The Annals of Applied Probability
1998, Vol. 8, No. 2, 438–495

MEASURE-VALUED PROCESSES AND INTERACTING PARTICLE SYSTEMS. APPLICATION TO NONLINEAR FILTERING PROBLEMS[1]

BY P. DEL MORAL

Université Paul Sabatier

In the paper we study interacting particle approximations of discrete time and measure-valued dynamical systems. These systems have arisen in such diverse scientific disciplines as physics and signal processing. We give conditions for the so-called particle density profiles to converge to the desired distribution when the number of particles is growing. The strength of our approach is that is applicable to a large class of measure-valued dynamical systems arising in engineering and particularly in nonlinear filtering problems. Our second objective is to use these results to solve numerically the nonlinear filtering equation. Examples arising in fluid mechanics are also given.

1. Introduction.

1.1. *Measure-valued processes.* Let $(E, \beta(E))$ be a locally compact and separable metric space, endowed with a Borel σ-field, state space. Denote by $\mathscr{P}(E)$ the space of all probability measures on E with weak topology. The aim of this work is the design of a stochastic particle system approach for the computation of a general discrete time and measure-valued dynamical system η_n given by

$$\eta_n = \phi(n, \eta_{n-1}) \qquad \forall\, n \geq 1,\ \ \eta_0 = \eta, \tag{1}$$

where $\eta \in \mathscr{P}(E)$ and $\phi(n, \bullet)$: $\mathscr{P}(E) \to \mathscr{P}(E)$ is a continuous function.

Such systems have arisen in such diverse scientific disciplines as physics (see [44] and the references given there), nonlinear economic modelling and signal processing (see [29] and [41]). Solving (1) is in general an enormous task as it is nonlinear and usually involves integrations over the whole space E. To obtain a computationally feasible solution, some kind of approximation is needed.

On the other hand, particle methods have been developed in physics since World War II, mainly for the needs of fluid mechanics [32, 33, 43] and statistical mechanics [22, 31, 40]. During the past decade their application area has grown, establishing unexpected connections with a number of other fields.

Our major motivation is from advanced signal processing applications in engineering and particularly in optimal nonlinear filtering problems. Recall that this consists of computing the conditional distribution of internal states

Received April 1997; revised September 1997.

[1]Supported in part by CEC Contract ERB-FMRX-CT96-0075 and INTAS-RFBR project 95-0091.

AMS 1991 *subject classifications*. 60G57, 60K35, 60G35, 93E11.

Key words and phrases. Nonlinear filtering, measure-valued processes, interacting and branching particle systems, genetic algorithms.

in dynamical systems when partial observations are made, and random perturbations are present in the dynamics as well as in the sensors. With the notable exception of the linear-Gaussian situation or wider classes of models [1], optimal filters have no finitely recursive solution [7]. Nevertheless, guided by Bayes' rule, we will see that the dynamic structure of such conditional distributions can be viewed, under mild assumptions, as a special case of (1). In our formulation, the most important measure of complexity of the problem is now reduced to the infinite dimensionality of the state space $\mathscr{P}(E)$. The main advantage of dealing with (1) rather than the conditional expectation is that the solution of (1) is Markovian and the solution of the conditional expectation plant equation is not. Our claim that this formulation is the natural framework for formulating and solving problems in nonlinear filtering will be amply justified by the results which follow.

The paper has the following structure. After fixing the context within which we work, we introduce in Section 3 a stochastic particle system approach to solve (1). The particle systems described in this section will consist of finitely many particles and the systems as a whole will be Markov processes with product state space. In this framework, the transition for individual particles will depend on the entire configuration, so the particles are not independent and the evolution of an individual particle is no longer Markovian. The models to be treated in Section 3.1 are obtained by imposing various types of interactions on the motions of the particles. We begin in this section with the description of a simple particle system in which the interaction function depends only on the empirical measure of the particle system. Furthermore, it is shown that the so-called particle density profiles, that is, the random empirical measures of the particle system, converge to the solution of (1) as the number of particles is growing. The proof of convergence is based on the use of semigroup techniques in the spirit of Kunita [29] and Stettner [41].

Our next objective is to extend the above construction. In Section 3.2 we introduce general particle systems which include branchings and nonlinear interactions, and we also prove that the random empirical measures of the particle system weakly converge to the solution of (1) as the number of particles is growing. The proof of convergence involves essentially the same analytic techniques which are used in Section 3.1. We will examine as much of the theory as possible in a form applicable to general nonlinear filtering problems and to the master equations arising in fluid mechanics.

The strength of our approach is that it is applicable to a large family of measure-valued dynamical systems. Several examples are worked out in Section 3.3. One way in which our approach may be applied to fluid mechanics problems is discussed. We will only examine the convergence of the particle density profiles. In real problems, the crucial question is the convergence of the empirical measures of the particle system on the path space. These problems are quite deep and we shall only scratch the surface here. For a detailed discussion and a full treatment of the above questions, the reader is referred to [43, 44] and the references given there. However we emphasize that our proof uses only semigroup techniques and turns out to be more transparent.

In Section 4, we will use the results of Section 3 to describe several interacting particle approximations of the nonlinear filtering plant equation. This section is divided into two subsections. In Section 4.1 we present a general discussion of the nonlinear filtering theory and we formulate the filtering problem in such a way that the techniques of Section 3 can be applied. The design of our particle system approach is described in Section 4.2. We finish the paper with a description of general interacting particle resolutions.

1.2. *Nonlinear filtering.* In our development, special knowledge of nonlinear filtering theory is not assumed. For a detailed discussion of the filtering problem, the reader is referred to the pioneering paper of Stratonovich [42] and to the more rigorous studies of Shiryaev [39] and Kallianpur and Striebel [28]. More recent developments can be found in [34] and [35]. Some collateral readings such as [29], [41] and [7] will be helpful in appreciating the relevance of our approximations.

As far as the author knows, the various numerical methods based on fixed grid approximations, conventional linearization or determining the best linear filter (in the expected cost error sense) have never really coped with large-scale systems or unstable processes. They are usually far from optimal, particularly in high-noise environments, when there is significant uncertainty in the observations or when the nonlinearities are not suitably smooth. More precisely, it is in general impossible to identify an event space region R so that the trajectories of the state signal lie entirely in R; therefore, it is difficult to use fixed grid methods. Moreover, it is well known that large state dependent noise has destabilizing effects on the dynamics of the best linear filter and tends to increase the magnitude of its gain.

It has been recently emphasized that a more efficient way is to use random particle systems to solve the filtering problem numerically. That particle algorithms are gaining popularity is attested by the recent papers; see [4, 5, 9–11, 16–18, 20, 21, 24, 45]. Instead of hand-crafting algorithms, often on the basis of ad hoc criteria, particle systems approaches provide powerful tools for solving a large class of nonlinear filtering problems.

Let us now briefly survey some different approaches and motivate our work. In [15] and [17] the author introduced particle methods where particles are independent of each other and weighted by regularized exponentials and he proposes minimal conditions which ensure the convergence to the optimal filter uniformly with respect to time. Crisan and Lyons develop in [9] a sequence of branching processes in the spirit of Dawson and Watanabe constructions [12, 46] whose expectation converges to the solution of the Zakai equation. In [10] they construct a different sequence of branching particle systems converging in distribution to the solution of the Zakai equation. In their last work [11], they describe another sequence of branching particle systems converging in measure to the solution of the Kushner–Stratonovitch equation.

In the present paper we describe different interacting and branching particle systems, and we prove that the empirical measure converges to the desired conditional distribution. The connection between such particle systems and

genetic algorithms is given in [16]. These algorithms are an extension of the well-known sampling–resampling principles introduced by Gordon, Salmon and Smith in [24] and independently by Del Moral, Rigal, Noyer and Salut in [14], [18] and [21]. Roughly speaking, they consist of periodically redistributing the particle positions in accordance with a discrete representation of the conditional distribution. This procedure allows particles to give birth to more particles at the expense of light particles, which die. This guarantees an occupation of the state space regions proportional to their probability mass, thus providing an adaptative and stochastic grid.

We will prove the convergences of such approximations to the optimal filter, yielding what seems to be the first convergence results for such approximations of the nonlinear filtering equations. This new treatment was influenced primarily by the development of genetic algorithms [26, 6] and secondarily by [29, 41].

Examples of several practical problems which have been solved using these methods are given in [5], [20], [24], [21], [4] including problems in radar/sonar signal processing and GPS/INS integration.

There is an essential difference between our particle systems and the branching particle systems described in [9], [10] and [11]. What makes our results interesting and new is that the number of particles is fixed and the system of interacting particles is as a whole a Markov process which feels its environment through the observations. Moreover, the transition for an individual particle depends on the entire configuration of the system and not only on its current position. On the other hand, our constructions are explicit, the recursions have a simple form and they can be easily generated on a computer. Thus, armed with these algorithms and the worked out examples of Section 3.3, users should be able to handle a large class of nonlinear filtering problems. However, here the difficulties are well known: these stochastic algorithms use Monte Carlo simulations and they are usually slow when the state space is too large. We will not describe the so-called complexity of the interacting particle algorithms. The reader who wishes to know details about such questions is recommended to consult [4], [5], [24], [21]. There is no universal agreement at present on the choice of the most efficient particle algorithm. Obviously this is a topic on which no one can claim to have the final answer. Perhaps such problems will become relatively transparent only after we have developed a theory powerful enough to describe all of these particle algorithms.

Such algorithms have been made valuable in practice by advances in computer technology. They fall within the scope of new architecture computers such as vector processors, parallel computers and connection machines. Moreover, they are ready to cope with real-time constraints imposed by practical problems.

2. Preliminaries. In this paper we will consider stochastic processes with values in the set of all probability measures on E. Such types of processes appear naturally in the study of nonlinear filtering problems (see for

instance [9], [29] and [41]) and in the propagation of chaos theory (see, for instance, [38], [43] and [44]). In this short section, we summarize the key concepts and the various forms of convergence of probability measures which are used throughout the paper. For further information the reader is referred to [36] and [23].

2.1. *E-valued random variables.* Assume E is a locally compact and separable metric space. By $\beta(E)$ we denote the σ-algebra of Borel subsets in E and by $\mathscr{P}(E)$ the family of all probability measures on $(E, \beta(E))$. As usual, by $\mathscr{B}(E)$ we denote the space of all bounded Borel measurable functions $f\colon E \to \mathbb{R}$, by $\mathscr{C}_b(E)$ the subspace of all bounded continuous functions and by $\mathscr{U}(E)$ the subspace of all bounded and uniformly continuous functions. In these spaces the norm is

$$\|f\| = \sup_{x \in E} |f(x)|.$$

For an $f \in \mathscr{B}(E)$ and $\mu \in \mathscr{P}(E)$, we write

$$\mu f = \int f(x)\, \mu(dx).$$

We say that a sequence $(\mu_N)_{N \geq 1}$, $\mu_N \in \mathscr{P}(E)$, converges weakly to a measure $\mu \in \mathscr{P}(E)$ if

$$\forall\, f \in \mathscr{C}_b(S), \qquad \lim_{N \to +\infty} \mu_N f = \mu f.$$

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

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

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

A transition probability kernel K on E is said to be Feller if

$$f \in \mathscr{C}_b(E) \quad \Rightarrow \quad K f \in \mathscr{C}_b(E). \tag{4}$$

With a Markov kernel K and a measure $\mu \in \mathscr{P}(E)$ we associate a measure $\mu \times K \in \mathscr{P}(E \times E)$ by setting

$$\forall\, f \in \mathscr{C}_b(E \times E), \qquad (\mu \times K) f = \int \mu(dx_1)\, K(x_1, dx_2)\, f(x_1, x_2).$$

It is well known that $\mathscr{P}(E)$ with the topology of weak convergence can be considered as a metric space with metric ρ defined for $\mu, \nu \in \mathscr{P}(E)$ by

$$\rho(\mu, \nu) = \sum_{m \geq 1} 2^{-(m+1)} |\mu f_m - \nu f_m|, \tag{5}$$

where $(f_m)_{m \geq 1}$ is a suitable sequence of uniformly continuous functions such that $\|f_m\| \leq 1$ for all $m \geq 1$ ([36], Theorem 6.6, page 47). Moreover we can show that $\mathscr{P}(E)$ is a separable metric space with metric ρ ([36], Theorem 6.2).

2.2. *Measure-valued random variables.* Recall that $\mathscr{P}(E)$ is a separable metric space with metric ρ. By $\beta(\mathscr{P}(E))$ we denote the σ-algebra of Borel subsets in $\mathscr{P}(E)$ and by $\mathscr{P}(\mathscr{P}(E))$ the collection of all probability measures on $(\mathscr{P}(E), \beta(\mathscr{P}(E)))$. By $\mathscr{B}(\mathscr{P}(E))$ we denote the space of all bounded Borel measurable functions $F\colon \mathscr{P}(E) \to \mathbb{R}$, by $\mathscr{C}_b(\mathscr{P}(E))$ the subspace of all bounded continuous functions and by $\mathscr{U}(\mathscr{P}(E))$ the subspace of all bounded and uniformly continuous functions. As usual, in these spaces the norm is

$$\|F\| = \sup_{\mu \in \mathscr{P}(E)} |F(\mu)|.$$

For an $F \in \mathscr{B}(\mathscr{P}(E))$ and $\Phi \in \mathscr{P}(\mathscr{P}(E))$ we write

$$\Phi F = \int F(\mu)\, \Phi(d\mu).$$

We say that a sequence $(\Phi_N)_{N \geq 0}$, $\Phi_N \in \mathscr{P}(\mathscr{P}(E))$, converges weakly to a measure $\Phi \in \mathscr{P}(\mathscr{P}(E))$ if

$$\forall\ F \in \mathscr{C}_b(\mathscr{P}(E)), \qquad \lim_{N \to +\infty} \Phi_N F = \Phi F \tag{6}$$

and this will be denoted

$$\Phi_N \xrightarrow[N \to +\infty]{w} \Phi.$$

In this paper we will consider measure-valued stochastic processes; it is therefore convenient to recall the definition of the conditional expectation of a $\mathscr{P}(E)$-valued random measure relative to a σ-field (cf. [29]). Let $\pi(\omega)$ be a $\mathscr{P}(E)$-valued random variable defined on a probability space (Ω, F, P). The conditional expectation of π relative to a sub-σ-field $G \subset F$ is defined as a $\mathscr{P}(E)$-valued random variable $E(\pi/G)$ such that

$$F(E(\pi/G)) = E(F(\pi)/G)$$

holds for all continuous affine functions $F\colon \mathscr{P}(E) \to \mathbb{R}$ [$F \in \mathscr{C}_b(\mathscr{P}(E))$ is affine if there exists a real constant c and a function $f \in \mathscr{C}_b(E)$ such that for every $\mu \in \mathscr{P}(E)$, $F(\mu) = c + \mu(f)$].

A linear mapping $M\colon \mathscr{B}(\mathscr{P}(E)) \to \mathscr{B}(\mathscr{P}(E))$ such that

$$M(1_{\mathscr{P}(E)}) = 1_{\mathscr{P}(E)} \quad \text{and} \quad M(F) \geq 0\ \forall\ F \in \mathscr{B}(\mathscr{P}(E)),\ F \geq 0$$

is called a Markov operator or Markov transition on $\mathscr{P}(E)$. Then we may define a linear mapping, still denoted by M, by setting

$$\begin{aligned} &M\colon \mathscr{P}(\mathscr{P}(E)) \to \mathscr{P}(\mathscr{P}(E)) \\ &\qquad \Phi \mapsto \Phi M \quad \text{with } \Phi M(A) \stackrel{\text{def}}{=} \Phi(M(1_A))\ \forall\ A \in \beta(\mathscr{P}(E)). \end{aligned} \tag{7}$$

A Markov transition M is said to be Feller if

$$MF \in \mathscr{C}_b(\mathscr{P}(E)) \qquad \forall\ F \in \mathscr{C}_b(\mathscr{P}(E)).$$

For an $F \in \mathscr{C}_b(\mathscr{P}(E))$ and M_1, M_2, two Feller transitions on $\mathscr{P}(E)$, we write

$$M_1 F(\mu) = \int M_1(\mu, d\nu)\, F(\nu),$$

$$M_1 M_2 F(\mu) = \int M_1(\mu, d\nu_1) M_1(\nu_1, d\nu_2)\, F(\nu_2).$$

Now, we introduce the Kantorovitch–Rubinstein or Vaserstein metric on the set $\mathscr{P}(\mathscr{P}(E))$ defined by

$$D(\Phi, \Psi) = \inf\left\{ \int \rho(\mu, \nu)\, \Theta(d(\mu, \nu)) \colon \Theta \in \mathscr{P}(\mathscr{P}(E) \times \mathscr{P}(E))\ p_1 \circ \Theta = \Phi \right. \tag{8}$$
$$\left. \text{and } p_2 \circ \Theta = \Psi \right\}.$$

In other words, $D(\Phi, \Psi) = \inf E(\rho(\mu, \nu))$, where the lower bound is taken over all pairs of random variables (μ, ν) with values in $(\mathscr{P}(E), \beta(\mathscr{P}(E)))$ such that μ has the distribution Φ and ν the distribution Ψ. The metric ρ being a bounded function, formula (8) defines a complete metric on $\mathscr{P}(\mathscr{P}(E))$, which gives to $\mathscr{P}(\mathscr{P}(E))$ the topology of weak convergence (see Theorem 2 in [23]). The proof of this last statement is very simple. We quote here its outline for the convenience of the reader. Let $(\Phi_N)_{N \geq 0}$, $\Phi_N \in \mathscr{P}(\mathscr{P}(E))$, $N \geq 0$, be a sequence of probability measures such that

$$\lim_{N \to +\infty} D(\Phi_N, \Phi) = 0, \qquad \Phi \in \mathscr{P}(\mathscr{P}(E)).$$

For every $F \in \mathscr{U}(\mathscr{P}(E))$ and $\varepsilon > 0$ there exists $\varrho(\varepsilon) > 0$ such that

$$\forall\ (\mu, \nu) \in \mathscr{P}(E) \times \mathscr{P}(E), \qquad |\rho(\mu, \nu)| \leq \varrho(\varepsilon) \Longrightarrow |F(\mu) - F(\nu)| \leq \varepsilon.$$

Let $(\mu, \mu_N)_{N \geq 1}$ be a sequence of measure-valued random variables on some probability space such that μ_N, $N \geq 1$, have distributions $\Phi_N \in \mathscr{P}(\mathscr{P}(E))$, $N \geq 1$ and μ is a measure-valued random variable with distribution $\Phi \in \mathscr{P}(\mathscr{P}(E))$. For every $F \in \mathscr{U}(\mathscr{P}(E))$ and $\varepsilon > 0$, one gets

$$\begin{aligned} |\Phi_N F - \Phi F| &\leq E(|F(\mu_N) - F(\mu)|) \\ &\leq \varepsilon + 2\|F\|\, P(\rho(\mu_N, \mu) > \varrho(\varepsilon)) \leq \varepsilon + 2\|F\| \varrho(\varepsilon)^{-1}\, E(\rho(\mu_N, \mu)), \end{aligned}$$

so

$$|\Phi_N F - \Phi F| \leq \varepsilon + \frac{2\|F\|}{\varrho(\varepsilon)}\, D(\Phi_N, \Phi).$$

Letting N go to infinity and recalling that, in the weak convergence (6), only bounded and uniformly continuous functions can be considered ([36], Theorem 6.1, page 40), we obtain our claim.

On the other hand, we can apply the monotone convergence theorem to prove that

$$D(\Phi_N, \Phi) \leq \sum_{m \geq 1} 2^{-(m+1)}\, E(|\mu_N f_m - \mu f_m|), \tag{9}$$

where $(f_m)_{m\geq 1}$ is the sequence of bounded and uniformly continuous functions introduced in (5). Hence, by the dominated convergence theorem, it follows that

$$\forall\ f \in \mathscr{U}(E), \qquad \lim_{N\to+\infty} E(|\mu_N f - \mu f|) = 0 \quad \Rightarrow \quad \lim_{N\to+\infty} D(\Phi_N, \Phi) = 0. \tag{10}$$

REMARK 1. It is interesting to note that if, in addition, μ is a constant probability distribution, the functions

$$F(\nu) = |\nu f - \mu f|, \qquad f \in \mathscr{C}_b(E)$$

are continuous for the weak convergence topology in $\mathscr{P}(E)$. So

$$\forall\ f \in \mathscr{U}(E), \qquad \lim_{N\to+\infty} E(|\mu_N f - \mu f|) = 0 \quad \Leftrightarrow \quad \lim_{N\to+\infty} D(\Phi_N, \Phi) = 0 \tag{11}$$

$$\Leftrightarrow \quad \Phi_N \xrightarrow[N\to+\infty]{w} \Phi. \tag{12}$$

It is also easily seen that

$$\forall\ f \in \mathscr{U}(E),$$
$$\lim_{N\to+\infty} E(|\mu_N f - \mu f|) = 0 \quad \Leftrightarrow \quad \lim_{N\to+\infty} E(|\mu_N f - \mu f|^2) = 0. \tag{13}$$

2.3. *Convergence of empirical measures.* In the study of Markov models of interacting particles, one looks at a N particle system $(\xi^{1,N}, \ldots, \xi^{N,N})$ satisfying a stochastic dynamical equation or more generally evolving according to a Markov transition probability kernel on a product space E^N, $N \geq 1$. Such models are used in fluid mechanics to study the many particled nature of real systems with internal fluctuations (see [22], [38], [40] and [44]), and in [18] the author proposes one way to use such models to solve numerically the so-called nonlinear filtering equation.

The problem of weak convergence in both situations consists in studying the limiting behavior of the empirical measures

$$\eta_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi^{i,N}} \tag{14}$$

as the number of particles N is growing. In the first situation the measures η_N are shown to converge in law, as $N \to +\infty$, to a constant probability measure η, which is called the McKean measure (see, for instance, [38]). Therefore, to prove convergence it is enough to verify (11).

In nonlinear filtering problems we will prove that the measures η_N converge in law, as $N \to +\infty$, to the desired conditional distribution. In this situation it is convenient to work in a first stage with a given sequence of observations and we will formulate the conditional distributions as a *nonrandom* probability measure η parameterized by the given sequence of observation parameters and the solution of a measure-valued dynamical system, which is usually called the nonlinear filtering equation. To prove convergence it is

enough to verify (11) P-a.s. for every sequence of observations and then apply the dominated convergence theorem.

In statistical physics and fluid mechanics the dynamical system (1) usually describes the time evolution of the density profiles of McKean–Vlasov stochastic processes with mean field drift functions. McKean and Vlasov proposed approximating the corresponding equations by mean field interacting particle systems. A crucial practical advantage of this situation is that the dynamical structure of the nonlinear stochastic process can be used in the design of an interacting particle system in which the mean field drift is replaced by a natural interaction function. Such models are called masters equations in physics and/or weakly interacting particle systems. In this situation it is convenient to use the following lemma.

LEMMA 1. *Let (Ω, F, P) be a probability space on which is defined a pair sequence $((\xi^{i,N})_{1\le i\le N}, (\bar{\xi}^{i,N})_{1\le i\le N})$ of N-particle systems so that the distribution u_N of $(\xi^{i,N})_{1\le i\le N}$ is a symmetric probability measure on E^N and $(\bar{\xi}^{i,N})_{1\le i\le N}$ are N i.i.d. with common law η. For every $f \in \mathscr{B}(E)$, we have*

$$\lim_{N\to+\infty} E\big(d(\xi^{1,N}, \bar{\xi}^{1,\mathbb{N}})\big) = 0 \quad \Rightarrow \quad \lim_{N\to+\infty} E\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\xi^{i,N}) - \eta f\right|\right) = 0. \tag{15}$$

PROOF. Let $\varrho(\varepsilon)$ be the modulus of continuity of the function $f \in \mathscr{B}(E)$. Then for every $\varepsilon > 0$ we have

$$\begin{aligned} E\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\xi^{i,N}) - \mu f\right|\right) &\le E\left(\frac{1}{N}\sum_{i=1}^{N} \left|f(\xi^{i,N}) - f(\bar{\xi}^{i,N})\right|\right) \\ &\quad + E\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\bar{\xi}^{i,N}) - \mu f\right|\right) \\ &\le \varepsilon + \frac{2\|f\|}{\varrho(\varepsilon)} E(d(\xi^{1,N}, \bar{\xi}^{1,N})) + \frac{2\|f\|}{\sqrt{N}}. \end{aligned}$$

Letting $N \to +\infty$, the lemma is proved. □

This lemma gives a simple condition for the convergence in law of the empirical measures when the interacting particle systems are described by a stochastic dynamical equation. More precisely, this powerful result can be used when the desired distribution η is the distribution of a finite-dimensional stochastic process. It will be applied in Section 3.3.2 to study the convergence in law of the empirical measures of weakly interacting particle systems. In nonlinear filtering problems, the dynamical system (1) describes the time evolution of the conditional distribution of the internal states in dynamical systems when partial observations are made. In contrast to the situation described above, the conditional distributions cannot be viewed as the law of a finite-dimensional stochastic process which incorporates a mean field drift [7]. We therefore have to find a new strategy to define an interacting particle sys-

tem which will approximate the desired distributions. We propose hereafter a new interacting particle system approach and another method to prove convergence.

3. Measure-valued processes. The chief purpose of this section is the design of two special models of stochastic particle systems for the numerical solving of a discrete time and measure-valued dynamical system η_n given by

$$\eta_n = \phi(n, \eta_{n-1}) \qquad \forall\, n \geq 1,\ \eta_0 = \eta, \tag{16}$$

where $\eta \in \mathscr{P}(E)$ and $\phi(n, \bullet)$: $\mathscr{P}(E) \to \mathscr{P}(E)$ are continuous functions. This section has the following structure: in Section 3.1 we describe a natural particle approximation with a simple interaction function. In this situation the interaction depends on the current positions but it does not depend on their paths. In Section 3.2 we introduce a general particle system which includes branching and interaction mechanisms. We emphasize that in both situations the nature of the interaction function is dictated by the plant equation (16). To illustrate our approach, we finish this section with practical examples for which all assumptions are satisfied.

3.1. *Interacting particle systems*

3.1.1. *The particle system state space.* The particle system under study will be a Markov chain with state space E^N, where $N \geq 1$ is the size of the system. The N-tuple of elements of E, that is, the points of the set E^N, are called particle systems and will be mostly denoted by the letters x, y, z.

The local dynamics of the system will be described by a product transition probability measure. Thus, to clarify the notations, with $\nu \in \mathscr{P}(E)$ we associate a measure $\nu^{\otimes N} \in \mathscr{P}(E^N)$ by setting

$$\nu^{\otimes N} = \underbrace{\nu \otimes \cdots \otimes \nu}_{N \text{ times}}$$

and we note $m^N(x)$ the empirical measure associated to the point $x = (x^1, \ldots, x^N)$:

$$m^N(x) = \frac{1}{N} \sum_{j=1}^{N} \delta_{x^j}.$$

3.1.2. *The associate Markov process.* Let $(\Omega, F_n, (\xi_n)_{n\geq 0}, P)$ be the E^N-valued Markov process defined by

$$P(\xi_0 \in dx) = \eta^{\otimes N}(dx), \qquad P(\xi_n \in dx / \xi_{n-1} = z) = \phi^{\otimes N}(n, m^N(z))(dx), \tag{17}$$

where $dx \overset{\text{def}}{=} dx^1 \times \cdots \times dx^N$ is an infinitesimal neighborhood of $x = (x^1, \ldots, x^N) \in E^N$. It is clear from the construction above that $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$ can be viewed as a system of N particles with nonlinear interaction function $\phi(n, m^N(\xi_{n-1}))$. The algorithm constructed in this way will be called an interacting particle approximation of (1). The term *interacting* is intended to emphasize that the particles are not independent and the

evolution of an individual particle is no longer Markovian. Nevertheless, the system as a whole is Markovian. Much more is true; the above description enables us to consider the particle density profiles

$$\eta_n^N \stackrel{\text{def}}{=} m^N(\xi_n)$$

as a measure-valued Markov process (Ω, F_n, η^N, P) defined by

$$P(\eta_0^N \in d\mu) = M_0 C_N(d\mu), \qquad P(\eta_n^N \in d\mu / \eta_{n-1}^N = \nu) = M_n C_N(\nu, d\mu), \tag{18}$$

where $d\mu$ is an infinitesimal neighborhood of μ and C_N and M_n, $n \geq 0$, are the Markov transitions on $\mathscr{P}(E)$ given by

$$M_n F(\nu) = F(\phi(n,\nu)), \qquad C_N F(\nu) = \int_{E^N} F(m^N(x))\, \nu^{\otimes N}(dx) \tag{19}$$

for every $F \in \mathscr{C}_b(\mathscr{P}(E))$, $n \geq 0$ and $\nu \in \mathscr{P}(E)$, with the convention $\phi(0,\nu) = \eta$, for all $\nu \in \mathscr{P}(E)$. To see this claim, it suffices to note that

$$\begin{aligned} E(F(\eta_n^N)/\eta_{n-1}^N = \nu) &= \int F(m^N(x))\, \phi^{\otimes N}(n,\nu)(dx) \\ &= C_N F(\phi(n,\nu)) = M_n(C_N F)(\nu) \end{aligned}$$

for all $F \in \mathscr{C}_b(\mathscr{P}(E))$ and $\nu \in \mathscr{P}(E)$.

3.1.3. *Description of the algorithm.* At the time $n = 0$ the particle system consists of N independent random particles $\xi_0^1, \ldots, \xi_0^N$ with common law η. At the time $n \geq 1$ the empirical measure $m^N(\xi_{n-1})$ associated to the particle system ξ_{n-1} enters in the plant equation (1) so that the resulting measure $\phi(n, m^N(\xi_{n-1}))$ depends on the configuration of the system at the previous time $n-1$. Finally, the particle system at the time n consists of N independent (conditionally to ξ_{n-1}) particles $\xi_n^1, \ldots, \xi_n^N$ with common law $\phi(n, m^N(\xi_{n-1}))$. We refer to Section 3.3 for further discussions and detailed examples.

3.1.4. *Convergence of the algorithm.* The crucial question is, of course, whether η_n^N converges to η_n as N is growing. When the state space E is compact we show hereafter a slightly more general result.

THEOREM 1. *Let us suppose that E is compact. Let $M = (M_n)_{n\geq 1}$ denote a series of time-inhomogeneous and Feller–Markov transitions on $\mathscr{P}(E)$, $M_0 \in \mathscr{P}(\mathscr{P}(E))$ and let $M_n^{(N)} = M_n C_N$, $n \geq 0$. Using these notations we have*

$$M_0^{(N)} M_1^{(N)} \cdots M_n^{(N)} \quad \xrightarrow[N\to+\infty]{w} M_0 M_1 \cdots M_n, \qquad \forall\ M_0 \in \mathscr{P}(\mathscr{P}(E)). \tag{20}$$

More generally, (20) *holds when E is locally compact and $M = (M_n)_{n\geq 1}$ such that*

$$M_n F \in \mathscr{U}(\mathscr{P}(E)) \qquad \forall\ F \in \mathscr{U}(\mathscr{P}(E)). \tag{21}$$

In order to prepare for its proof, we begin with a lemma.

LEMMA 2. *If E is compact, then for every $N \geq 1$, C_N is a Feller transition kernel and we have, for every $F \in \mathscr{C}_b(\mathscr{P}(E))$,*

$$\lim_{N \to +\infty} \|C_N F - F\| = 0. \tag{22}$$

When E is locally compact, (22) holds for every $F \in \mathscr{U}(\mathscr{P}(E))$.

To throw some light on convergence (22) and its connection with the law of large numbers, assume that the function F is given by

$$F(\eta) = |\eta f - \mu f| \qquad \forall\ \eta \in \mathscr{P}(E),$$

where $f \in \mathscr{C}_b(E)$ and $\mu \in \mathscr{P}(E)$. By a direct calculation, it follows from (22) that

$$\sup_{\nu \in \mathscr{P}(E)} E_\nu\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\xi^i) - \nu f\right|\right) \xrightarrow[N \to +\infty]{} 0,$$

where $(E^{\mathbb{N}}, \mathscr{E}^{\otimes \mathbb{N}}, (\xi^i)_{i \geq 1}, P_\nu)$ is a sequence of E-valued and independent random variables with common law ν and $E_\nu(\bullet)$ denotes the expectation with respect to P_ν. So Lemma 2 can be regarded as a uniform weak law of large numbers.

The proof of Theorem 1 is comparatively short; therefore we give it first.

PROOF OF THEOREM 1. For every $F \in \mathscr{C}_b(\mathscr{P}(E))$ we observe that

$$\begin{aligned}
&\|C_N M_1^{(N)} \cdots M_n^{(N)} F - M_1 \cdots M_n F\| \\
&\quad \leq \|C_N M_1 (C_N M_2^{(N)} \cdots M_n^{(N)} - M_2 \cdots M_n) F\| \\
&\qquad + \|C_N M_1 \cdots M_n F - M_1 \cdots M_n F\| \\
&\quad \leq \|C_N M_2^{(N)} \cdots M_n^{(N)} F - M_2 \cdots M_n F\| \\
&\qquad + \|C_N M_1 \cdots M_n F - M_1 \cdots M_n F\| \\
&\quad \leq \sum_{p=0}^{n-1} \|C_N M_{p+1} \cdots M_n F - M_{p+1} \cdots M_n F\| + \|C_N F - F\|.
\end{aligned} \tag{23}$$

Thus, using Lemma 2 and recalling that, in the weak convergence, only bounded uniformly continuous functions can be considered, the proof of Theorem 1 is straightforward. □

We come to the proof of Lemma 2.

PROOF OF LEMMA 2. Let us set

$$f(x^1, \ldots, x^N) = \prod_{i=1}^{N} f_i(x^i), \qquad f_1, \ldots, f_N \in \mathscr{C}_b(E)$$

and assume $\nu_n \in \mathscr{P}(E)$ is a sequence which weakly converges to $\nu \in \mathscr{P}(E)$ when n tends to infinity. Then, we obtain

$$\lim_{n\to+\infty} \prod_{i=1}^{N} \nu_n(f_i) = \prod_{i=1}^{N} \nu(f_i).$$

Since linear combinations of such functions are dense in $\mathscr{C}_b(E^N)$, one can check easily that C_N is Feller. Let $\tilde{F}(\nu) \stackrel{\text{def}}{=} F(\nu(f_1), \ldots, \nu(f_p))$ with $f_i \in \mathscr{C}_b(E)$ and F globally Lipschitz, that is

$$|F(x_1, \ldots, x_p) - F(x'_1, \ldots, x'_p)| \leq A \sum_{i=1}^{p} |x_i - x'_i|, \qquad A < +\infty.$$

Then we obtain

$$\begin{aligned} |C_N\tilde{F}(\nu) - \tilde{F}(\nu)| &\leq A \sum_{i=1}^{p} \int \left| \frac{1}{N} \sum_{j}^{N} f_i(x_j) - \nu(f_i) \right| \nu(dx_1) \cdots \nu(dx_N) \\ &\leq \frac{A}{\sqrt{N}} \sum_{i=1}^{p} \left(\nu(f_i^2) - \nu(f_i)^2\right)^{1/2} \xrightarrow[N\to+\infty]{} 0. \end{aligned} \tag{24}$$

Therefore, for some constant $B > 0$,

$$|C_N\tilde{F}(\nu) - \tilde{F}(\nu)| \leq \frac{B}{\sqrt{N}}.$$

Now, if E is compact, such $\tilde{F}$ are dense in $\mathscr{C}_b(\mathscr{P}(E))$ and C_N is Feller, $C_N F$ converges to F for every $F \in \mathscr{C}_b(\mathscr{P}(E))$.

Finally let us assume that E is locally compact. Let $\varrho(\varepsilon)$ be the modulus of continuity of the function $F \in \mathscr{U}(\mathscr{P}(E))$. Then for every $\varepsilon > 0$ we have

$$|C_N F(\nu) - F(\nu)| \leq \varepsilon + \frac{2\|F\|}{\varrho(\varepsilon)} \int \rho(\mu, \nu)\, C_N(\nu, d\mu).$$

In view of (5) it follows that

$$\begin{aligned} |C_N F(\nu) - F(\nu)| \leq \varepsilon + \frac{2\|F\|}{\varrho(\varepsilon)} \sum_{m\geq 1} 2^{-(m+1)} \\ \times \int_{E^N} \left| \frac{1}{N} \sum_{i=1}^{N} f_m(x_i) - \nu f_m \right| \nu(dx_1) \cdots \nu(dx_N), \end{aligned}$$

where $(f_m)_{m\geq 1}$ is a suitable sequence of uniformly continuous functions such that $\|f_m\| \leq 1$ for all $m \geq 1$. In the same way exactly as in (24) we can prove that

$$|C_N F(\nu) - F(\nu)| \leq \varepsilon + \frac{2\|F\|}{\varrho(\varepsilon)\sqrt{N}}.$$

The above inequality immediately implies the last assertion. This ends the proof of Lemma 2. □

Recall that the functions $\phi(n, \bullet)$ are continuous, so that the transitions defined by

$$M_n F(\nu) = F(\phi(n, \nu)) \qquad \forall\ \nu \in \mathscr{P}(E),\ \forall\ F \in \mathscr{C}_b(\mathscr{P}(E))$$

are Feller transition probability kernels. The interpretation of Theorem 1 is clear. The theorem states that under rather wide conditions,

$$\forall\ n \geq 0, \qquad \forall\ F \in \mathscr{C}_b(\mathscr{P}(E)), \qquad \lim_{N \to +\infty} E(F(\eta_n^N)) = F(\eta_n).$$

Applying this, one can obtain the limit of the moments of the particle density profile error

$$\forall\ n \geq 0,\ \forall\ f \in \mathscr{C}_b(E),\ \forall\ p \geq 1, \qquad \lim_{N \to +\infty} E(|\eta_n^N f - \eta_n f|^p) = 0.$$

This result can be regarded as a convergence theorem which vindicates the approach by semigroup techniques to a fairly general class of measure-valued dynamical systems. It will be applied in Section 3.3 to the so-called master equations of fluid mechanics.

Unfortunately, when the state space E is not compact or when the Feller transitions M_n, $n \geq 1$, do not satisfy condition (21) the question of convergence is quite difficult. In this situation we must check as usual the tightness of the laws of the random measures $m^N(\xi_n)$, $N \geq 1$, and identify all limit points as being concentrated on the desired measure η_n [2]. Thus very few substantive statements can be made about the convergence in view of the generality of our dynamical system (1).

Our next objective is to study an intermediate situation. More precisely, we introduce an additional assumption with regard to the functions $\phi(n, \bullet)$ which enables us to develop some useful theorem. In a little while we will see one way in which this result may be applied in nonlinear filtering problems.

THEOREM 2. *Suppose that for every $f \in \mathscr{C}_b(E)$, $\nu \in \mathscr{P}(E)$ and $n \geq 1$ there exist some constant $\alpha_n(\nu, f)$ and a finite set of bounded functions $\mathscr{H}_n(\nu, f)$ such that*

$$\forall\ \mu \in \mathscr{P}(E), \qquad |\phi(n, \nu)f - \phi(n, \mu)f| \leq \alpha_n(\nu, f) \sum_{h \in \mathscr{H}_n(\nu, f)} |\nu h - \mu h|. \tag{25}$$

Then, for every $f \in \mathscr{C}_b(E)$ and $n \geq 1$, there exists $A_n(f) < +\infty$ such that

$$E(|\eta_n^N f - \eta_n f|^2) \leq \frac{A_n(f)}{N}. \tag{26}$$

Therefore, if Φ_n^N is the distribution of η_n^N and $\Phi_n F(\mu) = F(\eta_n)$ for all $\mu \in \mathscr{P}(E)$, we have

$$\forall\ n \geq 1, \qquad \lim_{N \to +\infty} D(\Phi_n^N, \Phi_n) = 0, \tag{27}$$

where D is the Vaserstein metric introduced in (8).

Condition (25) strongly depends on the nature of the function $\phi(n,\bullet)$ which governs the dynamics of the distributions $(\eta_n)_{n\geq 0}$. Although this seems to be a very general condition, it may not rule out a certain kind of system of practical interest. For instance, the same does not work for the so-called master equations arising in fluid mechanics. Nevertheless, Theorem 1 may be applied to study these equations. Before proceeding to the proof of the theorem, let us examine some typical situations for which condition (25) is met.

1. Let us suppose that our functions $\phi(n,\bullet)$, $n\geq 1$, have the form $\phi(n,\mu)=\mu K_n$, where $(K_n)_{n\geq 0}$ is a family of Feller transition on E. Then we have, for every $f\in\mathscr{C}_b(E)$ and $\nu\in\mathscr{P}(E)$,

$$|\phi(n,\nu)f-\phi(n,\nu)f|\leq|\nu(K_nf)-\mu(K_nf)| \tag{28}$$

   for every $f\in\mathscr{C}_b(E)$ and $\mu,\nu\in\mathscr{P}(E)$. So, condition (25) is satisfied with $\alpha_n(\nu,f)=1$ and $\mathscr{H}_n(\nu,f)=\{K_nf\}$. Note that the set $\mathscr{H}_n(\nu,f)=\{K_nf\}$ does not depend on the measure ν.
2. At this point it is already useful to give some examples of measure-valued dynamical systems which will appear in nonlinear filtering problems. Next, suppose that the measure-valued dynamical system of interest is described by the equations

$$\eta_n=\phi_n(\eta_{n-1})\qquad\forall\ n\geq 1,\ \eta_0=\eta,$$

   where $\eta\in\mathscr{P}(E)$ and the continuous functions $\phi_n\colon\mathscr{P}(E)\to\mathscr{P}(E)$ are given by

$$\phi_n(\mu)f=\frac{\mu(g_n\ T_nf)}{\mu(g_n)}\qquad\forall\ n\geq 1,\ \forall\ f\in\mathscr{C}_b(E),\ \forall\ \mu\in\mathscr{P}(E), \tag{29}$$

   where $(T_n)_{n\geq 1}$ is a family of Feller transitions on E and $(g_n)_{n\geq 1}$ is a family of continuous and bounded functions such that

$$0<a_n\leq g_n(x)\leq A_n\qquad\forall\ x\in E,\ \forall\ n\geq 1 \tag{30}$$

   for some constants a_n and A_n, $n\geq 1$.

We immediately notice that this example is a generalization of the previous example [if $g_n(x)=1$ for all $x\in E$ then $\phi_n(\mu)=\mu T_n$]. Moreover, we will see in the last part of the paper that the functions (29) prescribe the dynamics structure of the optimal filter in nonlinear filtering problems. Now, we observe that

$$\begin{aligned}\phi_n(\mu)f-\phi_n(\nu)f&=\big(\phi_n(\mu)f-\phi_n(\nu)f\big)\left(\frac{\mu(g_n)}{\nu(g_n)}+\left(1-\frac{\mu(g_n)}{\nu(g_n)}\right)\right)\\&=\mu(h_n^{(1)})+\big(\phi_n(\mu)f-\phi_n(\nu)f\big)\,\mu(h_n^{(2)})\end{aligned}$$

with

$$h_n^{(1)}=\frac{g_n}{\nu(g_n)}\left(T_nf-\frac{\nu(g_n\ T_nf)}{\nu(g_n)}\right),\qquad h_n^{(2)}=1-\frac{g_n}{\nu(g_n)}.$$

It follows that

$$|\phi_n(\mu)f - \phi_n(\nu)f| \le (1+2\|f\|)\left(|\mu(h_n^{(1)}) - \nu(h_n^{(1)})| + |\mu(h_n^{(2)}) - \nu(h_n^{(2)})|\right). \tag{31}$$

and condition (25) is satisfied with $\alpha_n(\nu, f) = 1 + 2\|f\|$ and $\mathscr{H}_n(\nu, f) = \{h_n^{(1)}, h_n^{(1)}\}$. To see (31) it suffices to note that

$$\nu h_n^{(1)} = \nu h_n^{(2)} = 0 \quad \text{and} \quad \left|\frac{\mu(g_n\, T_n f)}{\mu(g_n)} - \frac{\nu(g_n\, T_n f)}{\nu(g_n)}\right| \le 2\|f\|.$$

PROOF OF THEOREM 2. Let us show (26) by induction on $n \ge 0$. Consider the first case $n = 0$. By the very definition of η_0^N, one gets easily

$$\begin{aligned} E(|\eta_0^N f - \eta_0 f|^2) &= E(|\eta_0^N(f - \eta_0 f)|^2) \\ &\le \int \left(\frac{1}{N}\sum_{i=1}^N (f(x^i) - \eta_0 f)\right)^2 \eta_0(dx^1)\cdots\eta_0(dx^N) \\ &\le \frac{(2\|f\|)^2}{N} \stackrel{\text{def}}{=} \frac{A_0(f)}{N}. \end{aligned}$$

Suppose the result is true at rank $(n-1)$. The assumption (25) implies the existence of a constant $\alpha_n(\eta_{n-1}, f) > 0$ and a finite set of bounded functions $\mathscr{H}_n(\eta_{n-1}, f)$ such that

$$\begin{aligned} E(|\eta_n^N f - \eta_n f|^2) &= E(|\phi(n, \eta_{n-1}^N)f - \phi(n, \eta_{n-1})f|^2) \\ &\quad + E(|f(\xi_n^1) - \phi(n, \eta_{n-1}^N)f|^2)/N \\ &\le E(|\phi(n, \eta_{n-1}^N)f - \phi(n, \eta_{n-1})f|^2) + (2\|f\|)^2/N \\ &\le \alpha_n(\eta_{n-1}, f)^2\, |\mathscr{H}_n(\eta_{n-1}, f)| \sum_{h\in\mathscr{H}_n(\eta_{n-1}, f)} E(|\eta_{n-1}^N h - \eta_{n-1} h|^2) \\ &\quad + (2\|f\|)^2/N, \end{aligned}$$

where $|\mathscr{H}_n(\eta_{n-1}, f)|$ is the cardinality of the set $\mathscr{H}_n(\eta_{n-1}, f)$. Now, the induction hypothesis at rank $(n-1)$ implies

$$E(|\eta_n^N f - \eta_n f|^2) \le \frac{1}{N}\, A_n(f)$$

with

$$A_n(f) = \alpha_n(\eta_{n-1}, f)^2\, |\mathscr{H}_n(\eta_{n-1}, f)| \sum_{h\in\mathscr{H}_n(\eta_{n-1}, f)} A_{n-1}(h) + (2\|f\|)^2$$

So the desired inequality (26) is true at rank n and the induction is completed. □

Our aim is now to get some information about the constant $A_n(f)$ in a special case arising in nonlinear filtering problems. Consider the dynamical

system described by (29). The discussion below closely follows [18]. Define the continuous functions

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

with the convention $\gamma_{n/n}(\mu) = \mu$ for all $\mu \in \mathscr{P}(E)$. Observe that

$$\gamma_{n/p}(\mu) f = \frac{\mu(g_{n/p}\ T_{n/p} f)}{\mu(g_{n/p})} \qquad \forall\ 0 \leq p \leq n-1,\ \forall\ f \in \mathscr{B}(E),\ \forall\ \mu \in \mathscr{P}(E), \tag{32}$$

where

$$\begin{aligned} g_{n/p-1} &= g_p\ T_p(g_{n/p}), \\ T_{n/p-1} f &= \frac{T_p(g_{n/p}\ T_{n/p} f)}{T_p(g_{n/p})} \qquad \forall\ 0 \leq p \leq n-1,\ \forall\ f \in \mathscr{B}(E) \end{aligned} \tag{33}$$

with the conventions $g_{n/n-1} = g_n$ and $T_{n/n-1} = T_n$. To prove this claim we first note that (32) is obvious for $p = n-1$ because $\gamma_{n/n-1} = \phi_n$. Now, using backward induction on the parameter p, if (32) is satisfied for a given value of $p \geq 1$ then we have

$$\begin{aligned} \gamma_{n/p-1}(\mu) f = \gamma_{n/p}(\phi_p(\mu)) f &= \frac{\phi_p(\mu)(g_{n/p}\ T_{n/p} f)}{\phi_p(\mu)(g_{n/p})} \\ &= \frac{\mu(g_p\ T_p(g_{n/p}\ T_{n/p} f))}{\mu(g_p\ T_p(g_{n/p}))} = \frac{\mu(g_{n/p-1}\ T_{n/p-1} f)}{\mu(g_{n/p-1})}, \end{aligned}$$

where $g_{n/p-1}$ and $T_{n/p-1}$ are given by (33). In the following we retain the notations of Theorem 2. Returning to the inequality (23), we have

$$\begin{aligned} &\left| M_0^{(N)} \cdots M_n^{(N)} F - M_0 \cdots M_n F \right| \\ &\quad \leq \sum_{p=0}^{n} \sup_{\nu \in \mathscr{P}(E)} \left| \int \left(F(\gamma_{n/p}(\mu)) - F(\gamma_{n/p}(\nu)) \right) C_N(\nu, d\mu) \right| \end{aligned}$$

for all $F \in \mathscr{B}(\mathscr{P}(E))$. Suppose now that

$$F(\mu) = |\mu f - \eta_n f|, \qquad f \in \mathscr{B}(E).$$

Then, one gets easily

$$E\left(|\eta_n^N f - \eta_n f|\right) \leq \sum_{p=0}^{n} \sup_{\nu \in \mathscr{P}(E)} \int \left|\gamma_{n/p}(\mu) f - \gamma_{n/p}(\nu) f\right| C_N(\nu, d\mu). \tag{34}$$

Arguing exactly as before, we have

$$\gamma_{n/p}(\mu) f - \gamma_{n/p}(\nu) f = \mu(h_{n/p}^{(1)}) + \left(\frac{\mu(g_{n/p}\ T_{n/p} f)}{\mu(g_{n/p})} - \frac{\nu(g_{n/p}\ T_{n/p} f)}{\nu(g_{n/p})} \right) \mu(h_{n/p}^{(2)})$$

with

$$h_{n/p}^{(1)} = \frac{g_{n/p}}{\nu(g_{n/p})} \left(T_{n/p} f - \frac{\nu(g_{n/p}\ T_{n/p} f)}{\nu(g_{n/p})} \right), \qquad h_{n/p}^{(2)} = 1 - \frac{g_{n/p}}{\nu(g_{n/p})}.$$

On the other hand,

$$\left|\gamma_{n/p}(\mu)f - \gamma_{n/p}(\nu)f\right| \leq \left|\mu(h^{(1)}_{n/p})\right| + 2\|f\| \left|\mu(h^{(2)}_{n/p})\right|$$

and a short calculation shows that

$$\int \left|\mu h^{(1)}_{n/p}\right|^2 C_N(\nu, d\mu) \leq \frac{(2\|f\|)^2}{N}\, \nu\left(\frac{g_{n/p}}{\nu(g_{n/p})}\right)^2$$

and

$$\int \left|\mu h^{(2)}_{n/p}\right|^2 C_N(\nu, d\mu) \leq \frac{1}{N}\, \nu\left(\frac{g_{n/p}}{\nu(g_{n/p})}\right)^2.$$

From this and (34) it follows that

$$E\left(\left|\eta^N_n f - \eta_n f\right|\right) \leq \frac{4\|f\|}{\sqrt{N}} \sum_{p=0}^{n} \sup_{\nu \in \mathscr{P}(E)} \left(\nu\left(\frac{g_{n/p}}{\nu(g_{n/p})}\right)^2\right)^{1/2} \leq \frac{4\|f\|}{\sqrt{N}} \sum_{p=0}^{n} \prod_{k=p+1}^{n} \frac{A_k}{a_k}$$

with the convention $\prod_{\varnothing} = 1$. Finally, we have proved that

$$\forall\ f \in \mathscr{B}(E), \qquad E\left(\left|\eta^N_n f - \eta_n f\right|\right) \leq \frac{B_n(f)}{\sqrt{N}} \quad \text{with } B_n(f) \stackrel{\text{def}}{=} 4n\|f\| \prod_{k=1}^{n} \frac{A_k}{a_k}.$$

In this situation we have obtained a uniform upper bound over the class of all measurable functions f with norm $\|f\| \leq 1$. This uniformity allows us to prove the following extension of the classical Glivenko–Cantelli theorem for interacting particle systems.

COROLLARY 1. *When the measure-valued dynamical system is given by* (29) *and* $E = \mathbb{R}$ *we have*

$$\forall\ n \geq 0, \qquad \lim_{N \to +\infty} \sup_{t \in \mathbb{R}} \left|\eta^N_n(]-\infty, t]) - \eta_n(]-\infty, t])\right| = 0, \qquad P\text{-a.s.} \tag{35}$$

PROOF. To prove (35) we first use the upper bound

$$P(|\eta^N_n(]-\infty, t]) - \eta_n(]-\infty, t])| \leq \varepsilon/2) \geq 1 - \frac{8\ n\ B_n}{\varepsilon\sqrt{N}} \quad \text{with } B_n = \prod_{k=1}^{n} \frac{A_k}{a_k}$$

for every $\varepsilon > 0$, $n \geq 0$ and $t \in \mathbb{R}$. Consequently, for $N \geq (16 n B_n/\varepsilon)^2$ we have

$$P(|\eta^N_n(]-\infty, t]) - \eta_n(]-\infty, t])| \leq \varepsilon/2) \geq 1/2 \qquad \forall\ \varepsilon > 0,\ \forall\ n \geq 0,\ \forall\ t \in \mathbb{R}.$$

Applying the symmetrization lemma ([37], page 14) it follows that

$$\begin{aligned} &P\Big(\sup_t \left|\eta^N_n(]-\infty, t]) - \eta_n(]-\infty, t])\right| > \varepsilon\Big) \\ &\qquad \leq 4E\Big(P\Big(\sup_t \left|\eta^{o,N}_n(]-\infty, t])\right| > \varepsilon/4 / \xi^1_n, \ldots, \xi^N_n\Big)\Big) \end{aligned} \tag{36}$$

where $\eta_n^{o,N}$ denotes the signed measure

$$\eta_n^{o,N} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i\, \delta_{\xi_n^i}$$

with $(\sigma_i)_{i\geq 1}$ an i.i.d. sequence with $P(\sigma_i = 1) = P(\sigma_i = -1) = 1/2$. Now, using Pollard's maximal inequality [37] and Hoeffding's lemma [25], we obtain the exponential bounds

$$P\Big(\sup_{t\in\mathbb{R}} \big|\eta_n^{o,N}(]-\infty, t])\big| > \varepsilon/4/\xi_n^1, \ldots, \xi_n^N\Big)$$
$$\leq 2\,(N+1)\exp(-N\varepsilon^2/32), \qquad P\text{-a.s.}$$

and the Borel–Cantelli lemma together with (36) turns this into (35). □

We now turn to the asymptotic normality of the particle approximation errors.

PROPOSITION 1. *Suppose the assumptions of Theorem* 2 *are satisfied. Then for every* $n \geq 0$, $f \in \mathscr{C}_b(E)$ *and* $x \in \mathbb{R}$ *we have*

$$\lim_{N\to+\infty} P\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(f(\xi_n^i) - \phi(n, \eta_{n-1}^N)f) \leq x\Big)$$
$$= \int_{-\infty}^{x} \frac{1}{\sigma_n(f)\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma_n^2(f)}\right) dz, \tag{37}$$

where

$$\sigma_n(f) = \big(\eta_n(f - \eta_n f)^2\big)^{1/2} \qquad \forall\, f \in \mathscr{C}_b(E),\ \forall\, n \geq 0. \tag{38}$$

Before proceeding, we should be more precise about the difficulty which arises here: this result gives some indications on the asymptotic behavior of the particle estimators $(\eta_n^N)_{n\geq 0}$ but the essential and unsolved problem is to characterize the asymptotical nature of the random errors

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(f(\xi_n^i) - \eta_n f) \qquad \forall\, f \in \mathscr{C}_b(E),\ \forall\, n \geq 0.$$

Other results relating to the asymptotical normality of particle approximation errors for specific models in continuous time can be found in [38].

PROOF. Using the above notations, we first observe that

$$E\left(\exp\left(\frac{i}{\sqrt{N}}\sum_{k=1}^{N}(f(\xi_n^k) - \phi(n, \eta_{n-1}^N)f)\right)\right)$$
$$= E\left(E\left(\exp\left(\frac{i}{\sqrt{N}}(f(\xi_n^1) - \phi(n, \eta_{n-1}^N)f)\right)/\eta_{n-1}^N\right)^N\right)$$
$$\stackrel{\text{def}}{=} E(B(n, N, f)^N),$$

where $i^2 = -1$. Using the classical inequality

$$\left| e^{i\,t} - \left(1 + i\,t - \frac{t^2}{2}\right)\right| \leq \frac{t^3}{6} \qquad \forall\ t \geq 0$$

and the fact that

$$E\big(f(\xi_n^1) - \phi(n, \eta_{n-1}^N)f / \eta_{n-1}^N\big) = 0$$

we get

$$\left| B(n, N, f) - \left(1 - \frac{1}{2N}\phi(n, \eta_{n-1}^N)(f - \phi(n, \eta_{n-1}^N)f)^2\right)\right| \leq \frac{4\|f\|^3}{3N\sqrt{N}},$$

which allows us to obtain

$$\big|B(n, N, f) - a(n, N, f)\big| \leq \frac{1}{N} b(n, N, f)$$

with

$$a(n, N, f) \stackrel{\text{def}}{=} 1 - \frac{1}{2N}\eta_n(f - \eta_n f)^2,$$

$$b(n, N, f) \stackrel{\text{def}}{=} \frac{1}{2}\big|\eta_n(f - \eta_n f)^2 - \phi(n, \eta_{n-1}^N)(f - \phi(n, \eta_{n-1}^N)f)^2\big| + \frac{4\|f\|^3}{3\sqrt{N}}.$$

Thus, we have

$$\big|B(n, N, f)^N - a(n, N, f)^N\big| \leq b(n, N, f)$$

and

$$\big|E(B(n, N, f)^N) - a(n, N, f)^N\big| \leq E(b(n, N, f)^2)^{1/2}.$$

On the other hand, let us remark that

$$\begin{aligned}
&\big|\eta_n(f - \eta_n f)^2 - \phi(n, \eta_{n-1}^N)(f - \phi(n, \eta_{n-1}^N)f)^2\big| \\
&\quad \leq \big|\eta_n(f^2) - \phi(n, \eta_{n-1}^N)(f^2)\big| + \big|(\eta_n f)^2 - (\phi(n, \eta_{n-1}^N)f)^2\big| \\
&\quad \leq \big|\eta_n(f^2) - \phi(n, \eta_{n-1}^N)(f^2)\big| + 2\|f\|\big|\eta_n f - \phi(n, \eta_{n-1}^N)f\big| \\
&\quad \leq \big(\big|\eta_n(f^2) - \phi(n, \eta_{n-1}^N)(f^2)\big| + \big|\eta_n f - \phi(n, \eta_{n-1}^N)f\big|\big)(1 + 2\|f\|).
\end{aligned}$$

Then we can write

$$\begin{aligned}
b(n, N, f) \leq &\left(\frac{1}{2} + \|f\| + \frac{4\|f\|^3}{3}\right) \\
&\times\left(\frac{1}{\sqrt{N}} + |\eta_n(f^2) - \phi(n, \eta_{n-1}^N)(f^2)| + |\eta_n f - \phi(n, \eta_{n-1}^N)f|\right).
\end{aligned} \tag{39}$$

We point out that the middle term in the second parentheses goes to zero by (25) and (26). It follows that

$$\lim_{N\to+\infty} E(B(n, N, f)^N) = \lim_{N\to+\infty} a(n, N, f)^N = \exp -\tfrac{1}{2}\eta_n(f - \eta_n f)^2.$$

This ends the proof of the proposition. □

Our final step is to provide some exponential bounds and to prove that $\eta_n^N f$ converges P-a.s. to $\eta_n f$ as N is the size of the systems growing, for every $n \geq 0$ and $f \in \mathscr{C}_b(E)$.

PROPOSITION 2. *Under the same conditions as in Theorem* 2 *we have*

$$\forall\ \varepsilon > 0,\ \forall\ n \geq 0,\ \forall\ f \in \mathscr{C}_b(E),$$

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\xi_n^i) - \eta_n f\right| > \varepsilon\right) \leq A_1(n,f)\exp(-N\varepsilon^2 A_2(n,f)) \tag{40}$$

with $A_1(n,f)$ *and* $A_2(n,f)$ *positive and finite. Then we have*

$$\forall\ n \geq 0,\ \forall\ f \in \mathscr{C}_b(E), \qquad \lim_{N\to+\infty} \eta_n^N f = \eta_n f, \qquad P\text{-a.s.} \tag{41}$$

PROOF. We first use assumption (25) to prove by induction that for every $f \in \mathscr{C}_b(E)$ and $n \geq 0$ there exist some constant $A(n,f)$ and some finite subset $\mathscr{L}(n,f) \subset \mathbb{N} \times \mathscr{C}_b(E)$ such that

$$\left|\eta_n^N f - \eta_n f\right| \leq A(n,f) \sup_{(k,h)\in\mathscr{L}(n,f)} \left|\eta_k^N h - \phi(k, \eta_{k-1}^N)h\right| \tag{42}$$

with the convention $\phi(0, \eta_{-1}^N) = \eta_0$.

Consider the case $n = 0$. For every $f \in \mathscr{C}_b(E)$, we have

$$\left|\eta_0^N f - \eta_0 f\right| = \left|\eta_0^N f - \phi(0, \eta_{-1}^N)f\right|$$

and (42) is satisfied at rank $n = 0$. Suppose the result is true at rank $n-1 \geq 0$. Observe that

$$\left|\eta_n^N f - \eta_n f\right| \leq \left|\eta_n^N f - \phi(n, \eta_{n-1}^N)f\right| + \left|\phi(n, \eta_{n-1}^N)f - \phi(n, \eta_{n-1})f\right| \qquad \forall\ n \geq 1.$$

Using (25) we get

$$\begin{aligned}\left|\eta_n^N f - \eta_n f\right| \leq{}& \left|\eta_n^N f - \phi(n, \eta_{n-1}^N)f\right| \\ &+ C_n(\eta_{n-1}, f)|\mathscr{H}(\eta_{n-1}, f)| \sup_{h\in\mathscr{H}_n(\eta_{n-1},f)} |\eta_{n-1}^N h - \eta_{n-1} h|.\end{aligned}$$

To clarify the presentation, we will note $C_n(f)$ and $\mathscr{H}_n(f)$ instead of $C_n(\eta_{n-1}, f)$ and $\mathscr{H}_n(\eta_{n-1}, f)$. The induction hypothesis at rank $n-1$ implies

$$\left|\eta_n^N f - \eta_n f\right| \leq A(n,f) \sup_{(k,h)\in\mathscr{L}(n,f)} \left|\eta_k^N h - \phi(k, \eta_{k-1}^N)h\right|$$

with

$$\begin{aligned}A(n,f) &= 2\,(1 \vee C_n(f)|\mathscr{H}_n(f)| \sup_{h\in\mathscr{H}_n(f)} A(n-1,h)), \\ \mathscr{L}(n,f) &= \{(n,f)\} \cup \bigcup_{h\in\mathscr{H}_n(f)} \mathscr{L}(n-1,h).\end{aligned}$$

The induction is thus completed. Apply Hoeffding's [25] inequality to get the upper bound

$$P\big(|\eta_k^N h - \phi(k, \eta_{k-1}^N)h| > \varepsilon/\eta_{n-1}^N\big) \le 2\exp -\frac{N}{8}(\varepsilon/\|h\|)^2, \quad P\text{-a.s.} \tag{43}$$

Therefore we have

$$P\big(|\eta_k^N h - \phi(k, \eta_{k-1}^N)h| > \varepsilon\big) \le 2\exp -\frac{N}{8}(\varepsilon/\|h\|)^2. \tag{44}$$

Combining (42) and (44) we obtain

$$P\left(\left|\frac{1}{N}\sum_{i=1}^{N} f(\xi_n^i) - \eta_n f\right| > \varepsilon\right) \le A_1(n, f)\exp(-N\varepsilon^2 A_2(n,\ f))$$

with

$$A_1(n, f) = 2|\mathscr{L}(n, f)| \quad \text{and} \quad A_2(n, f) = 1\Big/\Big(8A(n, f)^2 \sup_{(k,h)\in\mathscr{L}(n,f)} \|h\|^2\Big).$$

Equation (67) is a clear consequence of the Borel–Cantelli lemma and the proof is complete. □

3.2. *Interacting particle systems with branchings.* The latter is only concerned with simple interaction mechanisms therefore avoiding situations in which the interaction depends on parts of the trajectories of the particles. In this section we introduce a general approximation of (1) which includes branching and interaction mechanisms. Such constructions will solve (1) numerically when the state space E is given by

$$E = E_1 \times \cdots \times E_r \quad \text{with } r \ge 1,$$

where $(E_p)_{1\le p\le r}$ is a finite sequence of locally compact separable spaces. The basic idea is to split a probability measure on E for dealing with random trees. Nevertheless, it should be kept in mind that the content of this section is nothing more than an extension of the results of Section 3.1.

3.2.1. *The particle systems space.*

The particle system state space. Let us introduce some new notations. For every $N_1, \ldots, N_r \ge 1$, we note

$$(N)_q = \{1, \ldots, N_1\} \times \cdots \times \{1, \ldots, N_q\} \quad \text{for all } 1 \le q \le r. \tag{45}$$

The points of the sets $(N)_q$ will be denoted by the letters (i) or (j). For $(i) \in (N)_q$ and $1 \le p \le q$, i_p denotes the pth component of (i), so $(i) = (i_1, \ldots, i_q)$. The particle system under study will be a Markov chain with state space

$$E^{(N)} \stackrel{\text{def}}{=} E_1^{(N)_1} \times \cdots \times E_r^{(N)_r}, \qquad (N) \stackrel{\text{def}}{=} (N_1, \ldots, N_r). \tag{46}$$

For every $1 \le p \le r$, each point $x_p \in E_p^{(N)_p}$ consists of $|(N)_p|$ particles $x_p^{(i)}$,

$$x_p = (x_p^{(i)})_{(i)\in(N)_p} \in E_p^{(N)_p}. \tag{47}$$

Each point $x \in E^{(N)}$ consists of r particle systems

$$x = (x_1, \ldots, x_r) \in E_1^{(N)_1} \times \cdots \times E_r^{(N)_r}. \tag{48}$$

The points of $x \in E^{(N)}$ are called random trees.

It is important to remark that the size of the particle systems increases at each step of the algorithm:

$$|(N)_1| = N_1 \leq |(N)_2| = N_1 N_2 \leq \cdots \leq |(N)_r| = N_1 \cdots N_r.$$

At each step $1 < p \leq r$ the transitions under study will be branching mechanisms. More precisely, during a transition, each particle $\xi_{p-1}^{(i)}$, living in the system ξ_{p-1}, will branch into N_p auxiliary particles:

$$\forall (i) \in (N)_{p-1}, \qquad \xi_{p-1}^{(i)} \to (\xi_p^{(i),1}, \ldots, \xi_p^{(i),N_p}). \tag{49}$$

So at the end of this time, the resulting particle system ξ_p consists of $|(N)_p|$ random particles.

Note that, for every $1 \leq q \leq r$, a point $(x_1, \ldots, x_q)$ of the product space $E^{(N)_1} \times \cdots \times E^{(N)_q}$ may be viewed as a finite sequence

$$(x_1, \ldots, x_q) = \left(x_1^{i_1}, x_2^{i_1, i_2}, \ldots, x_q^{i_1, i_2, \ldots, i_q}\right)_{1 \leq i_1 \leq N_1, \ldots, 1 \leq i_q \leq N_q}.$$

Thus, from the above description (49), given a random tree $x \in E^{(N)}$, each q-tuple $(i_1, \ldots, i_q)$ will describe the *history* of the particle $x_q^{(i)}$. In order to describe the history of the particle $x_q^{(i)}$, it is convenient to introduce the following notations. For every $(i) = (i_1, \ldots, i_q) \in (N)_q$ and $1 \leq p \leq q \leq r$, we denote

$$(i)_p = i_1, \ldots, i_p.$$

For instance,

$$(i) = (i_1, \ldots, i_r) \in (N)_r \;\Rightarrow\; (i)_1 = (i_1), \quad (i)_2 = (i_1, i_2), \quad (i)_3 = (i_1, i_2, i_3).$$

Using these notations and the description (49), the branching dynamics which describe the evolution of the particle $x_r^{(i)} = x_r^{i_1, \ldots, i_r}$ is given by

$$x_1^{(i)_1} = x_1^{i_1} \to x_2^{(i)_2} = x_2^{i_1, i_2} \to \cdots \to x_{r-1}^{(i)_{r-1}} = x_{r-1}^{i_1, \ldots, i_{r-1}} \to x_r^{(i)} = x_r^{i_1, \ldots, i_r}.$$

Transition probability kernels. Let ν be a probability measure on $E_1 \times \cdots \times E_r$. For every $1 \leq p \leq r$, we define, whenever they exist, a probability measure $\nu_p \in \mathscr{P}(E_1 \times \cdots \times E_p)$ and a transition probability kernel $\nu_{p/p-1}$ by setting, with some obvious abusive notations,

$$\begin{aligned} \nu(dx_1, \ldots, dx_r) &= \nu_1(dx_1)\, \nu_{2/1}(x_1, dx_2) \cdots \nu_{r/r-1}(x_1, \ldots, x_{r-1}, dx_r), \\ \nu_p(dx_1, \ldots, dx_p) &= \nu_{p-1}(dx_1, \ldots, dx_{p-1})\, \nu_{p/p-1}(x_1, \ldots, x_{p-1}, dx_p). \end{aligned} \tag{50}$$

The existence of such splittings is discussed in [13], Chapter III.

With $\nu \in \mathscr{P}(E)$ and (50) we associate a measure $\nu^{\otimes(N)} \in \mathscr{P}(E^{(N)})$ by

$$\nu^{\otimes(N)}(dx) = \prod_{i_1=1}^{N_1} \nu_1(dx_1^{i_1}) \prod_{i_2=1}^{N_2} \nu_{2/1}(x_1^{i_1}, dx_2^{i_1,i_2}) \cdots \times \prod_{i_r=1}^{N_r} \nu_{r/r-1}(x_1^{i_1}, \ldots, x_{r-1}^{i_1,\ldots,i_{r-1}}, dx_r^{i_1,\ldots,i_r}).$$

Let us see an example of the use of these formulas.

EXAMPLE 1. Let $r = 2$, $E_1 = E_2$ and $\nu = \mu \times K$ where $\mu \in \mathscr{P}(E_1)$ and K is a transition probability kernel on E_1. Since in this case

$$\nu_1(dx_1) = \mu(dx_1) \quad \text{and} \quad \nu_{2/1}(x_1, dx_2) = K(x_1, dx_2)$$

one obtains

$$\nu^{\otimes(N)}(dx) = \prod_{i_1=1}^{N_1} \mu(dx_1^{i_1}) \prod_{i_2=1}^{N_2} K(x_1^{i_1}, dx_2^{i_1,i_2}). \tag{51}$$

It follows from (51) that $\nu^{\otimes(N)}$ is the probability distribution of a random tree

$$\xi = (\xi_1^{i_1}, \xi_2^{i_1,i_2})_{1 \le i_1 \le N_1, 1 \le i_2 \le N_2},$$

where we have the following:

(i) here $(\xi_1^{i_1})_{1 \le i_1 \le N_1}$ are N_1 independent and identically distributed random variables with common law μ;

(ii) for each $1 \le i_1 \le N_1$, $(\xi_2^{i_1,i_2})_{1 \le i_2 \le N_2}$ are N_2 independent and identically distributed random variables with common law $K(\xi_1^{i_1}, \bullet)$.

We observe that the probability measure $\nu^{\otimes(N)}$ may be split as before and written in a more compact and simple form. By the very definition of the sets $(N)_1, \ldots, (N)_r$, we have

$$\nu^{\otimes(N)}(dx) = \nu_1^{\otimes(N)_1}(dx_1)\, \nu_{2/1}^{\otimes(N)_2}(x_1, dx_2) \cdots \nu_{r/r-1}^{\otimes(N)_r}(x_1, \ldots, x_{r-1}, dx_r)$$

with

$$\begin{aligned} \nu_{p/p-1}^{\otimes(N)_p}(x_1, \ldots, x_{p-1}, dx_p) &= \prod_{(i) \in (N)_p} \nu(x_1^{(i)_1}, \ldots, x_{p-1}^{(i)_{p-1}}, dx_p^{(i)_p}) \\ &= \prod_{(j) \in (N)_{p-1}} \prod_{i_p=1}^{N_p} \nu(x_1^{(j)_1}, \ldots, x_{p-1}^{(j)_{p-1}}, dx_p^{(j), i_p}) \end{aligned} \tag{52}$$

Using these notations we introduce the transition probability kernels $C_{(N)}$ as follows:

$$\begin{aligned} C_{(N)}F(\nu) &= \int_{E^{(N)}} F(m^{(N)}(x))\, \nu^{\otimes(N)}(dx) \quad \text{with} \\ m^{(N)}(x) &\overset{\text{def}}{=} \frac{1}{|(N)_r|} \sum_{(i) \in (N)_r} \delta_{(x_1^{(i)_1}, \ldots, x_r^{(i)_r})} \end{aligned} \tag{53}$$

for all $\nu \in \mathscr{P}(E)$ and $F \in \mathscr{C}_b(\mathscr{P}(E))$. Roughly speaking, starting with a measure $\nu \in \mathscr{P}(E)$, the transition probability $C_{(N)}$ chooses randomly a measure $m^{(N)}(\xi)$ where ξ is an $E^{(N)}$-valued random variable with law $\nu^{\otimes(N)}$. From our notations, $m^{(N)}(\xi) \in \mathscr{P}(E_1 \times \cdots \times E_r)$ and it has the form

$$m^{(N)}(\xi) = \frac{1}{|(N)_r|} \sum_{i_1,\ldots,i_r=1}^{N_1,\ldots,N_r} \delta_{\xi_1^{i_1},\ldots,\xi_r^{i_1,\ldots,i_r}}.$$

Further manipulations yield the decomposition

$$m^{(N)}(\xi)(dz) = m_1^{(N)_1}(\xi)(dz_1)\, m_{2/1}^{(N)_2}(\xi)(z_1, dz_2) \cdots m_{r/r-1}^{(N)_r}(\xi)(z_{r-1}, dz_r) \tag{54}$$

with

$$m_1^{(N)_1}(\xi)_1(dz_1) = \frac{1}{N_1} \sum_{i_1=1}^{N_1} \delta_{\xi_1^{i_1}}(dz_1) \tag{55}$$

and

$$\begin{aligned} &m_{p/p-1}^{(N)_p}(\xi)(z_{p-1}, dz_p) \\ &\quad = \sum_{(i)\in(N)_{p-1}} 1_{\xi_{p-1}^{(i)}}(z_{p-1}) \frac{1}{N_p} \sum_{i_p=1}^{N_p} \delta_{\xi_p^{(i),i_p}}(dz_p) \qquad \forall\ 1 < p < r. \end{aligned} \tag{56}$$

The decompositions (52) and (56) make clearer the nature of the transition probability kernel $C_{(N)}$. The recursive description of the random tree $\xi = (\xi_1, \ldots, \xi_r)$ with law $\nu^{\otimes(N)}$ is straightforward.

1. *Step* $p = 1$. The particle system ξ_1 consists of N_1 i.i.d. random variables with common law ν_1.
2. *Step* $1 < p \le r$. At the end of step $(p-1)$, the random tree consists of $(p-1)$ particle systems $\xi_1, \ldots, \xi_{p-1}$. In particular the system ξ_{p-1} contains $|(N)_{p-1}|$ particles. During this transition each particle $\xi_{p-1}^{(i)}$ branches into a fixed number N_p of independent particles $(\xi_p^{(i),1}, \ldots, \xi_p^{(i),N_p})$ with common law $\nu_{p/p-1}(\xi_1^{(i)_1}, \ldots, \xi_{p-1}^{(i)_{p-1}}, du)$:

$$\xi_{p-1}^{(i)} \in E_{p-1} \to (\xi_p^{(i),1}, \ldots, \xi_p^{(i),N_p}) \in E_p^{N_p} \quad \text{i.i.d.} \sim \nu_{p/p-1}(\xi_1^{(i)_1}, \ldots, \xi_{p-1}^{(i)_{p-1}}, du).$$

REMARK 2. If $(N) = (N_1, 1, \ldots, 1)$, then it is clear from the above that for every $\nu \in \mathscr{P}(E)$,

$$E^{(N)} = E_1^{N_1} \times \cdots \times E_r^{N_1} = E^{N_1},$$

$$m^{(N)}(x) = m^{N_1}(x) = \frac{1}{N_1} \sum_{i_1=1}^{N_1} \delta_{x_1^{i_1},\ldots,x_r^{i_1}} \quad \text{and} \quad C_{N_1} = C_{(N)},$$

$$\nu^{\otimes(N)}(dx) = \prod_{i_1=1}^{N_1} \nu_1(dx_1^{i_1})\nu_{2/1}(x_1^{i_1}, dx_2^{i_1}) \dots \nu_{r/r-1}(x_1^{i_1}, \dots, x_{r-1}^{i_1}, dx_r^{i_1})$$
$$= \prod_{i_1=1}^{N_1} \nu(dx^{i_1}) = \nu^{\otimes N_1}(dx).$$

3.2.2. *The associate Markov process.* We further require that for every $z \in E^{(N)}$ and $n \geq 1$, the measure $\phi(n, m^{(N)}(z))$ can be split as in (50). At this point, it is appropriate to give a special case which will appear in nonlinear filtering problems where the splitting of the measure $\phi(n, m^{(N)}(z))$ does not present much more difficulty.

EXAMPLE 2. Let us suppose that $r = 2$, $E_1 = E_2$ and $\phi(n, \bullet)$ is given by $\forall\ f \in \mathscr{C}_b(E^2),\ \forall\ \mu \in \mathscr{P}(E^2)$,

$$\phi(n, \mu)f = \frac{\int f(x_1, x_2)\ g_n(x_1)\ \mu(dx_0, dx_1)\ K(x_1, dx_2)}{\int g_n(z_1)\ \mu(dz_0, dz_1)}.$$

Since in this case

$$\forall\ z \in E_1^{(N)_1} \times E_2^{(N)_2}, \qquad m^{(N)}(z) = \frac{1}{N_1 N_2} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \delta_{(z_1^{i_1}, z_2^{i_1, i_2})},$$

we obtain

$$\phi(n, m^{(N)}(z))(dx_1, dx_2) = \sum_{i_1, i_2=1}^{N_1, N_2} \frac{g_n(z_2^{i_1, i_2})}{\sum_{j_1, j_2=1}^{N_1, N_2} g_n(z_2^{j_1, j_2})}\ \delta_{z_2^{i_1, i_2}}(dx_1), \quad K(x_1, dx_2)$$

and therefore

$$\phi(n, m^{(N)}(z))(dx_1, dx_2) = \phi_1(n, m^{(N)}(z))(dx_1)\ \phi_{2/1}(n, m^{(N)}(z))(x_1, dx_2)$$

with

$$\phi_1(n, m^{(N)}(z))(dx_1) = \sum_{i_1, i_2=1}^{N_1, N_2} \frac{g_n(z_2^{i_1, i_2})}{\sum_{j_1, j_2=1}^{N_1, N_2} g_n(z_2^{j_1, j_2})}\ \delta_{z_2^{i_1, i_2}}(dx_1)$$
$$\phi_{2/1}(n, m^{(N)}(z))(x_1, dx_2) = K(x_1, dx_2)$$

We refer again to Section 3.3 for more detailed examples which explain the *splitting assumption* imposed in the construction of the branching transitions.

We are now ready to introduce a particle approximation of (1) which includes branching and interaction mechanisms. Let $N_1 \geq 1, \dots, N_r \geq 1$ and let $(\Omega, (F_n)_n, (\xi_n)_n, P)$ be the $E^{(N)}$-valued Markov process defined by

$$P(\xi_0 \in dx) = \eta^{\otimes(N)}(dx),$$
$$P(\xi_n \in dx / \xi_{n-1} = z) = \phi^{\otimes(N)}(n, m^{(N)}(z))(dx). \tag{57}$$

In Section 4 we will apply the above constructions to solve nonlinear filtering problems. In such a framework, the above transitions will have an explicit and simple form. The algorithm constructed in this way will be called an interacting particle system with branchings. The term *branching* is intended to emphasize that the points of the state space of the Markov chain are random trees.

It is useful at this point to stress the Markov description of the empirical measures

$$\eta_n^{(N)} \stackrel{\text{def}}{=} m^{(N)}(\xi_n).$$

Arguing as in Section 3.2.2, the above description enables us to consider the random measures $\eta_n^{(N)}$ as the measure-valued random process $(\Omega, (F_n)_n, (\eta_n^{(N)})_n, P)$ defined by

$$\begin{aligned} P\big(\eta_0^{(N)} \in d\mu\big) &= M_0 C_{(N)}(d\mu), \\ P\big(\eta_n^{(N)} \in d\mu / \eta_{n-1}^{(N)} = \nu\big) &= M_n C_{(N)}(\nu, d\mu) \end{aligned} \tag{58}$$

with $M_0 F(\nu) = F(\eta)$ for all ν.

3.2.3. *Description of the algorithm.* If we want to think in terms of branching and interaction mechanisms, it is essential to recall that at each time n and for every $(i) = (i_1, \ldots, i_r) \in (N)_r$ we have

$$\xi_n^{(i)} = \big(\xi_{n,1}^{i_1}, \xi_{n,2}^{i_1, i_2}, \ldots, \xi_{n,r}^{i_1, i_2, \ldots, i_r}\big).$$

Therefore $\xi_n^{(i)}$ can be viewed as the nth part of the trajectory of an individual particle

$$\xi_{n,1}^{(i)_1} \to \xi_{n,2}^{(i)_2} \to \ldots \to \xi_{n,r-1}^{(i)_{r-1}} \to \xi_{n,r}^{(i)_r}.$$

In addition, in view of (52), at each step $1 < p \le r$, each particle $\xi_{n,p-1}^{(i)}$ will branch into N_p auxiliary particles:

$$\forall\, (i) \in (N)_{p-1}, \qquad \xi_{n,p-1}^{(i)} \to (\xi_{n,p}^{(i),1}, \ldots, \xi_{n,p}^{(i),N_p}).$$

In the same spirit, the $E^{(N)}$-valued Markov process as a whole can be viewed as a branching process.

Notice that each particle system $\xi_{n,p}$, $1 \le p \le n$, contains $|(N)_p| = N_1 \ldots N_p$ particles. So the size of the particle systems $\xi_{n,1}, \ldots, \xi_{n,r}$ increase at each step of the $p = 1, \ldots, n$, but at the end of the interval, the next particle system $\xi_{n+1,1}$ only contains $|(N)_1| = N_1$ particles. Probabilistically and in more precise language, we may describe its evolution in time as follows:

1. At the time $n = 0$:

   Step $p = 1$. The particle system $\xi_{0,1}$ consists of $N_1 = |(N)_1|$ random particles $\xi_{0,1} = (\xi_{0,1}^{(i)})_{(i)\in(N)_1}$ with the same distribution η_1.

Step $1 < p \leq r$. At the end of step $p-1$ the random tree consists of $p-1$ particle systems $\xi_{0,1}, \ldots, \xi_{0,p-1}$. In particular, the particle system $\xi_{0,p-1}$ contains $|(N)_{p-1}|$ particles $\xi_{0,p-1}^{(i)}$. In the very beginning, each particle $\xi_{0,p-1}^{(i)}$ branches into a fixed number N_p of particles:

$$\xi_{0,p-1}^{(i)} \in E_{p-1} \to (\xi_{0,p}^{(i),1}, \ldots, \xi_{0,p}^{(i),N_p}) \in E_p^{N_p},$$
$$\text{i.i.d.} \sim \eta_{p/p-1}(\xi_{0,1}^{(i)_1}, \ldots, \xi_{0,p-1}^{(i)_{p-1}}, du).$$

Therefore at the end of these mechanisms the particle system $\xi_{0,p}$ contains $|(N)_p|$ particles.

2. At time $n \geq 1$.

At time $n-1$ the random tree ξ_{n-1} consists of r particle systems $\xi_{n-1} = (\xi_{n-1,1}, \ldots, \xi_{n-1,r})$. For all $p = 1, \ldots, r$, each particle system $\xi_{n-1,p}$ contains $|(N)_p|$ particles $\xi_{n-1,p} = (\xi_{n-1,p}^{(i)})_{(i)\in(N)_p}$.

Step $p = 1$. The particle system $\xi_{n,1}$ consists of $|(N)_1|$ random particles $\xi_{n,1}^{(i)}$ with the same distribution $\phi_1(n, m^{(N)}(\xi_{n-1}))$.

Step $1 < p \leq r$. At the end of step $p-1$ the random tree consists of $p-1$ particle systems $\xi_{n,1}, \ldots, \xi_{n,p-1}$. In particular, the particle system $\xi_{n,p-1}$ consists of $|(N)_{p-1}|$ particles $\xi_{n,p-1}^{(i)}$. In the very beginning each particle $\xi_{n,p-1}^{(i)}$ branches into a fixed number N_p of random particles:

$$\xi_{n,p-1}^{(i)} \to (\xi_{n,p}^{(i),1}, \ldots, \xi_{n,p}^{(i),N_p})$$
$$\text{i.i.d.} \sim \phi_{p/p-1}\big(n, m^{(N)}(\xi_{n-1})\big)(\xi_{n,1}^{(i)_1}, \ldots, \xi_{n,p-1}^{(i)_{p-1}}, du).$$

So at the end of these mechanisms, the particle system $\xi_{n,p}$ contains $|(N)_p|$ particles.

3.2.4. *Convergence of the algorithm.* The following discussion is an easy generalization of that given in Section 3.1 and for these reasons proofs will only be sketched.

PROPOSITION 3. *Let us suppose that E is compact. Let $M = (M_n)_{n\geq 1}$ denote a series of time-inhomogeneous and Feller–Markov transitions on $\mathscr{P}(E)$, $M_0 \in \mathscr{P}(\mathscr{P}(E))$ and let $M_n^{(N)} = M_n C_{(N)}$, $n \geq 0$. Using these notations we have*

$$M_0^{(N)} M_1^{(N)} \ldots M_n^{(N)} \xrightarrow[N\to+\infty]{w} M_0 M_1 \cdots M_n \qquad \forall\, \Phi_0 \in \mathscr{P}(\mathscr{P}(E)). \tag{59}$$

More generally, (59) *holds when E is locally compact and $M = (M_n)_{n\geq 1}$ such that*

$$M_n F \in \mathscr{U}(\mathscr{P}(E)) \qquad \forall\, F \in \mathscr{U}(\mathscr{P}(E)). \tag{60}$$

SKETCH OF PROOF. The proof which we sketch here is based on the same kind of arguments used in the proof of Theorem 1. Arguing as before, we can

show that the transition probability kernel $C_{(N)}$ is a Feller transition. Now, let $\tilde{F}(\nu) =_{\text{def}} F(\nu(f_1), \ldots, \nu(f_q))$ with $f_k \in \mathscr{C}_b(E)$ and F globally Lipschitz, that is:

$$|F(x_1, \ldots, x_q) - F(x'_1, \ldots, x'_q)| \leq A \sum_{k=1}^{q} |x_i - x'_i|, \qquad A < +\infty.$$

By the very definition of $C_{(N)}$ one gets easily the system of inequalities

$$\begin{aligned}
&|C_{(N)}\tilde{F}(\nu) - \tilde{F}(\nu)| \\
&\quad \leq A \sum_{k=1}^{q} \int_{E^{(N)}} |m^{(N)}(x) f_k - \nu(f_k)| \, \nu^{\otimes(N)}(dx) \\
&\quad \leq A \sum_{k=1}^{q} \int \left| \frac{1}{|(N)_r|} \sum_{(i)\in(N)_r} f_k(x_1^{(i)_1}, \ldots, x_r^{(i)_r}) - \nu(f_k) \right| \nu^{\otimes(N)}(dx_1, \ldots, dx_r).
\end{aligned}$$

Therefore using the form of $(x_1^{(i)_1}, \ldots, x_r^{(i)_r})$, one gets after some standard computations,

$$\begin{aligned}
|C_{(N)}\tilde{F}(\nu) - \tilde{F}(\nu)| &\leq A \sum_{k=1}^{q} \left(\sum_{p=1}^{r} \frac{1}{|(N)_p|} \nu_p((\nu_{r/p} f_k - \nu_{r/p-1} f_k)^2) \right)^{1/2} \\
&= A \sum_{k=1}^{q} \left(\sum_{p=1}^{r} \frac{1}{|(N)_p|} \nu_p(\nu_{r/p} f_k)^2 - \nu((\nu_{r/p-1} f_k)^2) \right)^{1/2} \\
&\leq \frac{A}{\sqrt{N_1}} \sum_{k=1}^{q} \left(\sum_{p=1}^{r} \nu_p(\nu_{r/p} f_k)^2 - \nu((\nu_{r/p-1} f_k)^2) \right)^{1/2} \\
&= \frac{A}{\sqrt{N_1}} \sum_{k=1}^{q} (\nu(f_k - \nu f_k)^2)^{1/2}
\end{aligned} \tag{61}$$

with the convention $\nu_{r/r} f = f$ for all $f \in \mathscr{C}_b(E)$. The proof of (61), while straightforward, is somewhat lengthy so it is omitted. Then there exists some constant B such that

$$|C_{(N)}\tilde{F}(\nu) - \tilde{F}(\nu)| \leq \frac{B}{\sqrt{N_1}}.$$

The statement follows using (23) and recalling that such functions $\tilde{F}$ are dense in $\mathscr{C}_b(\mathscr{P}(S))$. □

In particular, when $M_n F(\nu) = F(\phi(n, \nu))$, the above result implies that

$$\forall\, n \geq 1, \qquad \forall\, f \in \mathscr{C}_b(E),$$

$$\lim_{N_1 \to +\infty} E(|\eta_n^{(N)} f - \eta_n f|) = 0 \quad \text{where } (N) = (N_1, \ldots, N_r).$$

The problem is now to find an explicit upper bound for the rate of convergence. Similarly to Theorem 2 and using inequality (61), a crude upper bound may be derived when condition (25) is satisfied.

PROPOSITION 4. *Suppose that for every* $f \in \mathscr{C}_b(E)$, $\nu \in \mathscr{P}(E)$ *and* $n \geq 1$ *there exist some constant* $C_n(\nu, f)$ *and a finite set of bounded functions* $\mathscr{H}_n(\nu, f)$ *such that*

$$\forall\ \mu \in \mathscr{P}(E), \qquad |\phi(n,\nu)f - \phi(n,\mu)f| \leq C_n(\nu,f) \sum_{h \in \mathscr{H}_n(\nu,f)} |\nu h - \mu h|. \tag{62}$$

Then, for every $f \in \mathscr{C}_b(E)$ *and* $n \geq 1$, *there exists* $A_n(f) < +\infty$ *such that*

$$E(|\eta_n^{(N)} f - \eta_n f|^2) \leq \frac{A_n(f)}{N_1}. \tag{63}$$

Therefore, if $\Phi_n^{(N)}$ *is the distribution of* $\eta_n^{(N)}$ *and* $\Phi F(\mu) = F(\eta_n)$ *for all* $\mu \in \mathscr{P}(E)$ *we have*

$$\forall\ n \geq 1, \qquad \lim_{N_1 \to +\infty} D(\Phi_n^{(N)}, \Phi_n) = 0. \tag{64}$$

To get exponential bounds in this situation, we prove an extension of Hoeffding's inequality [25].

LEMMA 3. *Let* $(\xi^{(i)})_{(i)\in(N)}$ *be an* $E^{(N)}$*-valued random tree with law* $\nu^{\otimes(N)}$ *where* $(N) = (N_1, \ldots, N_r)$, $N_1 \geq 1, \ldots, N_r \geq 1, r \geq 1$ *and* $\nu \in \mathscr{P}(E_1 \times \cdots \times E_r)$. *For each* $f \in \mathscr{C}_b(E_1 \times \cdots \times E_r)$ *and* $\varepsilon > 0$,

$$\begin{aligned} &P\left(\left|\frac{1}{|(N)|} \sum_{(i)\in(N)} (f(\xi^{(i)}) - \nu f)\right| > \varepsilon\right) \\ &\qquad \leq 2\exp\left(-\frac{1}{8}(\varepsilon/\|f\|)^2 \left(\sum_{k=1}^{r} \frac{1}{|(N)_k|}\right)^{-1}\right) \end{aligned} \tag{65}$$

PROOF. Our method of proof follows that of Hoeffding. For each $f \in \mathscr{C}_b(E_1 \times \cdots \times E_r)$, $(i) \in (N)$ and $1 \leq k \leq r$, set

$$f_k^{(i)_k} = E(f(\xi^{(i)})/\xi^{(i)_k}) - E(f(\xi^{(i)})/\xi^{(i)_{k-1}}), \qquad \xi^{(i)_k} \stackrel{\text{def}}{=} (\xi^{i_1}, \xi^{i_1,i_2}, \ldots, \xi^{i_1,\ldots,i_k}).$$

Using these notations, we have

$$f(\xi^{(i)}) - \nu f = \sum_{k=1}^{r} f_k^{(i)_k} \qquad \forall\ (i) \in (N).$$

Write $L_N(t,f)^r$ the moment generating functions

$$\begin{aligned} L_N(t,f)^r &\stackrel{\text{def}}{=} E\left(\exp \frac{t}{|(N)_r|} \sum_{(i)_r \in (N)_r} (f(\xi^{(i)}) - \nu f)\right) \\ &= E\left(\prod_{(i)_r \in (N)_r} \exp\left(\frac{t}{|(N)_r|} \sum_{k=1}^{r} f_k^{(i)_k}\right)\right). \end{aligned}$$

It is well known that $E(\exp(tX)) \leq \exp((tb)^2/2)$ for every $t \in \mathbb{R}$ and every real-valued random variable X with zero mean and bounded ranges $|X| \leq b$.

Applying this inequality to each $f_r^{(i)_r}$, $(i)_r \in (N)$, conditionally to $\xi^{(i)_{r-1}}$ we obtain

$$L_{(N)}(t,f)^r \leq L_{(N)}(t,f)^{r-1} \exp(2t^2\|f\|^2/|(N)_r|).$$

Using the same technique, repeatedly we obtain the upper bound

$$L_{(N)}(t,f)^r \leq \exp\left(2\,t^2\,\|f\|^2 \sum_{k=1}^{r} \frac{1}{|(N)_k|}\right).$$

Thus, for each $\varepsilon > 0$ and $t \in \mathbb{R}$,

$$P\left(\frac{1}{|(N)|}\sum_{(i)\in(N)}(f(\xi^{(i)}) - \nu f) > \varepsilon\right) \leq \exp\left(-\varepsilon\, t + 2\, t^2\, \|f\|^2 \sum_{k=1}^{r} \frac{1}{|(N)_k|}\right).$$

To minimize the quadratic, let $t = \varepsilon/(4\|f\|^2 \sum_{k=1}^{r} 1/|(N)_k|)$. This yields

$$P\left(\frac{1}{|(N)|}\sum_{(i)\in(N)}(f(\xi^{(i)}) - \nu f) > \varepsilon\right) \leq \exp\left(-\frac{1}{8}(\varepsilon/\|f\|)^2\left(\sum_{k=1}^{r} \frac{1}{|(N)_k|}\right)^{-1}\right).$$

The end of the proof is now straightforward. □

Using the same lines of arguments as in the proof of Proposition 2, Lemma 3 leads immediatly to the following result.

PROPOSITION 5. *Under the same conditions as in Theorem* 2, *for every* $\varepsilon > 0$, $n \geq 0$ *and* $f \in \mathscr{C}_b(E)$, *we have*

$$P\big(|\eta_n^{(N)} f - \eta_n f| > \varepsilon\big) \leq A_1(n,f) \exp\left(-\varepsilon^2 A_2(n,f)\left(\sum_{k=1}^{r} \frac{1}{|(N)_k|}\right)^{-1}\right) \tag{66}$$

with $A_1(n,f)$ *and* $A_2(n,f)$ *positive and finite. Then we have*

$$\forall\, n \geq 0, \quad \forall\, f \in \mathscr{C}_b(E), \qquad \lim_{N_1 \to +\infty} \eta_n^{(N)} f = \eta_n f, \qquad P\text{-a.s.} \tag{67}$$

Our goal now is to discuss the connections between Proposition 4 and Theorem 2. Let $(\eta_n^{N_1})_n$ be the density profiles associated to the particle approximation with simple interactions and N_1 particles and let $(\eta_n^{(N)})_n$ be the density profiles associated to the interacting particle resolution with branchings, where $(N) = (N_1, \ldots, N_r)$ and $N_1, \ldots, N_r \geq 1$. We have already remarked that the particle systems with simple interactions and the interacting particle systems with branchings are exactly the same when the number of auxiliary branching particles at each step equals one. More precisely,

$$(N) = (N_1, 1, \ldots, 1) \quad \Rightarrow \quad C_{(N)} = C_{N_1} \quad \text{and} \quad \eta_n^{(N)} = \eta_n^{N_1}\ \forall\, n \geq 0.$$

Let us discuss the relationship between the above approximations in the situation where the functions $\phi(n, \bullet)$, $n \geq 1$, are defined by (29). The discussion of Example 2 can be extended in an obvious way to the interacting particle

approximation with branchings and it can be seen easily that the moments of the corresponding particle density profile errors are given by

$$E(|\eta_n^{N_1} f - \eta_n f|) \le \sum_{p=0}^{n} \sup_{\nu \in \mathscr{P}(E)} I_{n/p}^{N_1}(\nu, f),$$

$$E(|\eta_n^{(N)} f - \eta_n f|) \le \sum_{p=0}^{n} \sup_{\nu \in \mathscr{P}(E)} I_{n/p}^{(N)}(\nu, f)$$

with

$$I_{n/p}^{N_1}(\nu, f) = \int |\gamma_{n/p}(\mu) f - \gamma_{n/p}(\nu) f| \, C_{N_1}(\nu, d\mu),$$

$$I_{n/p}^{(N)}(\nu, f) = \int |\gamma_{n/p}(\mu) f - \gamma_{n/p}(\nu) f| \, C_{(N)}(\nu, d\mu).$$

In order to estimate the terms $I_{n/p}^{N_1}(\nu, f)$ we introduced previously two functions $h_{n/p}^{(1)}$ and $h_{n/p}^{(2)}$ such that $\nu h_{n/p}^{(1)} = 0 = \nu h_{n/p}^{(2)}$ and we proved that

$$\begin{aligned} I_{n/p}^{N_1}(\nu, f) &\le \int |\mu h_{n/p}^{(1)}| \, C_{N_1}(\nu, d\mu) + 2\|f\| \int |\mu h_{n/p}^{(2)}| \, C_{N_1}(\nu, d\mu) \\ &\le \frac{1 + 2\|f\|}{\sqrt{N_1}} \left(\nu((h_{n/p}^{(1)})^2)^{1/2} + \nu((h_{n/p}^{(2)})^2)^{1/2}\right). \end{aligned} \tag{68}$$

Similarly, one gets

$$I_{n/p}^{(N)}(\nu, f) \le \int |\mu h_{n/p}^{(1)}| \, C_{(N)}(\nu, d\mu) + 2\|f\| \int |\mu h_{n/p}^{(2)}| \, C_{(N)}(\nu, d\mu)$$

and the same computations as in (61) lead to

$$I_{n/p}^{(N)}(\nu, f) \le (1 + 2\|f\|) \sum_{i=1}^{2} \left(\sum_{q=1}^{r} \frac{1}{N_1 \cdots N_q} \nu\left(\nu_{r/q} h_{n/p}^{(i)} - \nu_{r/q-1} h_{n/p}^{(i)}\right)^2 \right)^{1/2}. \tag{69}$$

One way to see that the term on the right-hand side of (69) is lower than the one of (68) is to remark that

$$\begin{aligned} \sum_{q=1}^{r} \frac{1}{N_1 \cdots N_q} \nu\left(\nu_{r/q} h_{n/p}^{(i)} - \nu_{r/q-1} h_{n/p}^{(i)}\right)^2 &\le \frac{1}{N_1} \sum_{q=1}^{r} \nu\left(\nu_{r/q} h_{n/p}^{(i)} - \nu_{r/q-1} h_{n/p}^{(i)}\right)^2 \\ &= \frac{1}{N_1} \sum_{q=1}^{r} \left(\nu(\nu_{r/q} h_{n/p}^{(i)})^2 - \nu(\nu_{r/q-1} h_{n/p}^{(i)})^2\right) \\ &= \frac{1}{N_1} \nu(h_{n/p}^{(i)} - \nu h_{n/p}^{(i)})^2 = \frac{1}{N_1} \nu(h_{n/p}^{(i)})^2. \end{aligned}$$

We conclude that, using the branching mechanisms, it is possible to get a lower bound for the particle density moment errors but whether or not much loss of performance is incurred by one of the above algorithms is an interesting but unsolved theoretical question. Really effective methods for attacking such a problem are apparently not known.

3.3. *Examples.* One important and useful application of our techniques is the situation in which the dynamical system (1) has the form

$$\phi(n,\mu)f = \frac{\psi(n,\mu)(f)}{\psi(n,\mu)(1)} \qquad \forall\ f \in \mathscr{C}_b(E),\ \forall\ \mu \in \mathscr{P}(E), \tag{70}$$

where $\psi(n,\bullet)$ are continuous and measure-valued linear functions; that is,

$$\forall\ \mu_1,\mu_2 \in \mathscr{P}(E),\ \forall\ \alpha_1,\alpha_2 \in \mathbb{R},\ \forall\ n \geq 1,$$
$$\psi(n,\alpha_1\,\mu_1 + \alpha_2\,\mu_2) = \alpha_1\ \psi(n,\mu_1) + \alpha_2\ \psi(n,\mu_2).$$

This property is of particular interest because of its relation to the nonlinear filtering problem.

In addition to the role of interacting particle systems in nonlinear filtering theory, there are several important points of contact between our approach and nonlinear systems arising in fluid mechanics. The second part of this section is devoted to the study of such systems. These simple Markovian models of particles are called master equations in physics. Incidentally, our approach provides convergence results for the empirical measures of such interacting particle systems. For a more thorough treatment of these equations see [44] and the references given there.

The setting is the same as in the previous sections. The particle systems will be a Markov chain with state space E^N where $N \geq 1$ indicates the size of the system. The N-tuple of elements of E, that is, the points of the set E^N, are called systems of particles and will be mostly denoted by the letters x, y, z. As usual $m^N(x)$ denotes the empirical measure associated to the point $x = (x^1,\ldots,x^N) \in E^N$:

$$m^N(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}.$$

3.3.1. *Linear systems compositions.* The following examples illustrate our interacting particle system approach and highlight issues specific to nonlinear filtering problems. Let $(K_n)_{n\geq 0}$ be a family of Feller–Markov kernels and let $(g_n)_{n\geq 1}$ be a sequence of continuous functions $g_n\colon E \to \mathbb{R}$.

1. If $\phi(n,\eta) = \eta K_n$ for all $n \geq 1$ then for every $x \in E^N$

$$\phi(n, m^N(x)) = m^N(x)K_n = \frac{1}{N}\sum_{i=1}^{N} K_n(x^i,\bullet)$$

and, the transition probability kernels of the corresponding particle system are given by

$$P(\xi_n \in dz/\xi_{n-1} = x) = \prod_{p=1}^{N} \frac{1}{N}\sum_{i=1}^{N} K_n(x^i, dz^p)$$

In other words the particles are chosen randomly and independently in the previous system and, in a second step, they move independently of each other according to the transitions $(K_n)_n$.

2. Our second example concerns the dynamical plant equation (1) when the functions $\phi(n, \bullet)$ are given by

$$\phi(n, \mu)f = \mu(f\ g_n)/\mu(g_n) \qquad \forall\ f \in \mathscr{C}_b(E),\ \forall\ \mu \in \mathscr{P}(E).$$

In this situation, for every $x \in E^N$,

$$\phi(n, m^N(x)) = \sum_{i=1}^{N} \frac{g_n(x^i)}{\sum_{j=1}^{N} g_n(x^j)}\, \delta_{x^i}.$$

Thus, the transition probability kernels of the corresponding particle system are given by

$$P(\xi_n \in dz/\xi_{n-1} = x) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(x^i)}{\sum_{j=1}^{N} g_n(x^j)}\, \delta_{x^i}(dx^p).$$

In other words, at the time n the particles are chosen randomly and independently in the previous system according to the fitness functions g_n.

3. Let us study a way of combining situations (1) and (2). Let us set $E = E_1 \times E_2$, where E_1 and E_2 are two locally compact separable spaces. Suppose the dynamical plant equation (1) is given with the functions $\phi(n, \bullet)$ defined by

$$\phi(n, \mu)f = \frac{\int f(x_1, x_2)\ g_n(x_1)\ \mu(dx_0, dx_1)\ K(x_1, dx_2)}{\int g_n(z_1)\ \mu(dz_0, dz_1)} \tag{71}$$

for every $f \in \mathscr{C}_b(E_1 \times E_2)$ and $\mu \in \mathscr{P}(E_1 \times E_2)$. In this situation, for every $x = (x_1, x_2) \in E^N = E_1^N \times E_2^N$,

$$\phi(n, m^N(x))(du, dv) = \sum_{i=1}^{N} \frac{g_n(x_2^i)}{\sum_{j=1}^{N} g_n(x_2^j)} \delta_{x_2^i}(du) K(u, dv)$$

and the particles will move according to the transitions

$$P(\zeta_n \in d(x_1, x_2)/\zeta_{n-1} = (z_1, z_2)) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(z_2^i)}{\sum_{j=1}^{N} g_n(z_2^j)} \delta_{z_2^i}(dx_1^p) K(x_1^p, dx_2^p).$$

To be more precise, let us set $\zeta_n = (\widehat{\xi}_n, \xi_{n+1})$. Using this notation and the above description, the motion of the particles is decomposed into two separate mechanisms:

$$P(\xi_n \in dx_2/\widehat{\xi}_{n-1} = x_1) = \prod_{p=1}^{N} K(x_1^p, dx_2^p), \tag{72}$$

$$P(\widehat{\xi}_n \in dx_1/\xi_n = x_2) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(x_2^i)}{\sum_{j=1}^{N} g_n(x_2^j)} \delta_{x_2^i}(dx_1^p). \tag{73}$$

4. Finally we examine the above situation (71) when we use a particle system which includes branching and interaction mechanisms. When considering the dynamical system (71) with state space $E = E_1 \times E_2$, the particle system is modeled by a Markov chain with state space $E^{(N)} = E_1^{(N)_1} \times E_2^{(N)_2}$ where $(N) = (N_1, N_2)$ with $N_1, N_2 \geq 1$.

Recall that $m^{(N)}(x)$ is the empirical measure associated to the point $x = (x^{(i)})_{(i)\in(N)_2} \in E^{(N)}$:

$$m^{(N)}(x) = \frac{1}{|(N)_2|} \sum_{(i)\in(N)_2}, \qquad \delta_{x^{(i)}} = \frac{1}{N_1 N_2} \sum_{i_1}^{N_1} \sum_{i_2=1}^{N_2} \delta_{(x_1^{i_1}, x_2^{i_1,i_2})}.$$

Observe that, for every $x \in E^{(N)}$,

$$\phi(n, m^{(N)}(x))(du, dv) = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \frac{g_n(x_2^{i_1,i_2})}{\sum_{j_1=1}^{N_1} \sum_{i_2=1}^{N_2} g_n(x_2^{j_1,j_2})} \delta_{x_2^{i_1,i_2}}(du) K(u, dv)$$

Thus, the transition probability kernels of the corresponding $E_1 \times E_2$-valued particles is given by

$$P(\zeta_n \in dz / \zeta_{n-1} = x) = \prod_{p_1=1}^{N_1} \sum_{i_1,i_2=1}^{N_1,N_2} \frac{g_n(x_2^{i_1,i_2})}{\sum_{j_1,j_2=1}^{N_1,N_2} g_n(x_2^{j_1,j_2})} \delta_{x_2^{i_1,i_2}}(dz_1^{p_1}) \prod_{p_2=1}^{N_2} K(z_1^{p_1}, dz_2^{p_1,p_2}).$$

Let us set $\zeta_n = (\widehat{\xi}_n, \xi_{n+1}) \in E^{N_1} \times E_2^{N_1 N_2}$. Using this notation and the above description, the motion of the particles is decomposed into two separate mechanisms:

$$P(\xi_n \in dx_2 / \widehat{\xi}_{n-1} = x_1) = \prod_{p_1=1}^{N_1} \prod_{p_2=1}^{N_2} K(x_1^{p_1}, dx_2^{p_1,p_2}), \tag{74}$$

$$P(\widehat{\xi}_n \in dx_1 / \xi_n = x_2) = \prod_{p_1=1}^{N_1} \sum_{i_1,i_2=1}^{N_1,N_2} \frac{g_n(x_2^{i_1,i_2})}{\sum_{j_1,j_2=1}^{N_1,N_2} g_n(x_2^{j_1,j_2})} \delta_{x_2^{i_1,i_2}}(dx_1^{p_1}). \tag{75}$$

To be more precise, let us remark that

$$P(\widehat{\xi}_n \in dx_1 / \xi_n = x_2) = \prod_{p_1=1}^{N_1} \sum_{i_1=1}^{N_1} \frac{\sum_{k_2=1}^{N_2} g_n(x_2^{i_1,k_2})}{\sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} g_n(x_2^{j_1,j_2})} \sum_{i_2=1}^{N_2} \frac{g_n(x_2^{i_1,i_2})}{\sum_{k_2=1}^{N_2} g_n(x_2^{i_1,k_2})} \delta_{x_2^{i_1,i_2}}(dx_1^{p_1}).$$

Roughly speaking, each particle $\widehat{\xi}_n^{p_1}$ chooses a subsystem $(x_2^{i_1,k})_{1\leq k\leq N_2}$, with $1 \leq i_1 \leq N_1$, at random with probability

$$\sum_{k_2=1}^{N_2} g_n(x_2^{i_1,k_2}) \Big/ \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} g_n(x_2^{j_1,j_2})$$

and moves to the site $x_2^{i_1,i_2}$ with probability $g_n(x_2^{i_1,i_2}) / \sum_{k_2=1}^{N_2} g_n(x_2^{i_1,k_2})$.

3.3.2. *Master equations.* The following systems are called in physics weakly interacting systems, because the interaction depends only on a fixed function of the empirical measures m^N. We have only considered here very elementary equations which can easily be generalized. The continuous versions, with not necessarily compact state space, are studied in [3], [33], [32], [43], [44] and the references given there.

Compact state space. In order to use Theorem 1, we first made the sanguine assumption that the state space is compact. Although such is generally not the case, these artificial examples will serve their purpose in illuminating the effect of interaction in real systems.

(1) Our first example concerns the dynamical system (1) when $E = [0, 1]$ and the functions $\phi(n, \bullet)$ are given by

$$\phi(n, \mu)f = \int f(x + F(x, V * \mu(x)) + w)\, d\mu(x)\, d\Gamma(w) \qquad \forall\ f \in \mathscr{C}_b([0, 1]),$$

where the sign + means summation modulo 1, $\Gamma \in \mathscr{P}([0, 1])$ and V and F are continuous functions $V\colon [0, 1] \to \mathbb{R}$, $F\colon [0, 1] \times \mathbb{R} \to [0, 1]$.

Roughly speaking, the solution of this dynamical system

$$\eta_n = \phi(n, \eta_{n-1}) \qquad \forall\ n \geq 1,\ \eta_0 \in \mathscr{P}(E) \tag{76}$$

describes the time marginal distribution of the time inhomogeneous, $[0, 1]$-valued, and nonlinear process ξ defined by the recursive equation

$$\begin{aligned} &X_n - X_{n-1} = F(X_{n-1}, V * \eta_{n-1}(X_{n-1})) + W_n, n \geq 1, \\ &X_0 \sim \eta_0, \end{aligned} \tag{77}$$

where η_{n-1} is the distribution of X_{n-1} and W_n is a $[0, 1]$-valued random variable with law Γ. As we shall see, this process describes the limit behavior of the trajectory of an individual particle in an interacting particle system as the number of particles is growing. It is usually called, in propagation of chaos theory, the tagged particle process. To be more precise, it is well known that there exists a measure $\eta_\bullet$ in the path space $\in \mathscr{P}(E^{n+1})$, called a McKean measure (or McKean process) corresponding to the set of transitions $\{K_u;\ u \in \mathscr{P}(E)\}$ defined by

$$K_u f(x) = \int f(x + F(x, V * u(x)) + w)\, d\Gamma(w) \qquad \forall\ f \in \mathscr{C}_b(E)$$

such that $(E^{n+1}, \mathscr{E}^{\otimes n+1}, X_\bullet = (X_k)_{0 \leq k \leq n}, \eta_\bullet)$ is a time inhomogeneous Markov process with transitions $(K_{u_k})_{0 \leq k \leq n}$ and under η, the probability distribution of x_k is u_k for all $0 \leq k \leq n$.

In our settings, this measure is clearly given by

$$\eta(dx_0, \ldots, dx_{n+1}) = u_0(dx_0)\, K_{u_0}(x_0, dx_1) \cdots K_{u_n}(x_n, dx_{n+1})$$

with

$$u_0 = \eta_0 \quad \text{and} \quad \forall\ 1 \leq k \leq n, \qquad u_k = u_{k-1} K_{u_{k-1}}.$$

The description of such models in continuous time may be found in [32]. In this situation, the McKean probability measure on the path space is defined as the solution of a classical martingale problem. Therefore, the time marginal distributions η_n, $n \geq 0$, of the McKean measure η satisfy (76). The existence of a unique McKean measure is discussed in [43] and [38].

One classical problem is to estimate the the time marginal distributions η_n. Using (17) and Theorem 1 we are able to construct an interacting particle approximation. First, we observe that for every $x = (x^1, \ldots, x^N) \in E^N$,

$$\phi(k, m^N(x))f = \frac{1}{N}\sum_{i=1}^{N} \int f(x^i + F(x^i, (V * m^N(x))(x^i)) + w)\, d\Gamma(w).$$

With regard to (17), the system of particles is driven by the mechanisms

$$\begin{aligned} &P(\widehat{\xi}_k \in dx / \xi_k = z) = \prod_{p=1}^{N} \frac{1}{N}\sum_{i=1}^{N} \delta_{z^i}(dx^p) \\ &\xi_k^i = \widehat{\xi}_{k-1}^{\,i} + F\big(\widehat{\xi}_{k-1}^{\,i}, (V * m^N(\xi_{k-1}))(\widehat{\xi}_{k-1}^{\,i})\big) + W_k^i, \qquad 1 \leq i \leq N, \end{aligned} \tag{78}$$

where $(W_k^i)_{1 \leq i \leq N}$ are i.i.d. with common law Γ, $\widehat{\xi}_0 = (\widehat{\xi}_0^{\,1}, \ldots, \widehat{\xi}_0^{\,N})$ are i.i.d. with common law η_0, $\xi_k = (\xi_k^1, \ldots, \xi_k^N)$, $\widehat{\xi}_k = (\widehat{\xi}_k^{\,1}, \ldots, \widehat{\xi}_k^{\,N}) \in [0,1]^N$.

There remains the question of convergence of the particle density profiles

$$\eta_n^N \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^i}, \qquad n \geq 1.$$

Using Theorem 1, we conclude easily that for every $n \geq 1$ the random measures η_n^N converge in law to η_n when the size of the system is growing.

(2) Our second example concerns the dynamical system (1) when the functions $\phi(n, \bullet)$ are given by

$$\phi(n, \mu)f = \int f\left(x + \int a(x, z)\, \mu(dz) + \int b(x, z)\, \mu(dz)\, w\right) d\mu(x)\, d\Gamma(w)$$

$$\forall\, f \in \mathscr{C}_b([0,1]),$$

where the sign $+$ means summation modulo 1, $\Gamma \in \mathscr{P}([0,1])$ and a and b are continuous functions a: $[0,1]^2 \to \mathbb{R}$, b: $[0,1]^2 \to \mathbb{R}$. In this situation, the solution of the dynamical system

$$\eta_n = \phi(n, \eta_{n-1}) \qquad \forall\, n \geq 1, \qquad \eta_0 \in \mathscr{P}(E)$$

is the density profile of the McKean measure associated to the time inhomogeneous and $[0,1]$-valued nonlinear process

$$\begin{aligned} &X_n - X_{n-1} = \overline{a}_n(X_{n-1}) + \overline{b}_n(X_{n-1})\, W_n, \\ &X_0 \sim \eta_0, \end{aligned} \tag{79}$$

where we have the following:

(a) $\overline{a}_n(z) = E(a(z, X_{n-1})) = \int a(z, x)\ \eta_{n-1}(dx)$ and $\overline{b}_n(z) = E(b(z, X_{n-1})) = \int b(z, x)\ \eta_{n-1}(dx)$ for all $n \geq 1$ and $x \in [0, 1]$;
(b) η_{n-1} is the distribution of X_{n-1};
(c) W_n is an $[0, 1]$-valued random variable with law Γ.

The description of such models in continuous time may be found in [44]. Arguing as before, the corresponding interacting particle system is given by

$$
\begin{aligned}
&P(\widehat{\xi}_k \in dx/\xi_k = z) = \prod_{p=1}^{N} \frac{1}{N} \sum_{i=1}^{N} \delta_{z^i}(dx^p), \\
&\xi_n^i = \widehat{\xi}_{n-1}^i + \frac{1}{N} \sum_{j=1}^{N} a(\widehat{\xi}_{n-1}^{\,i}, \xi_{n-1}^j) + \frac{1}{N} \sum_{j=1}^{N} b(\widehat{\xi}_{n-1}^{\,i}, \xi_{n-1}^j)\ W_n^i, \qquad 1 \leq i \leq N
\end{aligned}
\tag{80}
$$

with $(W_n^i)_{1 \leq i \leq N}$ i.i.d. variables with common law Γ.

Using Theorem 1, the random measures $\eta_n^N \stackrel{\text{def}}{=} 1/N \sum_{i=1}^{N} \delta_{\xi_n^i}$ converge in law to η_n when the size of the system is growing.

(3) Let E be a compact separable state space and let $\eta_\bullet$ be a McKean process corresponding to a given set of Feller transitions $\{K_u;\ u \in \mathscr{P}(E)\}$ and to a given distribution $\eta \in \mathscr{P}(E)$. In addition, we assume that the maps

$$u \in \mathscr{P} \to \delta_x\, K_u \in \mathscr{P}$$

are continuous, for all $x \in E$. Recalling the above observations, we see that the density profile η_n is a solution of the dynamical system

$$\eta_n = \phi(n, \eta_{n-1}) \qquad \forall\ n \geq 1,\ \ \eta_0 = \eta \in \mathscr{P}(E),$$

where

$$\forall\ n \geq 1, \quad \forall\ u \in \mathscr{P}(E),\ \forall\ f \in \mathscr{C}_b(E), \qquad \phi(n, u)f \stackrel{\text{def}}{=} \int u(dx)\, K_u(x, dy)\, f(y).$$

Now, in view of (17), the transition probability kernels of the interacting particle system are defined by

$$P(\xi_n \in dx/\xi_{n-1} = z) = \prod_{p=1}^{N} \frac{1}{N} \sum_{i=1}^{N} K_{m^N(z)}(z^i, dx^p)$$

with

$$m^N(z) = \frac{1}{N} \sum_{j=1}^{N} \delta_{z^j} \in \mathscr{P}(E),$$

where $x = (x^1, \ldots, x^N)$, $z = (z^1, \ldots, z^N) \in E^N$ and $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$.

Finally, using Theorem 1 and the above conditions we can show easily that the random measures

$$\eta_n^N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i}, \qquad n \geq 1$$

converge in law to η_n when the size of the system is growing.

Noncompact state space. In this last example we study in a different, but more classical way, the asymptotic behavior of a system of interacting particles when the state space is not necessarily compact. The continuous version without interaction through the perturbation can be founded in [44]. Let $E = \mathbb{R}^d$, $d \geq 1$, and let η be the McKean measure on the path space E^{T+1}, $T \in \mathbb{N}^\star$, corresponding to the time-inhomogeneous and E-valued nonlinear process.

$$\begin{aligned} &X_n = \overline{F}_n(X_{n-1}, W_n), \qquad 1 \leq n \leq T, \\ &X_0 \quad \text{with law } \eta_0, \end{aligned} \tag{81}$$

where we have the following:

1. for every $z, w \in E$, $\overline{F}_n(z,w) = \int F_n(z,x,w)\,\eta_{n-1}(dx)$ where $F_n\colon E^3 \to E$ are bounded Lipschitz functions for all $1 \leq n \leq T$;
2. η_n is the distribution of X_n, for all $0 \leq n \leq T$;
3. $W = (W_n)_n$ is a sequence of E-valued and independent random variables;
4. X_0 is an independent of W random variable with distribution $\eta_0 \in \mathscr{P}(E)$.

In this situation, one looks at a pair system of N-particles,

$$\begin{aligned} &\xi_n^{i,N} = \tfrac{1}{N}\textstyle\sum_{j=1}^N F_n(\xi_{n-1}^{i,N}, \xi_{n-1}^{j,N}, W_n^i), \qquad 1 \leq i \leq N, \qquad 1 \leq n \leq T, \\ &\overline{\xi}_n^{i,N} = \overline{F}_n(\overline{\xi}_{n-1}^{i,N}, W_n^i), \qquad 1 \leq i \leq N, \qquad 1 \leq n \leq T, \end{aligned}$$

where W^i, $1 \leq i \leq N$, are independent copies of the perturbation W and $\overline{\xi}_0^{i,N} \overset{\text{def}}{=} \xi_0^{i,N}$, $1 \leq i \leq N$, are independent of W^i i.i.d. random variables with common law η_0.

Let us prove that

$$E\Big(\sup_{0\leq n\leq T} |\overline{\xi}_n^{1,N} - \xi_n^{1,N}|\Big) \leq \frac{A(T)}{\sqrt{N}} \tag{82}$$

for some finite constant $A(T) > 0$. Therefore, using Lemma 1, we conclude that the empirical measures

$$\eta^N \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{(\xi_0^{i,N},\ldots,\xi_T^{i,N})}$$

converge in law as N is growing to the McKean measure $\eta \in \mathscr{P}(E^{T+1})$.

To prove (82) we use the following decomposition:

$$\begin{aligned} |\overline{\xi}_n^{1,N} - \xi_n^{1,N}| \leq{}& \frac{1}{N}\sum_{j=1}^N \big|F_n(\xi_{n-1}^{1,N}, \xi_{n-1}^{j,N}, W_n^1) - F_n(\overline{\xi}_{n-1}^{1,N}, \xi_{n-1}^{j,N}, W_n^1)\big| \\ &+ \frac{1}{N}\sum_{j=1}^N \big|F_n(\overline{\xi}_{n-1}^{1,N}, \xi_{n-1}^{j,N}, W_n^1) - F_n(\overline{\xi}_{n-1}^{1,N}, \overline{\xi}_{n-1}^{j,N}, W_n^1)\big| \\ &+ \left|\frac{1}{N}\sum_{j=1}^N F_n(\overline{\xi}_{n-1}^{1,N}, \overline{\xi}_{n-1}^{j,N}, W_n^1) - \overline{F}_n(\overline{\xi}_{n-1}^{1,N}, W_n^1)\right|. \end{aligned}$$

Since the functions F_n are bounded and Lipschitz there exists a finite constant $C < +\infty$ such that

$$E\big(|\bar{\xi}_n^{1,N} - \xi_n^{1,N}|\big) \leq C\left(E\big(|\bar{\xi}_{n-1}^{1,N} - \xi_{n-1}^{1,N}|\big) + E\left|\frac{1}{N}\sum_{j=1}^{N} F_n(\bar{\xi}_{n-1}^{1,N}, \bar{\xi}_{n-1}^{j,N}, W_n^1) - \overline{F}_n(\bar{\xi}_{n-1}^{1,N}, W_n^1)\right|\right).$$

Using a discrete time version of Gronwall's lemma, one gets

$$E\big(|\bar{\xi}_T^{1,N} - \xi_T^{1,N}|^\star\big) \leq C(T)\sum_{n=1}^{T} E\left|\frac{1}{N}\sum_{j=1}^{N} F_n(\bar{\xi}_{n-1}^{1,N}, \bar{\xi}_{n-1}^{j,N}, W_n^1) - \overline{F}_n(\bar{\xi}_{n-1}^{1,N}, W_n^1)\right|,$$

where $|\bar{\xi}_n^{1,N} - \xi_n^{1,N}|^\star \stackrel{\text{def}}{=} \sup_{0\leq k\leq n} |\bar{\xi}_k^{1,N} - \xi_k^{1,N}|$ and $C(T) < +\infty$. To bound the terms of the right-hand side we observe that, by the law of large numbers,

$$E\left(\left|\frac{1}{N}\sum_{j=1}^{N} F_n(\bar{\xi}_{n-1}^{1,N}, \bar{\xi}_{n-1}^{j,N}, W_n^1) - \overline{F}_n(\bar{\xi}_{n-1}^{1,N}, W_n^1)\right|^2\right) \leq \frac{(2\|F_n\|)^2}{N}, \qquad 1 \leq n \leq T.$$

4. Application to the nonlinear filtering problem. The basic model for the general nonlinear filtering problem consists of a time inhomogeneous Markov process X and a nonlinear observation Y with observation noise V. Namely, let (X, Y) be the Markov process taking values in $S \times \mathbb{R}^d$, $d \geq 1$, and defined by the system

$$\mathscr{F}(X/Y) \qquad \begin{aligned} X &= (X_n)_{n\geq 0}, \\ Y_n &= h_n(X_n) + V_n, \qquad n \geq 1, \end{aligned} \tag{83}$$

where S is a locally compact and separable metric space, h_n: $S \to \mathbb{R}^d$, $d \geq 1$, are continuous functions and V_n are independent random variables with continuous and positive density g_n with respect to Lebesgue measure. The signal process X that we consider is assumed to be a non-inhomogeneous and S-valued Markov process with Feller transition probability kernel K_n, $n \geq 1$, and initial probability measure ν on S. We will assume the observation noise V and X are independent.

The classical filtering problem is to estimate the distribution of X_n conditionally to the observations up to time n. Namely,

$$\pi_n(f) \stackrel{\text{def}}{=} E(f(X_n)/Y_1, \ldots, Y_n) \tag{84}$$

for all $f \in \mathscr{C}_b(S)$. The nonlinear filtering problem has been extensively studied in the literature. With the notable exception of the linear-Gaussian situation or wider classes of models (Bènes filters [1]), optimal filters have no finitely recursive solution [7]. Although Kalman filtering [27, 30] is a popular tool in handling estimation problems, its optimality heavily depends on linearity. When used for nonlinear filtering (extended Kalman filter), its performance relies on and is limited by the linearizations performed on the concerned model. The interacting particle systems approach developed hereafter can be seen

as a nonlinear filtering method which discards linearizations. More precisely, these techniques use the nonlinear system model itself in order to solve the filtering problem. The evolution of this material may be seen quite directly through the following chain of papers: [4], [5], [16], [17], [21], [14], [20], [24]. Nevertheless, in most of these papers this method is applied as an heuristic approximation to specific models, its general nature is not emphasized and experimental simulations are the only guides for handling concrete problems.

The remainder of this paper is divided into three sections. In Section 4.1 we formulate the filtering problem in such a way that the techniques of Section 3 can be applied. The problem of assessing the distributions (84) is of course related to that of recursively computing the conditional distributions π_n, $n \geq 0$, which provides all statistical information about the state variables X_n obtainable from the observations $(Y_1, \ldots, Y_n)$, $n \geq 0$. The key idea is to study the filtering problem along the the lines proposed by Kunita [29] and Stettner [41]. Briefly stated, the essence of the present formulation is that, given the observations $Y = y$, the conditional distributions π_n, $n \geq 0$, are a solution of an explicit dynamical model with infinite-dimensional state space of the form studied in the first part of our development. Namely,

$$\begin{aligned} &\pi_n = \rho_n(y_n, \pi_{n-1}), \qquad n \geq 1, \\ &\pi_0 = \nu, \end{aligned} \tag{85}$$

where $y_n \in \mathbb{R}^d$ is the current observation at the time $n \geq 1$ and $\rho_n(y_n, \bullet)$ are continuous functions $\rho_n(y_n, \bullet)$: $\mathscr{P}(S) \to \mathscr{P}(S)$. It should be noted that in this form the optimal filter has a recursive but infinite-dimensional solution (except for the linear-Gaussian case, where the Kalman filter reduces to compute the mean and variance parameters). For illustration, recalling the constructions of the particle approximation of a measure-valued system of the form (85) described in Section 3, we will see that the local dynamics at the time $n \geq 1$ of the corresponding particle system are given by the distributions

$$\rho_n\left(y_n, \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}\right)(dz) = \sum_{i=1}^{N} \frac{g_{y,n}(x^i)}{\sum_{j=1}^{N} g_{y,n}(x^j)} K_{y,n}(x^i, dz), \tag{86}$$

$$x^1, \ldots, x^N \in S, \ n \geq 1,$$

where

$$g_{y,n}(x^i) = \int g_n(y_n - h(z))\, K_n(x^i, dz) \tag{87}$$

and

$$K_{y,n}(x^i, dz) = \frac{g_n(y_n - h(z))}{g_{y,n}(x^i)} K_n(x^i, dz) \tag{88}$$

for all $x^i \in S$, $1 \leq i \leq N$. What is remarkable is that the particle system motion is strongly influenced by the observations. More precisely, $K_{y,n}$ is exactly the conditional distribution of X_n given X_{n-1} and the observation Y_n. Intu-

itively speaking, when the observation of X_n becomes available, the particles are sampled around the real state X_n and this guarantees an occupation of the state space regions in accordance with the observations, thus providing a well-behaved adaptative grid. Unfortunately, the main difficulty in directly applying the random particle methods of Section 3 to (85) stems from the fact that this local dynamic still has the disadvantage of incorporating integrations over the space S. Thus, another kind of approximation is needed to simulate the motion of the particles. Nevertheless, we will work out an example in which the integrals (87) and (88) have an explicit and simple form.

This special case apart, such a computational difficulty will be solved by studying the conditional distributions of the pair process (X_n, X_{n+1}) with respect to the observations up to time n. Namely,

$$\eta_n(f) \stackrel{\text{def}}{=} E(f(X_n, X_{n+1})/Y_1, \ldots, Y_n) \qquad \forall\, f \in \mathscr{C}_b(S^2),\ \forall\, n \geq 0. \tag{89}$$

The advantage of this alternative formulation of the optimal filter is that it incorporates separately the so-called prediction and updating mechanisms. To be entirely precise, we will see that, given the observations $Y = y$, the conditional distributions η_n, $n \geq 0$, are solution of a new measure-valued dynamical system

$$\begin{aligned} &\eta_n = \varphi_n(y_n, \eta_{n-1}), \qquad n \geq 1, \\ &\eta_0 = \nu \times K_1, \end{aligned} \tag{90}$$

where $y_n \in \mathbb{R}^d$ is the current observation at the time $n \geq 1$ and $\varphi_n(y_n, \bullet)$ are continuous functions $\varphi_n(y_n, \bullet)$: $\mathscr{P}(S^2) \to \mathscr{P}(S^2)$. Then, it will follow easily that the local dynamic of the corresponding particle system with simple interactions is given by the distributions

$$\begin{aligned} &\varphi_n\left(y_n, \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_0^i, x_1^i)}\right)(du, dv) \\ &\qquad = \sum_{i=1}^{N} \frac{g_n(y_n - h(x_1^i))}{\sum_{j=1}^{N} g_n(y_n - h(x_1^j))}\, \delta_{x_1^i}(du)\, K_{n+1}(u, dv) \end{aligned} \tag{91}$$

with $n \geq 1$, $x_0^1, \ldots, x_0^N \in S$ and $x_1^1, \ldots, x_1^N \in S$. Roughly speaking, the prediction mechanism is introduced in the filtering model (89) in order to express in an explicit form the local dynamics of the associated random particle approximation.

The aim of Section 4.2 is to use the results of Section 3 to describe several interacting particle approximations of the nonlinear and measure-valued dynamical system (90). These approximations belong to the class of algorithms called genetic algorithms. These algorithms are based on the genetic mechanisms which guide natural evolution: exploration–mutation and updating–selection. They were introduced by Holland [26] to handle global optimization problems on a finite set. The first well-founded convergence theorem ensuring the convergence of the algorithm toward the desired set of the global minima of a given fitness function was obtained by Cerf in 1994 in his Ph.D. disser-

tation [6]. Another simple proof based on the use of Log-Sobolev inequalities and semigroup techniques can be found in [19].

In the beginning of this section we first describe a basic particle approximation with simple interaction. In this situation and in view of the distributions (91), the local dynamic of the corresponding particle systems is decomposed into two mechanisms. In the first one, each particle explores randomly the state space S, independently of each other, according to the transition probability kernel of the signal process X. Finally, when the observation is received, each particle examines the previous system and chooses a site randomly in accordance with the observation data.

In a second stage we describe a more general evolutionary scheme which includes branching mechanisms. Upon carefully examining the local dynamics of the particles, it will be shown that the corresponding transitions are themselves natural approximations of the distributions (86). Intuitively speaking, the integral form of the conditional distribution $K_{y,n}$ and the weights $g_{y,n}$ given by (87) are estimated at each step of the algorithm by auxiliary branching particles moving independently of each other according to the transition probability kernel of the signal process.

Computationally, the particle approximation with simple interactions is, of course, more time saving because it does not use branching mechanisms, but several numerical studies have revealed that its use is a more efficient way to solve the filtering problem. In fact, the choice of the number of auxiliary branching particles has considerable effect on the dynamics of the particle systems. The interested reader is referred to [4], [5] and [20].

There seems to be numerical evidence of this superiority. From an intuitive point of view, the physical reason of that seems to be that the particle systems are more likely to track the signal process by conditioning the exploration in accordance with the observations and thus avoid divergence from the real process X. This observation leads us to investigate more closely the relationship between these interacting particle resolutions. Although it is intuitively clear that a benefit can be obtained from the use of branchings, it is still an object of investigation to prove the superiority of such approximations. We remark that the estimates provided by Theorem 2 and Proposition 4 are in some ways rather crude; without some precise bound on the speed of convergence of such algorithms it is difficult to get a comparison argument between them. On the other hand, we have already noted in Remark 2 that the particle approximation with simple interactions can be viewed as a special case of the interacting particle approximation with branching mechanisms. To be entirely precise, when the number of auxiliary branching particles is equal to 1 these two algorithms are exactly the same. Thus, the interacting particle approximation with branchings generalizes the particle approximation with simple interactions.

We finish the paper with some natural generalizations of the elementary stochastic algorithms described in Section 4.2. Briefly stated, the particle system described in Section 4.3 will track the signal process by considering exploration paths of a given length $r \geq 1$ and limited sections of the obser-

vation path. In the case $r = 1$, these constructions will reduce to those described above. This enables a unified description of our particle approximations in terms of three parameters: the population size, the number of auxiliary branching particles and the length of exploration paths.

Several numerical investigations [4], [5] and [20], have also revealed that the introduction of exploration paths also tends to re-center the particles around the signal trajectory. Of course we have touched in this paper only a limited number of questions. For instance, we left open the practical question of the best choice of population size, the number of auxiliary branching particles and the length of exploration paths in accordance with the nonlinear filtering problem at hand. These simple questions turn out to be surprisingly hard to answer satisfactorily and no firm results concerning the choice of parameters have been available. Another question we left in the dark is the study of the asymptotic behavior of these algorithms in terms of the ergodic properties of the signal semigroup. In this direction, something was done in [16] when the state space is compact but many questions have yet no answers.

4.1. *Formulation of the nonlinear filtering problem.* The object of this section is to introduce the filtering model in such a way that the techniques of Section 3 can be applied. We emphasize that several presentations are available and here we follow rather closely the paper of Stettner [41]. For simplicity the nonlinear filtering problem in discrete time are treated throughout. The main virtue of these problems is that the theory is very simple and the ideas transparent. It is thus a good starting point.

Let $X = (\Omega_1 = S^{\mathbb{N}}, (F_n^1)_{n\geq 0}, (X_n)_{n\geq 0}, P_X^0)$ be a time-inhomogeneous discrete time Markov process with transition operators K_n, $n \geq 1$, initial distribution ν and let $Y = (\Omega_2 = (\mathbb{R}^d)^{\mathbb{N}}, (F_n^2)_{n\geq 0}, (Y_n)_{n\geq 0}, P_Y^0)$ be a sequence of independent of X independent random variables with continuous and positive density g_n with respect to Lebesgue measure.

On the canonical space $(\Omega = \Omega_1 \times \Omega_2,\ F_n = F_n^1 \times F_n^2,\ P^0 = P_X^0 \otimes P_Y^0)$ the signal process X and the observation process Y are P_0-independent. Let $(h_n)_{n\geq 1}$ be a family of continuous functions $h_n\colon S \to \mathbb{R}^d$, $n \geq 1$. Let us set

$$L_n = \prod_{k=1}^{n} g_k(Y_k - h_k(X_k))/g_k(Y_k), \qquad n \geq 0 \tag{92}$$

with the convention $\prod_{\varnothing} = 1$. Note that L is a $(P_0, (F_n)_{n\geq 0})$-martingale. Then we can define a new probability measure P on $(\Omega, (F_n)_{n\geq 0})$ such that the restrictions P_n^0 and P_n to F_n satisfy

$$P_n = L_n\, P_n^0, \qquad n \geq 0. \tag{93}$$

One can check the following easily.

LEMMA 4. *Under P, X is a time-inhomogeneous Markov process with transition operators K_n, $n \geq 1$ and initial distribution ν. Then $V_n = Y_n - h_n(X_n)$, $n \geq 1$, are independent of X and independent random variables with continuous and positive density g_n with respect to Lebesgue measure.*

We will use $E(\bullet)$ to denote the expectations with respect to P on Ω. The following well-known result gives a functional integral representation for the conditional expectation, which is known as the Kallianpur–Striebel formula [28] (see also [41] and [29]).

LEMMA 5 (Stettner [41]). *The conditional distributions $(\pi_n)_{n\geq 0}$ form a time-inhomogeneous and $(\sigma(Y^n), P)$-Markov process on $\mathscr{P}(S)$ with transition operators*

$$\Pi_n F(\mu) = \int F(\rho_n(y,\mu))\, g_n(y - h_n(x))\, dy\, (\mu K_n)(dx) \qquad \forall\ F \in \mathscr{C}_b(\mathscr{P}(S)),\ \forall\ \mu \in \mathscr{P}(S), \tag{94}$$

where $\rho_n(y,\bullet)\colon \mathscr{P}(S) \to \mathscr{P}(S)$, $n \geq 1$, is the continuous function given by

$$\rho_n(y,\mu)f = \frac{\int f(x)\, g_n(y - h_n(x))\, (\mu K_n)(dx)}{\int g_n(y - h_n(z))\, (\mu K_n)(dz)} \qquad \forall\ f \in \mathscr{C}_b(S) \tag{95}$$

for all $\mu \in \mathscr{P}(S)$, $y \in \mathbb{R}^d$ and $n \geq 1$. Therefore, given the observations $Y = y$, the distributions π_n, $n \geq 0$, are a solution of the $\mathscr{P}(S)$-valued dynamical system

$$\pi_n = \rho_n(y_n, \pi_{n-1}), \qquad n \geq 1,\ \pi_0 = \nu. \tag{96}$$

Equation (96) is called the nonlinear filtering equation. Even if it looks innocent, it requires extensive computation and can rarely be solved analytically. It is thus necessary to resort to numerical solutions. This lemma is proved in [41]. We quote its outline for the convenience of the reader.

PROOF OF LEMMA 5. In view of (93) we have

$$\pi_n f = \frac{E_0(f(X_n)\, L_n / Y^n)}{E_0(L_n / Y^n)} = \frac{E_0^X(f(X_n)\, L_n)}{E_0^X(L_n)},$$

where $E_0(\bullet)$ denotes the expectation with respect to P^0 and $E_0^X(\bullet)$ denotes the integration of the paths of the Markov process X and the variable X_0. Then we obtain

$$\pi_n f = \frac{E_0^X(E_0^X(f(X_n)\, g_n(Y_n - h_n(X_n))/X_{n-1})\, L_{n-1})}{E_0^X(E_0^X(g_n(Y_n - h_n(X_n))/X_{n-1})\, L_{n-1})}$$

and, finally,

$$\begin{aligned}\pi_n f &= \frac{E_0^X(E_0^X(f(X_n)\, g_n(Y_n - h_n(X_n))/X_{n-1})\, L_{n-1})/E_0^X(L_{n-1})}{E_0^X(E_0^X(g_n(Y_n - h_n(X_n))/X_{n-1})\, L_{n-1})/E_0^X(L_{n-1})} \\ &= \frac{\int f(x)\, g_n(Y_n - h_n(x))\, (\pi_{n-1}K_n)(dx)}{\int g_n(Y_n - h_n(z))\, (\pi_{n-1}K_n)(dz)}.\end{aligned}$$

□

The temptation is to apply immediately the random particle approximations described in Section 3.1. Recalling the construction of the interacting particle

system (17), we see that the transition of individual particles at the time $n \geq 1$ will be specified, in this situation, by the transition probability kernels

$$\rho_n\Big(y, \frac{1}{N}\sum_{i=1}^{N}\delta_{x^i}\Big)(dx_1) = \sum_{i=1}^{N}\frac{g_n(y_n - h_n(x_1))\ K_n(x^i, dx_1)}{\sum_{j=1}^{N}\int g_n(y_n - h_n(z_1))\ K_n(x^j, dz_1)}, \tag{97}$$

where N is the size of the particle systems, $Y_n = y_n$ is the current observation data, $x_1 \in S$ and $(x^1, \ldots, x^N) \in S^N$. To be more precise, let us put, for all $x_0, x_1 \in S$,

$$K_{y,n}(x_0, dx_1) = \frac{g_n(y_n - h_n(x_1))}{\int g_n(y_n - h_n(z_1))\ K_n(x_0, dz_1)}\ K_n(x_0, dx_1) \tag{98}$$

$$g_{y,n}(x_0) = \int K_n(x_0, dz_1) g_n(y_n - h_n(z_1)). \tag{99}$$

Using Bayes' rule, we note that $K_{y,n}(x_0, dx_1)$ is the density under P of the distribution of X_n conditionally to $X_{n-1} = x_0$ and $Y_n = y_n$, and $g_{y,n}(x_0)$ is the density under P of the distribution of Y_n conditionally to $X_{n-1} = x_0$.

With these notations, the transition probability kernel (97) becomes

$$\rho_n\Big(y, \frac{1}{N}\sum_{i=1}^{N}\delta_{x^i}\Big)(dx_1) = \sum_{i=1}^{N}\frac{g_{y,n}(x^i)}{\sum_{j=1}^{N} g_{y,n}(x^j)}\ K_{y,n}(x^i, dx_1). \tag{100}$$

Let us work out an example in which the desired transition (97) has a simple form.

EXAMPLE 3. Let (X, Y) be the Markov process taking values in $\mathbb{R} \times \mathbb{R}$ and defined by the system

$$\begin{aligned} X_n &= f_n(X_{n-1}) + W_n, \\ Y_n &= C_n\ X_n + V_n, \qquad n \geq 1, \end{aligned} \tag{101}$$

where $f_n\colon \mathbb{R} \to \mathbb{R}$, $C_n \in \mathbb{R}$ and W, respectively, V, is a discrete time Gaussian process with zero mean and variance function q, respectively, r. In this specific situation one easily gets

$$\begin{aligned} K_{y,n}(x_0, dx_1) = \frac{1}{\sqrt{2\,\pi\,|s_n|}}\exp\Big(-\frac{1}{2\,|s_n|}\Big(x_1 - \Big[f_n(x_0) + s_n\,C_n\,r_n^{-1} \\ \times \big(y_n - C_n\,f_n(x_0)\big)\Big]\Big)^2\Big) \quad \text{and} \end{aligned} \tag{102}$$

$$\begin{aligned} g_{y,n}(x_0) = \frac{1}{\sqrt{2\,\pi\,|q_n|\,|r_n|/|s_n|}} \\ \times \exp\Big(-\frac{1}{2\,|q_n|\,|r_n|/|s_n|}\big(y_n - C_n\,f_n(x_0)\big)^2\Big) \end{aligned}$$

with $s_n = (q_n^{-1} + C_n\ r_n^{-1}\ C_n)^{-1}$ and $x_0, x_1 \in \mathbb{R}$. Once more the verification of (102), while straightforward, is somewhat lengthy and is omitted.

Unfortunately, in most cases it is not possible to exhibit explicitly the form of transition (100). In a little while we shall see one way to approximate the transition probability kernel (100). Roughly speaking, the idea is to replace, in the definition of $K_{y,n}$, each transition $K_n(x^i, dx_1)$, $1 \leq i \leq N$, by the empirical measure

$$\frac{1}{N'} \sum_{j=1}^{N'} \delta_{x^{i,j}} \tag{103}$$

where $(x^{i,1}, x^{i,2}, \ldots, x^{i,N'})$ are i.i.d. random variables with common law $K_n(x^i, dx_1)$, $1 \leq i \leq N$. With these notations we have formally

$$\sum_{i,j=1}^{N,N'} \frac{g_n(y_n - h_n(x^{i,j}))}{\sum_{k,l=1}^{N,N'} g_n(y_n - h_n(x^{k,l}))} \delta_{x^{i,j}} \underset{N' \sim +\infty}{\sim} \rho_n\left(y, \frac{1}{N} \sum_{i=1}^{N} \delta_{x^i}\right).$$

Using such a *local approximation*, it is not obvious that the empirical measure of the particle system will converge to the conditional distribution since it is not clear what condition on the system size N' will guarantee the convergence of the algorithm. Fortunately, this vexing technical difficulty will disappear when we model this *local approximation* by a branching mechanism.

The difficulty with the recursion (96) is that it involves two separate mechanisms. Namely, the first one,

$$\mu \mapsto \mu K_n,$$

does not depends on the current observation and it is usually called the prediction and the second one

$$\mu \mapsto \frac{g_n(Y_n - h_n(\bullet))}{\int g_n(Y_n - h_n(z))\, \mu(dz)} \mu$$

updates the distribution given the current observation. It is therefore essential to find a dynamical system formulation which incorporates separately the prediction and the updating mechanisms. A natural idea is to study the distribution of the pair process (X_n, X_{n+1}) conditionally to the observations up to time n. Namely,

$$\eta_n(f) \overset{\text{def}}{=} E(f(X_n, X_{n+1})/Y_1, \ldots, Y_n) \qquad \forall\, f \in \mathscr{C}_b(S^2),\ \forall\, n \geq 0. \tag{104}$$

Lemma 6 is a modification of Lemma 5.

LEMMA 6. *Given the observations $Y = y$, η_n is a solution of the $\mathscr{P}(S^2)$-valued dynamical system*

$$\begin{aligned} &\eta_n = \varphi_n(y_n, \eta_{n-1}), \qquad n \geq 1, \\ &\eta_0 = \nu \times K_1, \end{aligned} \tag{105}$$

where $\varphi_n(y, \bullet)$: $\mathscr{P}(S^2) \to \mathscr{P}(S^2)$ *is the continuous function given by*

$$\varphi_n(y, \eta)f = \frac{\int f(x_1, x_2)\, g_n(y - h_n(x_1))\, d\eta(x_0, x_1)\, K_{n+1}(x_1, dx_2)}{\int g_n(y - h_n(z_1))\, d\eta(z_0, z_1)} \qquad \forall\ f \in \mathscr{C}_b(S^2) \tag{106}$$

for all $\eta \in \mathscr{P}(S^2)$, $y \in \mathbb{R}^d$ *and* $n \geq 1$. *Moreover* $(\eta_n)_n$ *is a time-inhomogeneous and* $(\sigma(Y^n), P)$*-Markov process on* $\mathscr{P}(S^2)$ *with transition operators*

$$\overline{\Pi}_n F(\mu) = \int F(\varphi_n(y, \mu))\, g_n(y - h_n(x))\, dy\, d\mu(z, x) \qquad \forall\ F \in \mathscr{C}_b(\mathscr{P}(S^2)),\ \forall\ \mu \in \mathscr{P}(S^2). \tag{107}$$

PROOF. It is easily checked from (96) that

$$\eta_n = \pi_n \times K_{n+1} = \rho_n(y_n, \pi_{n-1}) \times K_{n+1} = \varphi_n(y_n, \eta_{n-1}).$$

So (105) and Lemma 5 end the proof of the lemma. □

Returning once more to the random particle system described in Section 3.1, we see that the local dynamics of an individual particle will now be given by the transition probability kernels

$$\begin{aligned} &\varphi_n\left(y_n, \frac{1}{N}\sum_{i=1}^{N} \delta_{(z_1^i, z_2^i)}\right)(d(x_1, x_2)) \\ &\quad = \sum_{i=1}^{N} \frac{g_n(y_n - h_n(z_2^i))}{\sum_{j=1}^{N} g_n(y_n - h_n(z_2^j))} \delta_{z_2^i}(dx_1)\, K_{n+1}(x_1, dx_2), \end{aligned} \tag{108}$$

where N is the size of the particle system and $Y_n = y_n$ is the current observation data.

4.2. *Interacting particle resolutions.* This section covers stochastic particle methods for the numerical solving of the nonlinear filtering equation (105) based upon the simulation of interacting and branching particle systems. The technical approach presented here is to work with a given sequence of observations $Y = y$. With regard to (105) and (106), the nonlinear filtering problem is now reduced to the infinite dimensionality of the state space $\mathscr{P}(S^2)$. This assumption enables us to formulate the conditional distributions as probabilities parameterized by the observation parameters and the solution of a measure-valued dynamical system. The design of our particle system approach is described in Section 3 for the numerical solving of general measure-valued dynamical systems. We shall use the notions and notations introduced in Sections 3.1 and 3.2.

4.2.1. *Particle systems with simple interactions.* The algorithm presented in (17) was referred to as a particle system approximation with simple interaction function. With regard to the nonlinear filtering equation (105), the inter-

acting particle system is modeled by a Markov chain $(\Omega', (F'_n)_{n\geq 0}, (\zeta_n)_n, P_{[y]})$ with state space S^{2N}, where $N \geq 1$ is the size of the system.

The N-tuple of elements of S^2, that is, the points of the set S^{2N}, are called particle systems and will be denoted by the letters z, x. Recalling the description (17), this chain is defined by

$$\begin{aligned} P_{[y]}(\zeta_0 \in dx) &= \prod_{p=1}^{N} \eta_0(dx^p), \\ P_{[y]}(\zeta_n \in dx/\zeta_{n-1} = z) &= \prod_{p=1}^{N} \varphi_n\left(y_n, \frac{1}{N}\sum_{i=1}^{N} \delta_{z^i}\right)(dx^p) \end{aligned} \tag{109}$$

In view of (105) and (108), we have that

$$P_{[y]}(\zeta_0 \in dx) = \prod_{i=1}^{N} \nu(dx_1^i)\, K_1(x_1^i, dx_2^i) \tag{110}$$

$$\begin{aligned} &P_{[y]}(\zeta_n \in dx/\, \zeta_{n-1} = z) \\ &\quad = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y_n - h_n(z_2^i))}{\sum_{j=1}^{N} g_n(y_n - h_n(z_2^j))} \delta_{z_2^i}(dx_1^p)\, K_{n+1}(x_1^p, dx_2^p), \end{aligned} \tag{111}$$

where we have the following:

1. the expression y_n is the observation data at the time n;
2. $x = (x_1, x_2)$, $z = (z_1, z_2) \in S^N \times S^N$ and $x^i = (x_1^i, x_2^i)$, $z^i = (z_1^i, z_2^i) \in S^2$ for all $1 \leq i \leq N$;
3. $\zeta_n \in S^N \times S^N$ is the system of S^2-valued particles at the time n.

To be more precise, it is convenient to introduce additional notations. Let us set

$$\zeta_n \stackrel{\text{def}}{=} (\widehat{\xi}_n, \xi_{n+1}) \in S^N \times S^N \qquad \forall\, n \geq 0.$$

Now the points of the set S^N will be denoted by the letters x and z. Using these notations, (109) together with (108) lead to the following Markov model:

$$P_{[y]}(\xi_n \in dx/\widehat{\xi}_{n-1} = z) = \prod_{p=1}^{N} K_n(z^p, dx^p), \tag{112}$$

$$P_{[y]}(\widehat{\xi}_n \in dx/\, \xi_n = z) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y_n - h_n(z^i))}{\sum_{j=1}^{N} g_n(y_n - h_n(z^j))} \delta_{z^i}(dx^p), \tag{113}$$

where y_n is the observation data at time n and $x, z \in S^N$. Equations (112) and (113) resemble a genetic algorithm [6, 19, 26]. The advantage of this formulation is that it incorporates separately the prediction $\widehat{\xi}_{n-1} \rightsquigarrow \xi_n$ and the updating $\xi_n \rightsquigarrow \widehat{\xi}_n$ mechanisms. Thus, we see that the particles move according the following rules.

1. *Prediction*. Before the updating mechanism each particle evolves according to the transition probability kernel of the signal process.
2. *Updating*. When the observation $Y_n = y_n$ is received, each particle *examines* the system of particles $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$ and chooses randomly a site ξ_n^i with probability

$$\frac{g_n(y_n - h_n(\xi_n^i))}{\sum_{j=1}^N g_n(y_n - h_n(\xi_n^j))}.$$

Our purpose is now to understand why the second mechanism (113) plays a very special role in the behavior of the particle filter. What is important is that each particle interacts selectively with the system in accordance with the observation data. Roughly speaking, a given particle which takes a given value is more likely to choose another site if its position generally *disagrees* with the current observation than if it agrees with it. These observations point to the very interesting dynamical role played by the updating mechanism. It stabilizes the particles' motion around certain values of the real signal which are determined by the observations, thus providing a well-behaved adaptative grid.

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 $(\tilde{\Omega}, (\tilde{F}_n)_{n\geq 0}, \tilde{P})$ defined as follows.

1. Recall (Ω, F_n, P) is the canonical space for the signal observation pair (X, Y).
2. We define $\tilde{\Omega} = \Omega' \times \Omega$ and $\tilde{F}_n = F'_n \times F_n$ and, for every $\omega \stackrel{\text{def}}{=} (\omega^1, \omega^2, \omega^3) \in \overline{\Omega}$ we define

$$\zeta_n(\omega) = \omega_n^1, \qquad X_n(\omega) = \omega_n^2, \qquad Y_n(\omega) = \omega_n^3.$$

3. For every $A \in F'_n$ and $B \in F_n$, we define $\tilde{P}$ as follows:

$$\tilde{P}(A \times B) \stackrel{\text{def}}{=} \int_B P_{[Y]}(A)\, dP. \tag{114}$$

As usual we use $\tilde{E}(\bullet)$ to denote expectations with respect to $\tilde{P}$. The approximation of the conditional distribution η_n by the empirical measure

$$\eta_n^N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \delta_{(\widehat{\xi}_n^i, \xi_{n+1}^i)}$$

is guaranteed by Theorem 2. Moreover, applying the dominated convergence theorem, we find

$$\forall\ f \in \mathscr{C}_b(S^2),\ \forall\ n \geq 1, \qquad \lim_{N\to+\infty} \tilde{E}(|\eta_n^N f - \eta_n f|^2) = 0.$$

To see this claim, it clearly suffices to prove that condition (25) is satisfied for every sequence of observation $Y = y$. By the very definition of the functions

$\varphi_n(y_n, \bullet)$ one gets easily

$$\forall\, n \geq 1,\ \forall\, \mu \in \mathscr{P}(S^2),\ \forall\, f \in \mathscr{C}_b(S^2), \qquad \varphi_n(y_n, \mu)f = \frac{\mu(g_n\, T_n f)}{\mu(g_n)}$$

with

$$g_n(x_0, x_1) = g_n(y_n - h_n(x_1)) \quad \text{and} \quad T_n f(x_0, x_1) = \int f(x_1, x_2)\, K_{n+1}(x_1, dx_2).$$

Notice that this type of measure valued process has been examined in (29) Section 3.1.4. Arguing as before it is straightforward to see that condition (25) is satisfied.

4.2.2. *Interacting particle systems with branchings.* The method described above is the crudest of the random particle methods. When applied to the nonlinear filtering equation (105), the density profiles might lack some of the statistical details of the conditional distributions which one would get when using the branching refinement method introduced in Section 3.2. We shall use the notions and notations introduced in Section 3.2.1. This refinement is also relatively easy to program and it has been used with success in many practical situations [4, 5, 21], but its use still leaves open the optimal choice of the number of auxiliary particles. Let us start with a few remarks. When each particle branches into one particle, these algorithms are exactly the same. On the other hand, when the number of auxiliary branching particles is growing, the transition probability density of an individual particle tends to the transition (100). Even if the first algorithm is more timesaving because it does not use branching simulations, several numerical simulations have revealed that a clear benefit can be obtained by using branching particles. Unfortunately, one cannot quantify this superiority. Some attempts in this direction have been done in the end of Section 3.2.

When considering the nonlinear filtering equation (105) with a given sequence of observations $Y = y$, the corresponding interacting particle system with branchings is modeled by a Markov chain $(\Omega', (F'_n)_{n\geq 0}, (\zeta_n)_n, P_{[y]})$ with state space

$$E^{(N)} = S^{(N)_1} \times S^{(N)_2}, \qquad (N) \stackrel{\text{def}}{=} (N_1, N_2),$$

where $(N)_1 = \{1, \ldots, N_1\}$ and $(N_2) = \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$, $N_1, N_2 \geq 1$. Each point

$$x = (x_1, x_2) = (x_1^{p_1}, x_2^{p_1, p_2})_{1\leq p_1\leq N_1,\, 1\leq p_2\leq N_2} \in S^{(N)_1} \times S^{(N)_2},$$

consists of two particles systems called random particle trees; they will be denoted by the letters z, x. Now, recalling the description (57), the transition probability densities of this Markov chain are given by

$$P_{[y]}(\zeta_0 \in dx) = \prod_{p_1=1}^{N_1} \nu(dx_1^{p_1}) \prod_{p_2=1}^{N_2} K_1(x_1^{p_1}, dx_2^{p_1, p_2}), \tag{115}$$

$$
\begin{aligned}
&P_{[y]}(\zeta_n \in dx / \zeta_{n-1} = z) \\
&\quad = \prod_{p_1=1}^{N_1} \sum_{i_1,\, i_2=1}^{N_1,\, N_2} \frac{g_n(y_n - h_n(z_2^{i_1,\, i_2}))}{\sum_{j_1,\, j_2=1}^{N_1,\, N_2} g_n(y_n - h_n(z_2^{j_1,\, j_2}))}\, \delta_{z_2^{i_1,\, i_2}}(dx_1^{p_1}) \\
&\qquad \times \prod_{p_2=1}^{N_2} K_{n+1}(x_1^{p_1}, dx_2^{p_1,\, p_2})
\end{aligned} \tag{116}
$$

where y_n is the observation data at time n. This algorithm generalizes the one given above by allowing at each step N_2 auxiliary particles. For the moment, let us merely note that if $N_2 = 1$ then the transitions above are exactly the same as those given before.

To be more precise, it is now convenient to introduce additional notations:

$$
\zeta_n \stackrel{\text{def}}{=} (\widehat{\xi}_n, \xi_{n+1}) \in S^{(N)_1} \times S^{(N)_2} \qquad \forall\ n \geq 0.
$$

To clarify the notations, the N_1-tuple of elements of S, that is, the points of the set $S^{(N)_1}$ will be denoted by $z = (z^{p_1})_{p_1}$ and, the points of the set $S^{(N)_2}$ will be denoted by $x = (x^{p_1,\, p_2})_{p_1,\, p_2}$. For brevity, we will also write $g_n(x)$ instead of $g_n(y_n - h_n(x))$. Using these notations, we obtain the following Markov model:

$$
\begin{aligned}
P_{[y]}(\widehat{\xi}_0 \in dz) &= \prod_{p_1=1}^{N_1} \nu(dz^{p_1}), \\
P_{[y]}(\xi_n \in dx / \widehat{\xi}_{n-1} = z) &= \prod_{p_1=1}^{N_1} \prod_{p_2=1}^{N_2} K_n(z^{p_1}, dx^{p_1,\, p_2}), \\
P_{[y]}(\widehat{\xi}_n \in dz / \xi_n = x) &= \prod_{p_1=1}^{N_1} \sum_{i_1,\, i_2=1}^{N_1,\, N_2} \frac{g_n(x^{i_1,\, i_2})}{\sum_{j_1,\, j_2=1}^{N_1,\, N_2} g_n(x^{j_1,\, j_2})}\, \delta_{x^{i_1,\, i_2}}(dz^{p_1}).
\end{aligned}
$$

Continuing in the same vein, it is important to note that

$$
\begin{aligned}
P_{[y]}(\widehat{\xi}_n \in dz / \xi_n = x) &= \prod_{p_1=1}^{N_1} \sum_{i_1=1}^{N_1} \frac{\sum_{k_2=1}^{N_2} g_n(x^{i_1,\, k_2})}{\sum_{j_1,\, j_2=1}^{N_1,\, N_2} g_n(x^{j_1,\, j_2})} \\
&\qquad \times \sum_{i_2=1}^{N_2} \frac{g_n(x^{i_1,\, i_2})}{\sum_{k_2=1}^{N_2} g_n(x^{i_1,\, k_2})} \delta_{x^{i_1,\, i_2}}(dz^{p_1}).
\end{aligned} \tag{117}
$$

This formulation incorporates separately the prediction $\widehat{\xi}_{n-1} \rightsquigarrow \xi_n$ and the updating $\xi_n \rightsquigarrow \widehat{\xi}_n$ mechanism. More precisely, at each moment of time n, we see that the particles move according to the following rules.

1. *Prediction*. Before the updating mechanism, each particle $\widehat{\xi}_{n-1}^{\,i_1}$, $1 \leq i_1 \leq N_1$, branches into a fixed number N_2 of i.i.d. random particles $\xi_n = (\xi_n^{i_1,\, i_2})_{1 \leq i_2 \leq N_2}$ with law K_n.

2. *Updating*. When the observation $Y_n = y_n$ becomes available, each particle $\widehat{\xi}_{n-1}^{i_1}$, $1 \leq i_1 \leq N_1$, chooses a subsystem of auxiliary particles $(\xi_n^{i_1, i_2})_{1 \leq i_2 \leq N_2}$ at random with probability

$$\sum_{k_2=1}^{N_2} g_n(\xi_n^{i_1, k_2}) \Big/ \sum_{j_1, j_2=1}^{N_1, N_2} g_n(\xi_n^{j_1, j_2})$$

and moves to the site $\xi_n^{i_1, i_2}$ in the chosen subsystem with probability

$$g_n(\xi_n^{i_1, i_2}) \Big/ \sum_{k_2=1}^{N_2} g_n(\xi_n^{i_1, k_2}).$$

The fundamental difference between the so-called simple interaction and branching approaches lies in the fact that in the former each particle branches into a fixed number of auxiliary particles and the corresponding updating–selection procedure is itself decomposed into two different mechanisms. Intuitively speaking, the branchings are meant to discourage the particles from visiting bad state regions.

REMARK 3. The choice of the systems' size N_1 and N_2 is frequently a matter of judgment and intuition. We observe that each subsystem consists of N_2 i.i.d. random variables $(\xi_n^{i_1, i_2})_{1 \leq i_2 \leq N_2}$ with law $K_n(\widehat{\xi}_{n-1}^{i_1}, dz_1)$. So, formally

$$\frac{1}{N_2} \sum_{i_2=1}^{N_2} \delta_{\xi_n^{i_1, i_2}}(dz_1) \underset{N_2 \sim +\infty}{\sim} K_n(\widehat{\xi}_{n-1}^{i_1}, dz_1)$$

leads to

$$\sum_{i_2=1}^{N_2} \frac{g_n(\xi_n^{i_1, i_2})}{\sum_{k_2=1}^{N_1} g_n(\xi_n^{i_1, k_2})} \delta_{\xi_n^{i_1, i_2}}(dz_1) \underset{N_2 \sim +\infty}{\sim} K_{y, n}(\widehat{\xi}_{n-1}^{i_1}, dz_1)$$

and

$$\frac{1}{N_2} \sum_{i_2=1}^{N_2} g_n(\xi_n^{i_1, i_2}) \underset{N_2 \sim +\infty}{\sim} g_{y, n}(\widehat{\xi}_{n-1}^{i_1}) = \int g_n(z_1)\, K_{y, n}(\widehat{\xi}_{n-1}^{i_1}, dz_1).$$

This observation leads us to the following equivalence:

$$\tilde{P}_{[y]}(\widehat{\xi}_n \in dz / \widehat{\xi}_{n-1} = x) \underset{N_2 \sim +\infty}{\sim} \prod_{p_1=1}^{N_1} \rho_n(y_n, \frac{1}{N_1} \sum_{i_1=1}^{N_1} \delta_{x^{i_1}})(dz^{p_1}).$$

It follows that (117) is the right way to model the local approximation discussed in (103).

As in Section 4.2.1, to capture all randomness we list all outcomes into a canonical space $(\tilde{\Omega}, (\tilde{F}_n)_{n \geq 0}, \tilde{P})$ and we use $\tilde{E}(\bullet)$ to denote expectations with respect to $\tilde{P}$. Using the same line of arguments as in the end of Section 4.2.1,

the approximation of the desired conditional distribution η_n by the empirical measure

$$\eta_n^{(N)} \stackrel{\text{def}}{=} m^{(N)}((\widehat{\xi}_n, \xi_{n+1})) = \frac{1}{N_1 N_2} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \delta_{(\widehat{\xi}_n^{i_1}, \xi_{n+1}^{i_1,i_2})}$$

is guaranteed by Proposition 4. Moreover, arguing as before, we have

$$\lim_{N_1 \to +\infty} \tilde{E}(|\eta_n^{(N)} f - \eta_n f|^2) = 0.$$

4.2.2. *General interacting particle resolutions.* In the last part of this paper the above approximations are generalized. The prediction mechanism of the former particle filter will include exploration paths of a given length $r \geq 1$ and the corresponding updating procedure will then be used every r steps and will consider r observations. The idea is to study the pair process $(\mathscr{X}_n, \mathscr{X}_{n+1})$ given by

$$\mathscr{X}_n \stackrel{\text{def}}{=} (X_{nr}, \ldots, X_{nr+(r-1)}) \qquad \forall\, n \geq 0 \text{ and } r > 0. \tag{118}$$

For later convenience, let us set $\mathscr{Y}_n \stackrel{\text{def}}{=} (Y_{nr}, \ldots, Y_{nr+r-1})$ for all $n \geq 0$. Let μ_n be the distribution of the pair process $(\mathscr{X}_n, \mathscr{X}_{n+1})$ conditionally to the observations up to time nr. Namely,

$$\mu_n(f) \stackrel{\text{def}}{=} E(f(\mathscr{X}_n, \mathscr{X}_{n+1})/\mathscr{Y}_1, \ldots, \mathscr{Y}_n) \qquad \forall\, f \in \mathscr{C}_b(S^{2r}). \tag{119}$$

It is straightforward to see that this case may be reduced to the latter (104) through a suitable state space basis. More precisely, $(\mathscr{X}_n)_{n\geq 0}$ is an S^r-valued and time-inhomogeneous Markov process with transition operator

$$\mathscr{K}_n\big((x_0, \ldots, x_{r-1}), d(x_r, \ldots, x_{2r-1})\big) \stackrel{\text{def}}{=} \prod_{p=0}^{r-1} K_{nr+p}(x_{r+p-1}, dx_{r+p})$$

and initial distribution

$$\nu_0 = \nu \times K_1 \times \cdots \times K_{r-1}.$$

In this situation the corresponding observation process $(\mathscr{Y}_n)_{n\geq 0}$ may be written by

$$\mathscr{Y}_n = \mathscr{H}_n(\mathscr{X}_n) + \mathscr{V}_n \qquad \forall\, n \geq 0$$

with

$$\mathscr{H}_n(x_0, \ldots, x_{r-1}) = \big(h_{nr}(x_0), \ldots, h_{nr+(r-1)}(x_{r-1})\big)$$

and

$$\mathscr{V}_n = \big(V_{nr}, \ldots, V_{nr+(r-1)}\big), \qquad n \geq 0$$

a sequence of independent of $\mathscr{X}$ independent random variables with continuous and positive density

$$\mathscr{G}_n(v_0, \ldots, v_{r-1}) = \prod_{p=0}^{r-1} g_{nr+p}(v_p).$$

The same kind of arguments as the ones above show that μ_n is solution of the $\mathscr{P}(S^{2r})$-valued recursive equation

$$\begin{aligned} &\mu_n = \psi_n(\mathscr{Y}_n, \mu_{n-1}), \qquad n \geq 1, \\ &\mu_0 = \nu_0 \times \mathscr{K}_1, \end{aligned} \tag{120}$$

where $\psi_n(y, \bullet)$: $\mathscr{P}(S^{2r}) \to \mathscr{P}(S^{2r})$ is the continuous function given by

$$\psi_n(\mathscr{Y}_n, \eta)f = \frac{\int f(\mathscr{X}_1, \mathscr{X}_2)\, \mathscr{G}_n(\mathscr{Y}_n - \mathscr{H}(\mathscr{X}_1))\, d\eta(\mathscr{X}_0, \mathscr{X}_1)\, \mathscr{K}_{n+1}(\mathscr{X}_1, d\mathscr{X}_2)}{\int \mathscr{G}_n(\mathscr{Y}_n - \mathscr{H}_n(\mathscr{Y}_1))\, d\eta(\mathscr{Y}_0, \mathscr{Y}_1)} \tag{121}$$

for all $f \in \mathscr{C}_b(S^{2r})$ and $\eta \in \mathscr{P}(S^{2r})$. From (120) we can quickly deduce the design of general interacting particle resolutions. We will look at such algorithms in detail below and provide enough detail for the reader to be able to generate these stochastic algorithms on a computer. For brevity we will write $\mathscr{G}_n(\mathscr{X})$ instead of $\mathscr{G}_n(\mathscr{Y}_n - \mathscr{H}_n(\mathscr{X}))$. Now, the corresponding particle system is modeled by a Markov chain $(\Omega', (F'_n)_{n\geq 0}, \zeta = (\zeta_n)_{n\geq 0}, P_{[y]})$ with state space

$$E^{(N)} \stackrel{\text{def}}{=} (S^r)^{(N)_1} \times (S^r)^{(N)_2}, \qquad (N) \stackrel{\text{def}}{=} (N_1, N_2) \text{ and } N_1, N_2 \geq 1, \tag{122}$$

where

$$(N)_1 = \{1, \ldots, N_1\}, \qquad (N)_2 = \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}.$$

We will use the notations

$$\zeta_n \stackrel{\text{def}}{=} (\widehat{\xi}_n, \xi_{n+1}) \in (S^r)^{(N)_2} \times (S^r)^{(N)_2} \qquad \forall\, n \geq 0$$

and, to clarify the presentation, the points of the set $(S^r)^{(N)_1}$ will be denoted by $z = (z^{p_1})_{p_1}$ and the points of the set $(S^r)^{(N)_2}$ will be denoted by $x = (x^{p_1, p_2})_{p_1, p_2}$. Recalling the descriptions (115) and (117), this Markov chain has the following transition probability densities:

$$P_{[y]}(\widehat{\xi}_0 \in dz) = \prod_{p_1=1}^{N_1} \nu(dz_0^{p_1}) K_1(z_0^{p_1}, dz_1^{p_1}) \cdots K_{r-1}(z_{r-2}^{p_1}, dz_{r-1}^{p_1}),$$

$$\begin{aligned} P_{[y]}(\xi_n \in dx / \widehat{\xi}_{n-1} = z) = &\prod_{p_1=1}^{N_1} \prod_{p_2=1}^{N_2} K_{nr}(z_{r-1}^{p_1}, dx_0^{p_1, p_2}) \cdots \\ &K_{nr+(r-1)}(x_{r-2}^{p_1, p_2}, dx_{r-1}^{p_1, p_2}), \end{aligned}$$

$$\begin{aligned} P_{[y]}(\widehat{\xi}_n \in dz / \xi_n = x) = &\prod_{p_1=1}^{N_1} \sum_{i_1, i_2=1}^{N_1, N_2} \frac{\mathscr{G}_n(x_0^{i_1, i_2}, \ldots, x_{r-1}^{i_1, i_2})}{\sum_{j_0, j_1=1}^{N_1, N_2} \mathscr{G}_n(x_0^{j_0, j_1}, \ldots, x_{r-1}^{j_0, j_1})} \\ &\times\ \delta_{(x_0^{i_1, i_2}, \ldots, x_{r-1}^{i_1, i_2})}(d(z_0^{p_1}, \ldots, z_{r-1}^{p_1})). \end{aligned}$$

Finally, the approximation of the desired conditional distribution μ_n by the empirical measure

$$\mu_n^{(N)} \stackrel{\text{def}}{=} \frac{1}{N_1 N_2} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \delta_{(\widehat{\xi}_n^{i_1}, \xi_{n+1}^{i_1, i_2})}$$

is guaranteed by Proposition 4 and for every $f \in \mathscr{C}_b(S^{2r})$, $n \geq 1$,

$$\lim_{N_1 \to +\infty} \tilde{E}(|\mu_n^{(N)} f - \mu_n f|^2) = 0.$$

To check that this algorithm generalizes the one given above, observe that the former transition probability densities coincide with the transitions of the particle systems with simple interaction described in Section 4.2.1 when $r = 1$ and $N_2 = 1$, and they coincide with those of the interacting particle systems with branchings described in Section 4.2.1 when $r = 1$.

The choice of the parameters N_1, N_2, r requires a criterion for optimality. The result will probably depend on the dynamics of the state process and also on the dynamical structure of the observations. Unfortunately, no one has yet been able to give an effective method of solution.

Another idea is to study the performance of a modified version of the above algorithms which includes an interactive updating–selection schedule $r = r(n)$. For instance, we may choose to re-sample the particles when fifty percent of the weights are lower than A/N^p, with a convenient choice of the parameter $A > 0$ and $p \geq 2$.

Acknowledgments. I am deeply grateful to Professor Dominique Michel for useful discussions and her 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] BENES, V. E. (1981). Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* **5** 65–92.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[3] BRAUN, W. and HEPP, K. (1977). The Vlasov dynamics and its fluctuation in the $1/n$ limit of interacting particles. *Comm. Math. Phys.* **56** 101–113.

[4] CARVALHO, H. (1995). Filtrage optimal non linéaire du signal GPS NAVSTAR en racalage de centrales de navigation. Thèse de L'Ecole Nationale Supérieure de l'Aéronautique et de l'Espace.

[5] CARVALHO, H., DEL MORAL, P., MONIN, A. and SALUT, G. (1997). Optimal nonlinear filtering in GPS/INS integration. *IEEE Trans. Aerospace and Electron. Syst.* **33** 835–850.

[6] CERF, R. (1994). Une théorie asymptotique des algorithmes génétiques. Ph.D. dissertation, Univ. Montpellier II, Sciences et techniques du Languedoc.

[7] CHALEYAT-MAUREL, M. and MICHEL, D. (1983). Des résultats de non existence de filtres de dimension finie. *C.R. Acad. Sci. Paris Ser. I* **296**.
plus 0pt

[8] CRISAN, D. (1996). The nonlinear filtering problem. Ph.D. dissertation, Univ. Edinburgh.

[9] CRISAN, D. and LYONS, T. J. (1996). Nonlinear filtering and measure valued processes. Imperial College, London. Preprint.

[10] CRISAN, D. and LYONS, T. J. (1996). Convergence of a branching particle method to the solution of the Zakai equation. Imperial College, London. Preprint.
[11] CRISAN, D. and LYONS, T. J. (1996). A particle approximation of the solution of the Kushner–Stratonovitch equation. Imperial College, London. Preprint.
[12] DAWSON, D. A. (1975). Stochastic evolution equations and related measure-valued processes. *J. Multivariate Anal.* **5** 1–52.
[13] DELLACHERIE, C. and MEYER, P. A. (1978). *Probabilités et Potentiel*. Hermann, Paris.
[14] DEL MORAL, P. (1994). Résolution particulaire des problèmes d'estimation et d'optimisation nonlinéaires. Thèse, Univ. Paul Sabatier, Rapport LAAS 94269.
[15] DEL MORAL, P. (1995). Nonlinear filtering using random particles. *Theor. Probab. Appl.* **40**.
[16] DEL MORAL, P. (1996). Asymptotic properties of nonlinear particle filters. Publ. Laboratoire de Statistiques et Probabilités, Univ. Paul Sabatier, 11-96.
[17] DEL MORAL, P. (1996). Nonlinear filtering: Monte Carlo particle resolution. Publ. Laboratoire de Statistiques et Probabilités, Univ. Paul Sabatier, 02-96.
[18] DEL MORAL, P. (1996). Nonlinear filtering: interacting particle resolution. *Markov Processes and Related Fields* **2**.
[19] DEL MORAL, P. and MICLO, L. (1996). On the convergence and the applications of the generalized simulated annealing. Publ. Laboratoire de Statistiques et Probabilités, Univ. Paul Sabatier, 16-96.
[20] DEL MORAL, P., NOYER, J. C. and SALUT, G. (1995). Résolution particulaire et traitement non linéaire du signal: application radar/sonar. *Traitement du Signal*.
[21] DEL MORAL, P., RIGAL, G., NOYER, J. C. and SALUT, G. (1993). Traitement non-linéaire du signal par reseau particulaire: application radar. *Quatorzième colloque GRETSI*, Juan les Pins.
[22] DOBRUSHIN, R. L. (1971). Markov processes with a large number of locally interacting components: existence of a limit process and its ergodicity. *Problems Inform. Transmission* **7** 70–87.
[23] DOBRUSHIN, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.* **15** 459–485.
[24] GORDON, N. J., SALMON, D. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F* **140** 107–113.
[25] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.
[26] HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Univ. Michigan Press, Ann Arbor.
[27] JAZWINSKI, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.
[28] KALLIANPUR, G. and STRIEBEL, C. (1967). Stochastic differential equations occuring in the estimation of continuous parameter stochastic processes. Technical Report 103, Dept. Statistics, Univ. Minnesota.
[29] KUNITA, H. (1971). Asymptotic behavior of nonlinear filtering errors of Markov processes. *J. Multivariate Anal.* **1** 365–393.
[30] LIANG, D. and MCMILLAN, J. (1989). Development of a marine integrated navigation system. Kalman Filter Integration of Modern Guidance and Navigation Systems, AGARD-LS-166, OTAN.
[31] LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.
[32] MÉLÉARD, S. and ROELLY-COPPOLETTA, S. (1987). A propagation of chaos result for a system of particles with moderate interaction. *Stochastic Process. Appl.* **26** 317–332.
[33] MCKEAN, H. P. (1967). Propagation of chaos for a class of nonlinear parabolic equations. *Lecture Series in Differential Equations* **7** 41–57. Catholic Univ.
[34] OCONE, D. L. (1980). Topics in nonlinear filtering theory. Ph.D. dissertation, MIT.
[35] PARDOUX, E. (1989). Filtrage non-linéaire et equations aux dérivés partielles stochastiques associées. *Ecole d'Été de Probabilités de Saint-Flour XIX*, 1989. *Lecture Notes in Math.* **1464** 1–164. Springer, Berlin.
[36] PARTHASARATHY, K. R. (1968). *Probability Measures on Metric Spaces*. Academic Press, New York.

[37] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
[38] SHIGA, T. and TANAKA, H. (1985). Central limit theorem for a system of Markovian particles with mean field interactions. *Z. Wahrsch. Verw. Gebiete* **69** 439–459.
[39] SHIRYAEV, A. N. (1966). On stochastic equations in the theory of conditional Markov processes. *Theory Probab. Appl.* **11** 179–184.
[40] SPITZER, F. (1969). Random processes defined through the interaction of an infinite particle system. *Lecture Notes in Math.* **89** 201–223. Springer, Berlin.
[41] STETTNER, L. (1989). On invariant measures of filtering processes. In *Stochastic Differential Systems. Lecture Notes in Control and Inform. Sci.* **126**. Springer, Berlin.
[42] STRATONOVICH, R. L. (1960). Conditional Markov processes. *Theory Probab. Appl.* **5** 156–178.
[43] SZNITMAN, A. S. (1984). Nonlinear reflecting diffusion process and the propagation of chaos and fluctuations associated. *J. Functional Anal.* **56** 311–336.
[44] SZNITMAN, A. S. (1989). Topics in propagation of chaos. *Ecole d'Été de Probabilités de Saint-Flour XIX*, 1989. *Lecture Notes in Math.* **1464** 165–251.
[45] VAN DOOTINGH, M., VIEL, F., RAKOTOPARA, D. and GAUTHIER, J. P. (1991). Coupling of nonlinear control with a stochastic filter for state estimation: application on a free radical polimerization reactor. In *I.F.A.C. International Symposium ADCHEM'91, Toulouse, France.*
[46] WATANABE, S. (1968). A limit theorem of branching processes and continuous state branching processes. *J. Math. Kyoto Univ.* **8** 141–167.

LABORATOIRE DE STATISTIQUES ET PROBABILITÉS
CNRS UMR C55830, BAT 1R1
UNIVERSITÉ PAUL SABATIER
118 ROUTE DE NARBONNE
31062 TOULOUSE CEDEX
FRANCE
E-MAIL: delmoral@cict.fr