehat{\xi}_{t_{n-1}}^i} \xrightarrow{\text{Prediction}} \pi_{t_n}^{M,N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_{t_n}^i}.$$

The convergence of the genetic type interacting particle system under study to the solution of the time discretization approximating model (11) has been studied in [6, 7, 9]; quenched large deviations principles are developed in [10] and the associated fluctuations are presented in [11].

3.2 Error bounds and order of convergence

The aim of this section is to study the convergence of the particle density profiles to the solution of the Kushner-Stratonovitch equation as the number of particles $N \to \infty$ and the time discretization parameter $M \to \infty$.
A rough structure is as follows. In Section 3.2.1 we present two different ways to obtain quenched error bounds for the convergence of the genetic type approximating model to the solution of the discrete time and measure valued dynamical system (11) as N is growing. This section is based in general on a previous work of one of the authors [6, 7]. Here we present a simplified proof that captures the main ideas. We also emphasize that our strategy can be used to get explicit quenched error bounds for the more general branching and interacting particle systems presented in [3] or [1, 2].
In Section 3.2.2 we study the convergence of the particle approximating model to the solution of the Kushner-Stratonovitch equation and we propose several orders of convergence depending on the choice of the time dicretization parameter $M = M(N)$ with respect to the number of particles N.
The situation is different in continuous time settings and the quenched error bounds presented in Section 3.2.1 cannot be used directly. Nevertheless we will see that the strategy used to obtain these quenched error bounds can be nested to easily produce the order of convergence.

3.2.1 Quenched error bounds

To motivate our work we begin with a quenched error which can easily be derived from the very definition of the interacting particle system approximating model.

Proposition 3.1 *For any bounded measurable function f on $\mathbb{R}^p$ such that $\|f\| \leq 1$ and for any $M \geq 1$ we have P.a.s.*

$$\sup_{n=0,\ldots,M} E_Y(|\pi_{t_n}^M f - \pi_{t_n}^{M,N} f|) \leq \frac{4M}{\sqrt{N}} C(M,Y) \tag{14}$$

with

$$\log C(M,Y) = \frac{\|h\|^2}{2} + 2\|h\| \sum_{n=1}^{M} |\Delta Y_{t_n}|. \tag{15}$$

Proof:

Let us fix in what follows the observation process Y. In addition, to clarify the presentation, we simplify the notations suppressing, when there is no possible confusions, the observation parameter Y, the index parameters t and M, so that we simply note g_n, ϕ_n, P, π_n and π_n^N instead of $g_{t_n}^M(\Delta Y_{t_n}, \bullet)$, $\phi_{t_n}^M(\Delta Y_{t_n}, \bullet)$, $P^{(M)}$, $\pi_{t_n}^M$ and $\pi_{t_n}^{M,N}$.
To prove the quenched error bound (14) the key idea is to introduce the composite mappings

$$\phi_{n/p} \stackrel{\text{def}}{=} \phi_n \circ \phi_{n-1} \circ \ldots \circ \phi_{p+1} \qquad \forall 0 \leq p \leq n \leq M$$

with the convention $\phi_{n/n} = Id$.
A clear backward induction on the parameter $p (\leq n)$ leads to the formula

$$\phi_{n/p}(\pi) f = \frac{\pi\left(g_{n/p}\,(P_{n/p} f)\right)}{\pi\left(g_{n/p}\right)} \tag{16}$$

where, for any $f \in \mathcal{C}_b(\mathbb{R}^p)$ and $0 \leq p \leq n \leq M$

$$g_{n/p-1} = g_p\, P(g_{n/p}) \qquad P_{n/p-1} f = \frac{P\left(g_{n/p}\,(P_{n/p} f)\right)}{P\left(g_{n/p}\right)} \tag{17}$$

with the conventions $g_{n/n} = 1$ and $P_{n/n} = Id$.
For later use we immediately notice that for any $x \in \mathbb{R}^p$

$$a_n \leq g_n(x) \leq A_n$$

with

$$a_n \stackrel{\text{def}}{=} \exp(-\|h\||\Delta Y_{t_n}| - \frac{\|h\|^2}{2M}) \qquad \text{and} \qquad A_n \stackrel{\text{def}}{=} \exp(\|h\||\Delta Y_{t_n}|).$$

More generally, from the definition of the functions $g_{n/p}$ we note the following inequality for future use:

$$a_{n/p} \leq g_{n/p}(x) \leq A_{n/p} \qquad 0 \leq p \leq n \leq M$$

with

$$\begin{aligned} a_{n/p} &\stackrel{\text{def}}{=} \exp\,(-\|h\| \sum_{q=p+1}^{n} |\Delta Y_{t_q}| - \frac{(n-p)\|h\|^2}{2M}) \\ A_{n/p} &\stackrel{\text{def}}{=} \exp\,(\ \|h\| \sum_{q=p+1}^{n} |\Delta Y_{t_q}|) \end{aligned}$$

Using the above notation we have the decomposition

$$\pi_n^N f - \pi_n f = \sum_{p=0}^{n} \left(\phi_{n/p}(\pi_p^N) f - \phi_{n/p}(\phi_p(\pi_{p-1}^N)) f\right)$$

with the convention $\phi_0(\pi_{-1}^N) = \nu$.
By using the formula (16) we see that each term

$$|\phi_{n/p}(\pi_p^N) f - \phi_{n/p}(\phi_p(\pi_{p-1}^N)) f|$$

is bounded by

$$\frac{A_{n/p}}{a_{n/p}} \left(|\pi_p^N f_1 - \phi_p(\pi_{p-1}^N) f_1| + |\pi_p^N f_2 - \phi_p(\pi_{p-1}^N) f_2|\right)$$

where f_1 and f_2 are some bounded measurable functions on $I\!R^p$ such that $\|f_1\|, \|f_2\| \leq 1$ and with the convention that $A_{n/n} = a_{n/n} = 1$.
By recalling that π_p^N is the empirical measure associated to N conditionally independent random variables with common law $\phi_p(\pi_{p-1}^N)$ we have the quenched error bound

$$E_Y\left(|\pi_p^N f - \phi_p(\pi_{p-1}^N) f|\right) \leq \frac{1}{\sqrt{N}}$$

for any bounded measurable function f with $\|f\| \leq 1$.

Collecting the above inequalities one concludes that for any $n = 0, \ldots, M$

$$E_Y\left(|\pi_n^N f - \pi_n f|\right) \quad \leq \quad \frac{2}{\sqrt{N}}(\sum_{p=0}^{n} \frac{A_{n/p}}{a_{n/p}}) = \frac{2}{\sqrt{N}}(1 + n\frac{A_{n/0}}{a_{n/0}}).$$

This yields

$$\begin{aligned} \sup_{n=0,\ldots,M} E_Y\left(|\pi_n^N f - \pi_n f|\right) &\leq \frac{4M}{\sqrt{N}} \frac{A_{M/0}}{a_{M/0}} \\ &= \frac{4M}{\sqrt{N}} \exp\left(\frac{\|h\|^2}{2} + 2\|h\| \sum_{n=1}^{M} |\Delta Y_{t_n}|\right) \end{aligned}$$

and the proof of Proposition 3.1 is completed. ∎

The term $\frac{M}{\sqrt{N}}$ in the right hand side of (14) expresses the fact that our interacting particle algorithm has M steps and at each step the approximating empirical measure consists of N particles. We can obtain a sharper quenched error bound if we use the "martingale approach" presented independently in [6] and [2]. More precisely we have the following result

Theorem 3.2 *For any bounded measurable function f on $\mathbb{R}^p$ such that $\|f\| \leq 1$ and for any $M, N \geq 1$ we have P.a.s.*

$$\sup_{n=0,\ldots,M} E_Y(|\pi_{t_n}^M f - \pi_{t_n}^{M,N} f|) \leq 2\sqrt{2}\sqrt{\frac{M}{N}}\, C(M,Y) \qquad (18)$$

with

$$\log C(M,Y) = \frac{\|h\|^2}{2} + 2\|h\| \sum_{n=1}^{M} |\Delta Y_{t_n}|.$$

Proof:

Once again we fix the observation Y and we use of simplified notations introduced in the proof of Proposition 3.1.

Before getting down to serious business it is useful to note that the dynamical system (11), the composite mappings $\phi_{n/p}$ and the recursive formulas (16) remain the same if we replace all the functions g_n by the "normalized" ones

$$g_{n/n-1} \stackrel{\text{def}}{=} \frac{g_n}{\pi_{n-1}(g_n)},$$

but in this situation the functions $g_{n/p}$ are defined by the recursion

$$g_{n/p-1} = g_{p/p-1}\, P(g_{n/p}) \qquad p = 1, \ldots, n$$

with the convention $g_{n/n} = 1$.

Using the above notation we clearly have

$$\pi_p(g_{n/p}) = 1 \qquad \forall 0 \leq p \leq n. \qquad (19)$$

To see this it suffices to note that $\pi_{n-1}(g_{n/n-1}) = 1$ and, for any $1 \leq p \leq n$ we have

$$\pi_p(g_{n/p}) = \frac{\pi_{p-1}\left(g_{p/p-1}\, P(g_{n/p})\right)}{\pi_{p-1}\left(g_{p/p-1}\right)} = \pi_{p-1}(g_{n/p-1}).$$

On the other hand, when using these"normalized" functions we have the "normalized" bounds

$$B_{n/p}^{-1} \leq g_{n/p}(x) \leq B_{n/p} \qquad 0 \leq p \leq n \leq M$$

with

$$\log B_{n/p} \stackrel{\text{def}}{=} 2\|h\| \sum_{q=p+1}^{n} |\Delta Y_{t_q}| + \frac{n-p}{2M} \|h\|^2. \tag{20}$$

To clarify the presentation for any bounded test function f we write $f_{n/p}$ for the functions given by

$$f_{n/p} = g_{n/p}\, P_{n/p}(f) \qquad p = 0, \ldots, n.$$

Arguing as before we have

$$\pi_p(f_{n/p}) = \pi_n(f) \qquad \forall 0 \leq p \leq n.$$

To see this it suffices to note that (19) and the recursive formula (16) implies that

$$\pi_n(f) = \frac{\pi_p\left(g_{n/p}\,(P_{n/p}f)\right)}{\pi_p\left(g_{n/p}\right)} = \pi_p\left(g_{n/p}\,(P_{n/p}f)\right) = \pi_p(f_{n/p}).$$

As in [6] the next step is to introduce the random sequence $\{U_n\ :\ n = 0, \ldots, M\}$ defined by

$$U_n = \prod_{p=0}^{n-1} \pi_p^N(g_{p+1/p})$$

with the standard convention $\prod_{\emptyset} = 1$ and the decomposition

$$U_n\ \pi_n^N(f) = \sum_{p=1}^{n} \left(U_p \pi_p^N(f_{n/p}) - U_{p-1}\pi_{p-1}^N(f_{n/p-1})\right) + \pi_0^N(f_{n/0}). \tag{21}$$

Using the fact that

$$\pi_{p-1}^N(f_{n/p-1}) = \pi_{p-1}^N(g_{p/p-1})\ \frac{\pi_{p-1}^N\left(g_{p/p-1}\,(Pf_{n/p})\right)}{\pi_{p-1}^N\left(g_{p/p-1}\right)} = \pi_{p-1}^N(g_{p/p-1})\ \phi_p(\pi_{p-1}^N)(f_{n/p})$$

we arrive at

$$U_{p-1}\ \pi_{p-1}^N(f_{n/p-1}) = U_p\ \phi_p(\pi_{p-1}^N)(f_{n/p}),$$

so that (21) may be rewritten in the more tractable form

$$\begin{aligned} U_n\ \pi_n^N(f) - \pi_n f &= U_n\ \pi_n^N(f) - \pi_0(f_{n/0}) \\ &= \sum_{p=0}^{n} U_p\left(\pi_p^N(f_{n/p}) - \phi_p(\pi_{p-1}^N)(f_{n/p})\right) \end{aligned}$$

with the convention $\phi_0(\pi_{-1}^N) = \nu$.

Using the fact that U_p only depends on the random empirical measures $(\pi_0^N, \ldots, \pi_{p-1}^N)$ up to time $(p-1)$ we have that

$$E_Y\left((U_n\ \pi_n^N(f) - \pi_n f)^2\right) = \sum_{p=0}^{n} E_Y\left(U_p^2\left(\pi_p^N(f_{n/p}) - \phi_p(\pi_{p-1}^N)(f_{n/p})\right)^2\right).$$

Recalling that

$$E_Y\left(\left(\pi_p^N(f_{n/p}) - \phi_p(\pi_{p-1}^N)(f_{n/p})\right)^2 \Big| \pi_{p-1}^N\right) = \frac{1}{N}\phi_p(\pi_{p-1}^N)\left(\left(f_{n/p} - \phi_p(\pi_{p-1}^N)(f_{n/p})\right)^2\right)$$

and using the bounds (20) one gets

$$\begin{aligned} E_Y\left(|U_n\ \pi_n^N(f) - \pi_n f|^2\right) &\leq \frac{1}{N}\sum_{p=0}^{n} E_Y\left(U_p^2\phi_p(\pi_{p-1}^N)(f_{n/p}^2)\right) \\ &\leq \frac{\|f\|^2}{N}\sum_{p=0}^{n} E_Y\left(U_p^2\phi_p(\pi_{p-1}^N)(g_{n/p}^2)\right) \end{aligned} \tag{22}$$

and

$$\begin{aligned} E_Y\left(|U_n\ \pi_n^N(f) - \pi_n f|^2\right) &\leq \frac{\|f\|^2}{N}\sum_{p=0}^{n}\prod_{q=1}^{p} B_{q/q-1}^2\ B_{n/p}^2 \leq \frac{\|f\|^2}{N}\sum_{p=0}^{n} B_{n/p}^2\ B_{p/0}^2 \\ &\leq \frac{n+1}{N} B_{n/0}^2\ \|f\|^2. \end{aligned}$$

This implies that

$$\begin{aligned} E_Y\left(|\pi_n^N(f) - \pi_n f|\right) &\leq E_Y\left(|U_n\ \pi_n^N(f) - \pi_n f|\right) + \|f\|\ E_Y\left(|U_n - 1|\right) \\ &\leq 2\sqrt{\frac{n+1}{N}} B_{n/0}\ \|f\| \leq 2\sqrt{2}\sqrt{\frac{M}{N}}\ C(M,Y)\ \|f\|. \end{aligned}$$

and completes the proof of the theorem. ■

3.2.2 Order of convergence

For any $n = 0, \ldots, M-1$ and $t \in [t_n, t_{n+1})$ we denote by $\pi_t^{M,N}$, the empirical measures associated with the system ξ_{t_n}, namely

$$\pi_t^{M,N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_{t_n}^i}.$$

In view of Theorem 3.2 we may ask whether we can deduce the convergence of $\{\pi_t^{M,N} \;:\; t \in [0,1)\}$ to the solution of the Kushner Stratonovich equation. This question has of course a positive answer but the error bounds obtained using these quenched results are not really satisfactory. More precisely Theorem 3.2 has the following immediate consequence.

Corollary 3.3 *For any bounded Lipschitz test function f such that*

$$|f(x) - f(z)| \leq k(f)\, |x - z|$$

we have that

$$\sup_{t\in[0,1)} E\left(|\pi_t f - \pi_t^{M,N} f|\right) \leq \frac{\mathrm{C}}{\sqrt{\mathrm{M}}}\left(\|f\| + k(f)\right) + 2\sqrt{2}\|f\|\sqrt{\frac{M}{N}} E\left(C(M,Y)\right)$$

and

$$\frac{5}{2}\|h\|^2 + 2\|h\|\sqrt{\frac{2qM}{\pi}} \leq \log E\left(C(M,Y)\right) \leq 5q\|h\|^2 + 2q\|h\|\sqrt{\frac{2M}{\pi}} \qquad (23)$$

where the finite constant C is the one given in Theorem 2.3

Proof:
In a first place we note that

$$E\left(|\pi_t f - \pi_t^{M,N} f|\right) \leq E\left(|\pi_t f - \pi_t^{M} f|\right) + E\left(|\pi_t^M f - \pi_t^{M,N} f|\right).$$

Thus, in view of Theorem 2.3 and Theorem 3.2 it remains to check (23).
To clarify the presentation we will only treat the case $p = q = 1$ since the extension to the vector case is straightforward.

By definition of the observation process Y we first note that

$$\log C(M,Y) = \frac{5\|h\|^2}{2} + 2\|h\| \sum_{n=1}^{M} |\Delta V_{t_n}|.$$

On the other hand, it is well known that for any Gaussian variable U with $E(U) = 0$ and $E(U^2) = 1$ we have that

$$E(\exp(\alpha|U|)) \leq \left(1 + \alpha\sqrt{\frac{2}{\pi}}\right) \exp\frac{\alpha^2}{2} \qquad \forall \alpha \geq 0.$$

Applying this inequality and using the fact that

$$\left(1 + \frac{\alpha}{\sqrt{M}}\right)^M \leq \exp \alpha\sqrt{M} \qquad \forall \alpha \geq 0,$$

and after some elementary manipulations one gets

$$\log E(C(M,Y)) \leq 5\|h\|^2 + 2\|h\|\sqrt{\frac{2M}{\pi}}.$$

This yields

$$\sup_{t\in[0,1)} E(|\pi_t^M f - \pi_t^{M,N} f|) \leq C_1 \frac{M}{\sqrt{N}} \exp\left(C_2\sqrt{M}\right)$$

with

$$C_1 = 4\exp\left(5\|h\|^2\right) \qquad \text{and} \qquad C_2 = 2\|h\|\sqrt{\frac{2}{\pi}}.$$

Using Jensen's inequality we also have that

$$\log E(C(M,Y)) \geq \left(\frac{5\|h\|^2}{2} + 2\|h\|\sqrt{\frac{2M}{\pi}}\right).$$

The extension to the vector case is a clear consequence of the following chain of inequalities. If we write

$$\Delta V_{t_n} = \left(\Delta V_{t_n}^1, \ldots, \Delta V_{t_n}^q\right) \in I\!R^q,$$

then we have

$$\begin{aligned} |\Delta V_{t_n}| &= \left(\sum_{r=1}^{q} |\Delta V_{t_n}^r|^2\right)^{1/2} \leq \sum_{r=1}^{q} |\Delta V_{t_n}^r| \\ \exp\left(\alpha\, E(|\Delta V_{t_n}|)\right) &\leq E\left(\exp\left(\alpha\, |\Delta V_{t_n}|\right)\right) \leq E\left(\exp\left(\alpha\, |\Delta V_{t_n}^1|\right)\right)^q \qquad \forall \alpha \geq 0. \end{aligned}$$

■

Remark 3.4:
In view of (23) one cannot expect to obtain a better error bound using the above lemma. Roughly speaking the exponential term (23) represents the mean growth rate of the increasing sequence of variables $\{C(M,Y)\ :\ M \geq 1\}$.

We now present the main result of this section.

Theorem 3.5 *For any bounded Lipschitz test function f such that*

$$|f(x) - f(z)| \leq k(f)\ |x - z|$$

we have that

$$\sup_{t\in[0,1)} E\left(|\pi_t f - \pi_t^{M,N} f|\right) \leq \frac{C_1}{\sqrt{M}}\ (\|f\| + k(f)) + C_2\ \sqrt{\frac{M}{N}}\ \|f\| \qquad (24)$$

where C_1 is the finite constant defined in theorem 2.3 and $C_2 = 2\sqrt{2}\exp(12\|h\|^2)$. In addition, if $p = q = 1$ and a, b, f, h are four times continuously differentiable with bounded derivatives then we have

$$\sup_{t\in[0,1)} E\left(|\pi_t f - \pi_t^{M,N} f|\right) \leq \text{Cte}\left(\frac{1}{M} + \sqrt{\frac{M}{N}}\right). \tag{25}$$

Proof:
In view of (13) is suffices to check that

$$\sup_{t\in[0,1)} E\left(|\pi_t^M f - \pi_t^{M,N} f|\right) \leq 2\sqrt{2}\sqrt{\frac{M}{N}}\|f\| \exp(12\|h\|^2). \tag{26}$$

The proof of (26) relies strongly on the proof of Theorem 3.2. To clarify the presentation all notations introduced in the proof of Theorem 3.2 are in force.
We will also simplify the notation suppressing the time parameter t and writing X_n, ξ_n and ΔY_n instead of X_{t_n}, ξ_{t_n} and ΔY_{t_n}.

Our immediate goal is to replace the quenched error bounds in (22) by some explicit mean error bounds.

If we denote $\overline{g}_{n/p}$ the "unormalized" functions given by

$$\overline{g}_{n/p} = g_{p+1}\, P(\overline{g}_{n/p+1}) \tag{27}$$

with the convention $\overline{g}_{n/n} = 1$, we have the relation

$$g_{n/p}(x) = \overline{g}_{n/p}(x) \Big/ \prod_{q=p}^{n-1} \pi_q(g_{q+1}) .$$

Using the fact that $\|P_{n/p} f\| \leq \|f\|$ and $g_p \geq 1$ one can easily check that

$$U_p^2 \phi_p(\pi_{p-1}^N)(f_{n/p}^2) \leq U_n^2\, \|f\|^2\, \phi_p(\pi_{p-1}^N)(\overline{g}_{n/p}^2).$$

Up to now we have considered the observation records as a fixed series of parameters. Our aim is now to produce some mean error bounds. To this end is it necessary to evaluate the mean of the terms in the right hand side of (22).
The key idea consists in using in a first stage the reference probability measure P_0. This approach was presented by two of the authors in [1, 2]. Recalling that under P_0 the observation process Y is a standard Brownian motion the random sequence $\{\Delta Y_n \,:\, n \geq 0\}$ can be considered as a sequence of i.i.d. and $\mathbb{R}^q$-valued Gaussian variables with zero mean and $E((\Delta Y_1)^\star(\Delta Y_1)) = M^{-1}\, I_{q\times q}$.

In what follows we denote by $E_0(\bullet)$ the expectation associated with P_0. Using the above notation and the Cauchy-Schwartz inequality we arrive at

$$E_0\left(U_p^2\phi_p(\pi_{p-1}^N)(f_{n/p}^2)\right)^2 \leq \|f\|^4 \; E_0\left(U_n^4\right) \; E_0\left(\overline{g}_{n/p}^4(\xi_p^1)\right). \tag{28}$$

To complete the proof of the theorem it remains to estimate the terms

$$E_0\left(U_n^4\right) \qquad \text{and} \qquad E_0\left(\overline{g}_{n/p}^4(\xi_p^1)\right).$$

In view of (27) we clearly have for any $x \in I\!R^p$

$$E_0\left(\overline{g}_{n/p}^4(\xi_p^1) \,/\xi_p^1 = x\right) \leq E_0\left(g_{p+1}^4(X_p)\ldots g_n^4(X_{n-1})\,/X_p = x\right).$$

On the other hand for any $p+1 \leq q \leq n$ we have that

$$\begin{aligned} E_0\left(g_q^4(X_{q-1})\,|X_{q-1}\right) &\leq E_0\left(\exp\left(4h^\star(X_{q-1})\Delta Y_q\right)|X_{q-1}\right) \\ &= \exp\frac{8|h(X_{q-1})|^2}{M} \leq \exp\frac{8\|h\|^2}{M}. \end{aligned}$$

It follows that

$$E_0\left(\overline{g}_{n/p}^4(\xi_p^1)\right) \leq \exp\left(\frac{8(n-p)}{M}\|h\|^2\right). \tag{29}$$

Let us now estimate the term $E_0(U_n^4)$.

Using Jensen's formula we can check that U_n^4 is upper bounded by

$$I_n = \prod_{p=0}^{n-1} \pi_p^N\left(\exp\left(4\,\widetilde{l}_{p+1}\right)\right)$$

where

$$\widetilde{l}_{p+1}(x) = l_{p+1}(x) - \pi_p(l_{p+1}) \qquad \text{and} \qquad l_{p+1}(x) = \log g_{p+1}.$$

Noticing that

$$\widetilde{l}_n(x) = \left(h^\star(x) - \pi_{n-1}(h)\right)\Delta Y_{t_n} - \frac{1}{2M}\left(|h(x)|^2 - \pi_{n-1}(|h|^2)\right)$$

and for any $\alpha \geq 0$ and $x \in I\!R^p$

$$\log E_0\left(e^{\alpha(h^\star(x)-\pi_{n-1}(h))\Delta Y_{t_n}}\,|\pi_{n-1}\right) = \frac{\alpha^2}{2M}\,|h^\star(x) - \pi_{n-1}(h)|^2 \leq \frac{2\alpha^2\|h\|^2}{M},$$

we arrive at

$$\log E_0\left(\pi_{n-1}^N\left(e^{4\widetilde{l}_n}\right)|\Delta Y_{t_1},\ldots,\Delta Y_{t_{n-1}},\pi_{n-1}\right) \leq \frac{16}{2M}\,4\|h\|^2 + \frac{1}{M}\,\|h\|^2 = \frac{33}{M}\,\|h\|^2.$$

Thus, one easily gets the inequality

$$E_0\,(I_n) \leq \exp\left(\frac{33}{M}\,\|h\|^2\right)\,E_0\,(I_{n-1})\,,$$

so that

$$E_0\left(U_n^4\right) \leq \exp\left(33\,\|h\|^2\right) \qquad \forall 0 \leq n \leq M. \tag{30}$$

If we combine (28), (29) and (30) we conclude that

$$E_0\left(U_p^2\phi_p(\pi_{p-1}^N)(f_{n/p}^2)\right) \leq \|f\|^2\ \exp\,(17\|h\|^2 + 4\|h\|^2) = \|f\|^2 \exp\,(21\|h\|^2).$$

Hence, in view of (22) we have

$$E_0\left((U_n\ \pi_n^N(f) - \pi_n f)^2\right) \leq 2\frac{M}{N}\ \|f\|^2 \exp\,(21\|h\|^2).$$

Taking into consideration the inequality

$$\begin{aligned} E_0\left(\left(\pi_n^N(f) - \pi_n f\right)^2\right) \ &\leq\ E_0\left(\left[\left(U_n\ \pi_n^N(f) - \pi_n f\right) + \pi_n^N(f)\,(1-U_n)\right]^2\right) \\ &\leq\ 2\ E_0\left((U_n\ \pi_n^N(f) - \pi_n f)^2\right) + 2\|f\|^2\ E_0\left((U_n-1)^2\right), \end{aligned}$$

we conclude that

$$E_0\left(|\pi_n^N(f) - \pi_n f|^2\right) \leq 8\frac{M}{N}\ \|f\|^2 \exp\,(21\|h\|^2).$$

Finally, using the inequalities

$$E\left(|\pi_n^N(f) - \pi_n f|\right) \leq E_0\left(|\pi_n^N(f) - \pi_n f|^2\right)^{1/2}\ E_0\left(\left(\frac{dP}{dP_0}\right)^2\right)^{1/2}$$

and

$$E_0\left(\left(\frac{dP}{dP_0}\right)^2\right) = E\,(Z_1) \leq \exp\,(\|h\|^2),$$

after some elementary manipulations we arrive at the desired mean error bound

$$E\left(|\pi_n^N(f) - \pi_n f|\right) \leq 2\sqrt{2}\sqrt{\frac{M}{N}}\|f\|\exp\,(12\|h\|^2)$$

from which the end of proof of (24) is now straightforward.

When the signal X and the observation Y are real processes and the functions

a, b, f, h are four times continuously differentiable with bounded derivatives a more precise error bound has been obtained by Picard in [22], namely,

$$\sup_{t\in[0,1)} E\left(|\pi_t f - \pi_t^M f|\right) \leq \frac{\mathrm{Cte}}{\mathrm{M}}$$

Under these assumptions (25) is a clear consequence of (26) ∎

Remark 3.6:
Theorem 1.1 is a clear consequence of Theorem 3.5 and the fact that

$$\sqrt{\frac{M}{N}} \leq \frac{1}{\sqrt{M}} \Longleftrightarrow M \leq N^{1/2} \qquad \sqrt{\frac{M}{N}} \leq \frac{1}{M} \Longleftrightarrow M \leq N^{1/3}.$$

The basic space for the study of the weak convergence of a sequence of random measures is the set of all probability measures on $\mathbf{M}_1(I\!R^p)$ denoted by$\mathbf{M}_1\left(\mathbf{M}_1(I\!R^p)\right)$. On this set we can define the Kantorovitch-Rubinstein or Vasershtein metric

$$D(\Phi_1, \Phi_2) = \inf E\left(\|\mu_1 - \mu_2\|_{\mathcal{F}}\right) \qquad \Phi_1, \Phi_2 \in \mathbf{M}_1\left(\mathbf{M}_1(I\!R^p)\right)$$

where the infimum is taken over all pair of random variables (μ_1, μ_2) such that μ_1 has distribution Φ_1 and μ_2 has distribution Φ_2.
This metric gives to $\mathbf{M}_1\left(\mathbf{M}_1(I\!R^p)\right)$ the topology of weak convergence. If we note $\Phi_t^{M,N}$ the distribution of $\pi_t^{M,N}$ and Φ_t the distribution of π_t the conclusions of Theorem 3.5 yield

$$M \leq \sqrt{N} \Longrightarrow D(\Phi_t^{M,N}, \Phi_t) \leq \frac{\mathrm{Cte}}{\mathrm{N}^{1/4}}.$$

Concluding Remarks

We have proposed an approximating scheme for the numerical solution of the Kushner-Stratonovitch equation. We recall that our approach is based on the use of the classical time discretization scheme as presented in [17, 18, 22] and on the interacting particle scheme presented in [6, 9, 7]. This yields a natural line of proof for the convergence of our particle approximating models.

The main contribution here was to connect these two separate approximating schemes and to propose explicit error bounds associated to this combination.
The interacting particle system approach studied in this work is one of the crudest particle methods. For the sake of unity and to highlight issues specific to interacting and branching particle system theory we mention that the same line of proof can be used to get the convergence of more general particle schemes such as those presented in [1, 2] and [3] .

We also feel that our approach is more transparent and the proof of convergence is simpler than other studies on interacting particle approximations.
We have only treated a very special case of a non linear filtering problem. It is obvious that the situation becomes considerably more involved if one dispenses with the assumption that the observation process is solution of an Ito-type differential equation of the form (2). To deal with more general signal/observation pair it has recently been proposed in [12] an alternative interacting particle system approach which only uses the semigroup of the pair process $\{(X_t, Y_t) \ : \ t \in [0,1]\}$. The idea is to use particles which explore the whole state/observation space and to "select" particles when their sampled observation component is close to the current observation data.

References

[1] CRISAN D., LYONS T.J, Nonlinear Filtering And Measure Valued Processes. *Prob. Theory and Related Fields*, 109, 217-244, 1997.

[2] CRISAN D., J. GAINES, LYONS T.J., A Particle Approximation of the Solution of the Kushner-Stratonovitch Equation. *SIAM Journal of Applied Math.* (to appear).

[3] CRISAN D., DEL MORAL P., LYONS T.J. (1998), Discrete Filtering Using Branching and Interacting Particle Systems, *Publications du Laboratoire de Statistiques et Probabilités*, Université Paul Sabatier, **No 01-98**.

[4] DI MASI G.B., PRATELLI M., RUNGGALDIER W.G., An approximation for the Non linear filtering problem with error bounds. *Stochastics* 14 (1985), no. 4, 247–271.

[5] DEL MORAL P. (1996). Non linear filtering using random particles. *Theor. Prob. Appl.,* **Vol. 40**, *(4)*.

[6] DEL MORAL P. (1996). Non-linear Filtering: Interacting particle resolution. *Markov Processes and Related Fields,* **Vol.2**

[7] DEL MORAL P. (1996). Measure Valued Processes and Interacting Particle Systems. Application to Non Linear Filtering Problems. *Ann. Appl. Probab.* 8 (1998), no. 2, 438–495..

[8] DEL MORAL P., A Uniform Theorem for the Numerical Solving of Non Linear Filtering Problems. *Journal of Applied Probability* (to appear).

[9] DEL MORAL P (1997), Filtrage non linéaire par systèmes de particules en intéraction. *C.R. Acad. Sci. Paris, t. 325, Série I, p. 653-658.*

[10] DEL MORAL P., GUIONNET A., Large Deviations for Interacting Particle Systems. Applications to Non Linear Filtering Problems. *Stochastic Processes Appl.*, (to appear).

[11] DEL MORAL P., GUIONNET A. (1997). A Central Limit Theorem for Non Linear Filtering using Interacting Particle Systems. *Ann. Appl. Probab.*, (to appear).

[12] DEL MORAL P., JACOD J., PROTTER Ph. (1998). The Monte Carlo Method for Filtering with discrete-time observations. *Prépublication n° 453 du Laboratoire de Probabilités de L'Université Paris VI*, 1998.

[13] FORTET R., MOURIER B. (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. École Norm. Sup.*, 70, pp. 267-285.

[14] GORDON N.J., SALMON D.J., SMITH A.F.M. (1993). Novel Approach to non-linear/non-Gaussian Bayesian state estimation.,*IEE Proceedings 140*, 107-113.

[15] HOLLAND J.H. (1975). Adaptation in natural and artificial systems. *University of Michigan Press, Ann. Arbor*

[16] KOREZLIOGLU H. (1987). Computation of Filters by Sampling and Quantization. *Center for Stochastic Processes*, University of North Carolina, Technical Report No 208.

[17] KOREZLIOGLU H., MAZIOTTO G. (1983). Modelization and Filtering of Discrete Systems and Discrete Approximation of Continuous Systems. *VI Conférence INRIA, Modélisation et Optimisation des Systèmes*, 19-22/06/84, Nice.

[18] KOREZLIOGLU H., RUNGGALDIER W.J. (1992). Filtering for Non Linear Systems Driven by Nonwhite Noises : An Approximating Scheme. *Stochastics Stochastics Rep.* 44 (1993), no 1-2, 65-102.

[19] LEGLAND F. (1989). High order time discretization of nonlinear filtering equations, *28th IEEE CDC*, Tampa, pp. 2601-2606.

[20] VAN DOOTINGH M., VIEL F., RAKOTOPARA D., GAUTHIER J.P. (1991), Coupling of non-linear control with a stochastic filter for state estimation : Application on a free radical polimerization reactor,*I.F.A.C. International Symposium ADCHEM'91, Toulouse, France, October 14-15*

[21] PARDOUX E., TALAY D. Approximation ans simulation of solutions of stochastic differential equations, *Acta Applicandae Mathematicae* 3 (1985), no 1, 23-47.

[22] PICARD J. (1984). Approximation of non linear filtering problems and order of convergence. Filtering and Control of random processes, *Lecture Notes Control and Inf. Sc.*, 61, Springer. Filtering

[23] PICARD J. (1986). Nonlinear filtering of one-dimensional diffusions in the case of a high signal-to-noise ratio, *SIAM J. Appl. Math.*, **16**, pp. 1098-1125.

[24] PICARD J. (1986). An estimate of the error in time discretization of nonlinear filtering problems, Proceedings of the 7th MTNS — Theory and Applications of Nonlinear Control Systems, Stockholm 1985, C.I. Byrnes and A. Lindquist Eds. North–Holland Pub. Amsterdam, pp. 401-412.

[25] TALAY D. Efficient numerical schemes for the approximation of expectations of functionals of the solution of S.D.E. and applications, *Filtering and Control of random processes (Paris 1983)*, 294-313, Lecture Notes in Control and Inform. Sci., 61, *Springer, Berlin-New York*, 1984.