A Uniformly Convergent Adaptive Particle Filter

Anastasia Papavasiliou*

Columbia University

October 20, 2004

Abstract

Particle filters are Monte-Carlo methods that aim to approximate the optimal filter of a partially observed Markov chain. In this paper, we study the case where the transition kernel of the Markov chain depends on unknown parameters: we construct a particle filter for the simultaneous estimation of the parameter and the partially observed Markov chain (adaptive estimation) and we prove the convergence of this filter to the correct optimal filter, as time and the number of particles go to infinity. The filter presented here generalizes Del Moral's Monte-Carlo particle filter, presented in [5].

Key words : nonlinear filtering, particle filters, Bayesian estimation.

AMS subject classifications : 60G35, 93E10, 93E11, 65C50.

1   Introduction

In stochastic filtering, the goal is to estimate the distribution of a stochastic process at any time instant, given some partial information up to that time. This conditional distribution is usually referred to as the *optimal filter*. The basic model consists of a Markov chain X (also called the *state variable*) and possibly nonlinear observations Y with observational noise V independent of the signal X. In this case, the optimal filter is completely determined by the observations, the transition probability kernel, the distribution of the noise, and the initial distribution.

The two most important problems in stochastic filtering are how to treat unknown initial conditions or unknown parameters and how to compute the optimal filter. A common approach for dealing with unknown parameters in the system is to treat them as part of the state variable (see [13] for a historical perspective). In this case, the Bayesian posterior distribution of the parameters is a marginal of the optimal filter.

---

*Department of Applied Physics and Applied Mathematics, 200 Mudd Building, Columbia University, New York, NY-10027. Current address: Chemical Engineering, EQuad A217, Princeton, NJ 08544.

It turns out that under certain conditions on the transition kernels and the prior distribution, the Bayesian estimator is asymptotically consistent and in addition, the optimal filter of the augmented system, i.e. the one where the parameters are included in the state variable, is asymptotically stable with respect to the initial conditions. Thus, if we initialize the optimal filter with any prior distribution on the parameters, this will eventually converge to the true optimal filter, as long as the prior distribution satisfies some condition; namely, the mass put by the prior in the neighborhood of the true value should not disappear "too quickly". For a more detailed discussion regarding the asymptotic stability of the optimal filter of the augmented system, see [14].

In this paper, we discuss the problem of computing the optimal filter for the augmented system. It is well known that with the exception of very few cases (for example, linear Gaussian systems), an exact computation of the optimal filter is not possible and we have to resort to numerical methods. One of the most efficient schemes for the recursive computation of the optimal filter is the interacting particle filter (or sequential Monte-Carlo method), first suggested in [9] and [10], independently. The idea is to approximate the optimal filter by an empirical distribution of particles that evolve in a way that imitates the evolution of the optimal filter. It has been shown that as the number of particles grows, the empirical distribution on these particles converges to the optimal filter, at every time instant (see, for example, [2], [7] or [1]). Under suitable conditions, the uniform convergence of the particle filter has also been shown (see [6]). However, the uniform convergence results assume some ergodicity conditions on the transition kernel of the Markov chain, which will clearly not hold for the augmented system. The only particle filter, to the best of the author's knowledge, that has been shown to converge uniformly and with no assumption on the ergodicity of the transition kernel is the Monte-Carlo particle filter presented in [5]. Unfortunately, this algorithm is not practical, since it requires the dense sampling of the path space of the Markov chain.

In fact, in the case of adaptive estimation, the standard particle filters are not going to converge to the true optimal filter: there is a positive probability that the particles corresponding to the best estimate of the parameter will be lost and once they are lost, they cannot be recovered. Several variations of the interacting particle filter designed for the particular case of adaptive estimation have been proposed (see, for example, [3], [8] or [16]). These algorithms try to avoid this degeneracy by combining the particle filters with other methods of parameter estimation, or by introducing some possibility for the 'rebirth' of a particle. Yet, there is no proof of convergence for any of these filters. We study here a variation of the particle filter suggested in [8] and prove its convergence, under suitable conditions.

As we mentioned above, most adaptive algorithms are based on the idea of treating the estimation of the Markov chain and that of the parameter separately. We also follow a two-layer approach: first, we use a particle filter for the computation of the optimal filter for each fixed value of the parameter. Then, we use another particle filter to compute the posterior distribution of the parameter given the observations, for a given prior distribution on the parameter space. However, the computation of the posterior distribution of the parameter also involves the previous layer of compu-

tations, i.e. the optimal filters for fixed parameter values. Putting these together, we end up computing the weighted average of the optimal filters as functions of the parameter, with respect to posterior distribution of the parameters (see also [8]). Note, though, that in this case, pointwise convergence or convergence of pathwise average approximation errors is not sufficient, since there is an additional error coming from the inexact initialization of the parameter. We need particle filters that converge uniformly with respect to time. Also note that while we can assume the ergodicity of the transition kernel for the first layer of computations, this is not possible for the second layer. These limitations led us into choosing an algorithm that is a combination of the interacting particle filter (for the computation of the optimal filter for fixed parameter value) and a variation of the Monte-Carlo particle filter described in [5] (for the computation of the posterior distribution of the parameter). The computational inefficiency of the Monte-Carlo particle filter is not going to be of any concern in this case, since it is applied to the particles corresponding to the parameters and their paths are trivial (constant), since the parameters do not evolve. We prove that this algorithm will converge uniformly, under certain stability conditions. When the system is linear and Gaussian, we can actually take advantage of the existence of an exact solution for fixed parameter values and construct a simpler particle filter that is a combination of the Kalman-Bucy filter and the Monte-Carlo particle filter.

The structure of this paper is the following: we first describe the main assumptions and some previous results we will be using, regarding the asymptotic stability of the optimal filter. In section 3, we study the case where the system is linear and Gaussian, but it depends on unknown parameters. We describe an algorithm that is a combination of the Monte-Carlo particle filter and the Kalman-Bucy filter (or, in fact, any method that gives an exact computation of the optimal filter for fixed values of the parameter – if such a method exists). We show that this algorithm will actually converge uniformly with respect to time.

In section 4, we describe the full algorithm, which is a combination of the interacting and the Monte-Carlo particle filter – once again, any algorithm that converges uniformly in time for fixed parameter values can replace the interacting particle filter. Finally, in section 5, we show that such an algorithm will also converge uniformly with respect to time.

2   Setting

Let E be a Polish space, i.e., a complete separable metric space and let us denote by $\mathcal{B}(E)$ its Borel σ-field. We construct below a numerical scheme for the approximation of the conditional distribution of a Markov chain $\{X_n\}$ taking values in E, given some noisy partial information, when the kernel depends on an unknown parameter θ. More specifically, we want to approximate the optimal filter of the following system, which we will refer to as:

System 1. *Let $\{X_n\}$ be a homogeneous Markov chain taking values in $(E, \mathcal{B}(E))$. Let μ be its initial distribution and K_θ its transition probability kernel depending on*

a parameter $\theta \in \Theta$. Furthermore, we assume that for each $\theta \in \Theta$, K_θ is Feller and mixing, i.e. there exists a constant $0 < \epsilon_\theta \leq 1$ and a nonnegative measure $\lambda_\theta \in \mathcal{M}^+(E)$ ($\mathcal{M}^+(E)$ being the set of finite nonnegative measures on E), such that

$$\epsilon_\theta \lambda_\theta(A) \leq K_\theta(x, A) \leq \frac{1}{\epsilon_\theta} \lambda_\theta(A), \quad \forall x \in E \text{ and } \forall A \in \mathcal{B}(E). \tag{1}$$

The observation process is defined by

$$Y_n = h(X_n) + V_n,$$

where V_n are i.i.d. Gaussian($0,\sigma^2$) $\mathbb{R}^p$-valued random variables independent of X and $h : E \to \mathbb{R}^p$ is a bounded continuous function. We denote by g the Gaussian probability density function of the observational noise.

In practice, we usually take the parameter space Θ to be a compact subset of Euclidean space. More generally, we assume that it is a compact Polish space, with metric $d_\Theta(\cdot, \cdot)$.

Most problems are given in the form of System 1. Following a standard Bayesian technique, we rewrite the system, so that the parameter becomes part of the Markov chain, whose transition probability kernel is now completely known. Note, however, that since we do not know the value of the parameter, we assign to it some prior distribution u. The new system constructed in this way can be formally defined as follows:

System 2. *Suppose that $\{\tilde{X}_n = (X_n, \theta_n)\}$ is an $E \times \Theta$-valued homogeneous Markov chain, with transition probability*

$$\tilde{K}((x', \theta'), dx \otimes d\theta) = K_\theta(x', dx) \otimes \delta_{\theta'}(d\theta).$$

and initial distribution $\mu \otimes u$, where K_θ is Feller and mixing in the sense of (1). The observation process is defined by

$$Y_n = \tilde{h}(\tilde{X}_n) + V_n,$$

where $\tilde{h}(\tilde{x}) = h(x)$ and $\tilde{x} = (x, \theta)$.

System 2 can be thought of as a generalization of System 1 and will be the main object of study in this paper: our goal is to construct a particle filter that converges to the optimal filter of system 2 uniformly with respect to time.

The canonical space of the Markov chain X with kernel K_θ and initial distribution μ is denoted by $(\Omega_1 = E^{\mathbb{N}}, (\mathcal{F}_n^{(X)})_{n \geq 0}, P_{\mu, \theta})$, where $\mathcal{F}_n^{(X)} = \sigma(X_0, X_1, \ldots, X_n)$ is the σ-algebra constructed by the random variables $X_0, X_1, \ldots, X_n$. Similarly, the observation process is defined on the canonical space $(\Omega_2 = (\mathbb{R}^p)^{\mathbb{N}}, (\mathcal{F}_n^{(Y)})_{n \geq 0}, Q_{\mu, \theta})$, where $\mathcal{F}_n^{(Y)} = \sigma(Y_1, \ldots, Y_n)$. The law of the observation process $Q_{\mu, \theta}$ is given by

$$Q_{\mu, \theta}(dy_{k_1}, \ldots, dy_{k_n}) = \int_{E^{\otimes n}} \prod_{i=1}^{n} g(y_{k_i} - h(x_{k_i})) P_{\mu, \theta}(dx_{k_1}, \ldots, dx_{k_n}) dy_{k_1} \ldots dy_{k_n},$$

for any $n \geq 0$ and $k_1, \ldots, k_n \in \mathbb{R}_+$, where $E^{\otimes n} = E \times \cdots \times E$ is the product space of n copies of E. Then, we can define the pair process (X, Y) on the space $(\Omega = \Omega_1 \times \Omega_2, (\mathcal{F}_n = \mathcal{F}_n^{(X)} \times \mathcal{F}_n^{(Y)})_{n \geq 0}, \mathbb{P}_{\mu,\theta})$, where the measure $\mathbb{P}_{\mu,\theta}$ is such that its marginal distributions with respect to X and Y are $P_{\mu,\theta}$ and $Q_{\mu,\theta}$ respectively. It is not hard to show that this measure exists (see, for example, [5]). We will denote the expectation with respect to $\mathbb{P}_{\mu,\theta}$ by $\mathbb{E}_{\mu,\theta}$.

Similarly, we define the triplet (X, Y, θ) on the space $(\tilde{\Omega} = \Omega \times \Omega_3, (\tilde{\mathcal{F}}_n = \mathcal{F}_n \times \sigma(\theta))_{n \geq 0}, \tilde{\mathbb{P}}_{\mu,u})$, where θ is a Θ-valued random variable defined on $(\Omega_3, \sigma(\theta), u)$ and the marginals of $\tilde{\mathbb{P}}_{\mu,u}$ on (X, Y) and θ respectively are $\int_\Theta \mathbb{P}_{\mu,\theta} u(d\theta)$ and u. We will denote the expectation with respect to $\tilde{\mathbb{P}}_{\mu,u}$ by $\tilde{\mathbb{E}}_{\mu,u}$.

Furthermore, we denote by $\Psi_n^\theta(\mu)$ and $\Phi_n(\mu \otimes u)$ the optimal filters for Systems 1 and 2, with initial distributions μ and $\mu \otimes u$, respectively, defined as the posterior distribution of the state variable given the observations. The name "optimal filters" is due to the fact that they are the best estimators adapted to the available information (the σ-algebra constructed by the observations), with respect to the L_2-norm. They are random measures on the space E and $E \times \Theta$ respectively, defined as follows: for every $f \in \mathcal{C}_b(E \times \Theta)$,

$$\Phi_n(\mu \otimes u)(f) = \tilde{\mathbb{E}}_{\mu,u}[f(X_n, \theta)|Y_n, \ldots, Y_1] =$$

$$= \frac{\int_\Theta \int_{E^{\otimes n}} f(x_n, \theta) \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu,\theta}(dx_1, \ldots dx_n) u(d\theta)}{\int_\Theta \int_{E^{\otimes n}} \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu,\theta}(dx_1, \ldots, dx_n) u(d\theta)}. \quad (2)$$

Similarly, $\forall f' \in \mathcal{C}_b(E)$,

$$\Psi_n^\theta(\mu)(f') = \mathbb{E}_{\mu,\theta}[f'(X_n)|Y_n, \ldots, Y_1] =$$

$$= \frac{\int_{E^{\otimes n}} f'(x_n) \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu,\theta}(dx_1, \ldots, dx_n)}{\int_{E^{\otimes n}} \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu,\theta}(dx_1, \ldots, dx_n)}. \quad (3)$$

Clearly, $\Psi_n^\alpha(\mu)$ is the marginal of $\Phi_n(\mu \otimes \delta_\alpha)$ with respect to the state variable X, since $\Phi_n(\mu \otimes \delta_\alpha)(f) = \Psi_n^\alpha(\mu)(f_\alpha)$, where we used the notation $f_\alpha(x) = f(x, \alpha)$. Similarly, we define $\Psi_n^u(\mu)$ as the marginal of $\Phi_n(\mu \otimes u)$ with respect to X, i.e.,

$$\Psi_n^u(\mu)(dx) = \int_\Theta \Phi_n(\mu \otimes u)(dx, d\theta).$$

The mixing condition (1) implies the ergodicity of the signal. We denote by μ_θ the limiting distribution of the Markov chain whose transition kernel is K_θ. The measure μ_θ is uniquely defined in this way, as a result of the ergodic property of the signal. This property is also transferred to the observation process: its limiting distribution is uniquely defined by $\nu_\theta = (\mu_\theta \circ h^{-1}) * g$.

It has been shown that the optimal filters will be asymptotically stable with respect to the initial conditions in the setting of system 2, under suitable conditions on the prior distribution and the kernels (see [14]):

$$\lim_{n\to\infty} \mathbb{E}_{\mu,\alpha} \|\Psi_n^u(\mu') - \Psi_n^\alpha(\mu)\|_{tv} = 0.$$

In this paper, we will assume the slightly stronger result

$$\lim_{q\to\infty} \sup_{n\geq q} \mathbb{E}_{\mu,\alpha} \|\tilde{\Psi}_{n-q,n}^u(\mu') - \Psi_n^u(\mu)\|_{tv} = 0, \tag{4}$$

where $\tilde{\Psi}_{n-q,n}^u(\mu')(dx) = \int_\Theta \Psi_n^\theta(\mu')(dx)\mathbb{P}_{\Psi_{n-q}^\theta(\mu'),u}(d\theta|Y_n,\ldots,Y_{n-q+1})$. This definition is consistent with the way we approximate the posterior distribution of the parameter θ in (5) and (9). The proof of this is similar to the proof of asymptotic stability, once we note that $\mathbb{P}_{\Psi_{n-q}^\theta(\mu'),u}(d\theta|Y_n,\ldots,Y_{n-q+1}) = \mathbb{P}_{\mu',u}(d\theta, Y_{n-q},\ldots,Y_1|Y_n,\ldots,Y_{n-q+1})$.

We construct below a particle filter that approximates $\Psi_n^u(\mu)$ uniformly with respect to time. Then, it is an obvious corollary that the particle filter will converge to the true one, as time n and the number of particles N go to infinity, i.e. $(n, N) \to (\infty, \infty)$.

3 Adaptive Estimation of Linear Gaussian Systems

We will first construct a particle filter for the adaptive estimation of a linear Gaussian system, i.e. a system of the form

$$\begin{cases} X_n = F(\theta)X_{n-1} + c(\theta) + G(\theta)W_n \\ Y_n = HX_n + d + V_n \end{cases},$$

where for every $\theta \in \Theta$, the coefficients $F(\theta)$, $c(\theta)$, $G(\theta)$, H, and d are deterministic and they belong to the Euclidean spaces $\mathbb{R}^{p\times p}$, $\mathbb{R}^p$, $\mathbb{R}^{p\times p}$, $\mathbb{R}^{q\times p}$, $\mathbb{R}^q$ respectively. In addition, both the state and the observation noises, $W_n \in \mathbb{R}^p$ and $V_n \in \mathbb{R}^q$, are mean zero i.i.d. Gaussian sequences, with covariance $\mathcal{C}ov(W_n) = Q$ and $\mathcal{C}ov(V_n) = \Sigma$, independent of each other and of the initial distribution $\mathcal{L}(X_0)$ of X_0. Moreover, we assume that initial distribution of X_0 is also Gaussian, with mean $\mathbb{E}(X_0) = \bar{X}_0$ and covariance Q_0.

Then, for fixed θ, the optimal filter of this system is the Kalman-Bucy filter which can be computed exactly. However, once we enter the parameter in the system (as in system 2), the linearity is lost and the Kalman-Bucy filter cannot be applied. Below, we construct an algorithm which is a combination of the Kalman-Bucy filter for the state variables and the Monte-Carlo particle filter for the parameters and we show that this filter will converge uniformly with respect to time. In fact, both the algorithm and the theorem are given in the more general formulation of system 1, under the assumption that for each fixed value of the parameter θ, the optimal filter can be computed exactly.

3.1 The Algorithm

The following algorithm is based on the Monte-Carlo particle filter presented in [5], where the role of the state variable is played by the optimal filters.

At time n=0 (Initialization):

The particle system consists of M independent particles $\{\theta_j\}_{j=1,\ldots,M}$ with common law u. The weights $W_{0,0}^{j,M}$ of the parameters are set to be equal to $\frac{1}{M}$, i.e. $W_{0,0}^{j,M} = \frac{1}{M}$. Then, the particle filter is defined as

$$\Psi_0^{u,M}(\mu) := \sum_{j=1}^{M} W_{0,0}^{j,M} \Psi_0^{\theta_j}(\mu) = \frac{1}{M} \sum_{j=1}^{M} \mu = \mu,$$

where μ is the initial distribution of the Markov chain (a Gaussian, in the case where Ψ_n^θ is the Kalman-Bucy filter).

For $n \geq 0$ (Evolution):

We compute the new weights $W_{(n-q(M))^+,n}^{j,M}$, for q a non-decreasing function such that $\lim_{M\uparrow\infty} q(M) = \infty$, as follows:

$$W_{(n-q(M))^+,n}^{j,M} = \frac{Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_j}(\mu), Y)}{\sum_{k=1}^{M} Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y)}, \tag{5}$$

where $\hat{\Psi}_n^\theta(\mu)(dx) := \mathbb{P}_{\mu,\theta}(X_n \in dx | y_{n-1}, \ldots, y_1)$ is the optimal one-step predictor for fixed θ (the Kalman-Bucy predictor, in the case of the linear Gaussian system); for each $j = 1, \ldots, M$

$$Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_j}(\mu), Y) := \prod_{k=(n-q(M))^++1}^{n} \int_E (\sqrt{(2\pi)}\sigma) g(Y_k - h(z_k)) \hat{\Psi}_k^{\theta_j}(\mu)(dz_k) \tag{6}$$

and we used the notation $a^+ = \max\{a, 0\}$ for all $a \in \mathbb{R}$.

The particle filter is defined by

$$\Psi_n^{u,M}(\mu) := \sum_{j=1}^{M} W_{(n-q(M))^+,n}^{j,M} \Psi_n^{\theta_j}(\mu). \tag{7}$$

It is crucial for the convergence of the algorithm to consider a finite window of information: even though the denominator is always positive, it converges to zero very fast since it is a product of numbers less than one. Note here that multiplying g with $\sqrt{2\pi}\sigma$ in (6) only simplifies notation, since it cancels out in the computation of the weights (5). Thus, considering a window of information of length $q(M)$ allows us to control the way this converges to zero (see also [5]).

In practice, things are even worse: numbers below some threshold are treated as equal to zero in computer arithmetic and consequently, just a series of 'unlikely'

observations with respect to some θ can set the corresponding weight equal to zero. This weight will of course stay zero for ever, so we are back to the problem of 'losing' particles. To remedy that, one can set the weights to be $\max\{W^{j,M}_{(n-q(M))^+,n}, \varepsilon\}$ for some very small ε.

In order to account for the additional randomness in the choice of $\bar\theta^M = (\theta_1, \ldots, \theta_M)$, we define a new probability space $(\tilde\Omega' = \tilde\Omega \times \Omega_3^M, (\tilde{\mathcal{F}}'_n = \tilde{\mathcal{F}}_n \times \sigma(\theta_1, \ldots, \theta_M))_{n\geq 0}, \tilde{\mathbb{P}}'_{\mu,u} = \tilde{\mathbb{P}}_{\mu,u} \otimes u^M)$ and the corresponding expectation will be denoted by $\tilde{\mathbb{E}}'_{\mu,u}$. Similarly, we define the probability space $(\Omega' = \Omega \times \Omega_3^M, (\mathcal{F}'_n = \mathcal{F}_n \times \sigma(\theta_1, \ldots, \theta_M))_{n\geq 0}, \mathbb{P}'_{\mu,\theta,u} = \mathbb{P}_{\mu,\theta} \otimes u^M)$ and we denote by $\mathbb{E}'_{\mu,\theta,u}$ the respective expectation.

The following theorem extends the main theorem in [5]:

Theorem 3.1. *Suppose that (4) holds and that*

$$\sup_n \mathbb{E}_{\mu,\alpha} h(X_n)^2 < \infty \quad \text{and}$$
$$\sup_{n,\theta} \mathbb{E}_{\mu,\alpha} \mathbb{E}_{\mu,\theta}(h(X_n)|Y_n, \ldots, Y_1)^2 < \infty.$$

(This holds trivially in the case h is bounded, but we include this condition to also cover linear Gaussian systems with unknown parameters). Then, there exists a non-decreasing function $q : \mathbb{N} \to \mathbb{N}$ (for example, $q(M) = \frac{1}{2}\sqrt{\log(M)}$) for which the error made by the particle filter $\Psi_n^{u,M}(\mu)$ given by (7) will converge to zero uniformly with respect to time:

$$\lim_{M\uparrow\infty} \sup_{n\geq 0} \mathbb{E}'_{\mu,\alpha,u} | \sum_{j=1}^M W^{j,M}_{(n-q(M))^+,n} \Psi_n^{\theta_j}(\mu)(f) - \Psi_n^u(\mu)(f)| = 0. \tag{8}$$

Proof. The theorem can be proved by following the arguments of the proof of uniform convergence for the Monte-Carlo particle filter (see [5]).

$\square$

4   The Full Algorithm

In [6], Del Moral and Guionnet show the uniform convergence of the interacting particle filter, but under some strong ergodicity conditions that are not satisfied by system 2. In [5], Del Moral proves the uniform convergence of the Monte Carlo particle filter under weaker conditions. However, the Monte-Carlo particle filter is not very efficient computationally; it is a weighted sum over independent paths and thus, it requires a dense sampling of the path space.

We present below a combination of the interacting particle filter and the Monte-Carlo particle filter: we use the interacting particle filter to compute the posterior distribution of the first component of system 2, the Markov chain $\{X_n\}$, while we fix the second component, the parameter. Since for each fixed value of the parameter the kernels K_θ are mixing, we expect that the interacting particle filter will converge uniformly to the marginal of the optimal filter with respect to the first component.

In order to compute the posterior distribution of the parameter, we use the Monte-Carlo particle filter. Since the parameters do not evolve, their path space is trivial

and its dimension is the dimension of the parameter space. Thus, in this case, the two filters are computationally equivalent. On the other hand, it is crucial that we relax the ergodicity assumptions on the transition kernel – in this case, the kernel for the evolution of the parameter is $\delta_{\theta'}(d\theta)$ which is obviously not ergodic! We have already seen in theorem 3.1 that the Monte-Carlo particle filter on the parameters will converge uniformly with respect to time, given the posterior distribution of $\{X_n\}$. We will show that this will also be the case for the combination of the two particle filters: they will converge uniformly with respect to time.

Intuitively, the reasoning is the following. The interacting particle filter resamples the particles. Thus, an unlikely observation can result in the loss of the 'good particles' on the parameters, i.e. the ones close to the correct value of the parameter. Instead, the Monte-Carlo particle filter computes the weighted average. So, no particle is lost and the weight depends on several past observations rather than just the last one.

Below, we give a detailed description of the algorithm (q: a non-decreasing function, such that $\lim_{N\uparrow\infty} q(N) = \infty$).

At time n=0 (Initialization):

The particle system consists of M independent particles $\{\theta_j\}_{j=1,\ldots,M}$ with common law u and $N * M$ independent particles $\{\xi_i^0(\theta_j)\}_{i=1,\ldots,N;j=1,\ldots,M}$ with common law μ. The weights of the parameters are set to be equal, i.e. $W_{0,0}^{j,M,N} = \frac{1}{M}$. Then,

$$\Psi_0^{u,M,N}(\mu) := \sum_{j=1}^{M} W_{0,0}^{j,M,N} \Psi_0^{\theta_j,N}(\mu) = \frac{1}{NM} \sum_{i,j=1}^{N,M} \delta_{\xi_i^0(\theta_j)},$$

where $\Psi_0^{\theta_j,N}(\mu)$ is the interacting particle filter at time $n = 0$, corresponding to parameter value θ_j, i.e.

$$\Psi_0^{\theta_j,N}(\mu) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i^0(\theta_j)}.$$

For $n \geq 0$ (Evolution):

1. Each particle $\xi_i^n(\theta_j)$, $i = 1, \ldots, N$ evolves according to the kernel K_{θ_j}, for $j = 1, \ldots, M$, i.e.
   $$\hat{\xi}_i^n(\theta_j) \sim K_{\theta_j}(\xi_i^n(\theta_j), \cdot).$$

   We set
   $$\hat{\Psi}_n^{\theta_j,N}(\mu) := \frac{1}{N} \sum_{i=1}^{N} \delta_{\hat{\xi}_i^n(\theta_j)}.$$

2. We compute the new weights $W_{(n-q(M))^+,n}^{j,M,N}$, as follows:

   $$W_{(n-q(M))^+,n}^{j,M,N} = \frac{Z_{(n-q(M))^+,n}\big(\hat{\Psi}^{\theta_j,N}(\mu), Y\big)}{\sum_{k=1}^{M} Z_{(n-q(M))^+,n}\big(\hat{\Psi}^{\theta_k,N}(\mu), Y\big)}, \tag{9}$$

where $a^+ = \max\{a, 0\}, \forall a \in \mathbb{R}$ and for each $j = 1, \ldots, M$,

$$Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_j,N}(\mu), Y) := \prod_{k=(n-q(M))^++1}^{n} \int_E g(Y_k - h(z_k))\hat{\Psi}_k^{\theta_j,N}(\mu)(dz_k) =$$

$$= \prod_{k=(n-q(M))^++1}^{n} \left(\frac{1}{N}\sum_{i=1}^{N} g(Y_k - h(\hat{\xi}_i^k(\theta_j)))\right). \quad (10)$$

3. For each $j = 1, \ldots, M$, we resample the particles $\{\hat{\xi}_i^n(\theta_j)\}_{i=1,\ldots,N}$ according to weights $\frac{g(Y_n - h(\hat{\xi}_i^n(\theta_j)))}{\sum_{i=1}^{N} g(Y_n - h(\hat{\xi}_i^n(\theta_j)))}$ to get $\{\xi_i^{n+1}(\theta_j)\}_{i=1,\ldots,N}$.

Then, the particle filter is defined by

$$\Psi_n^{u,M,N}(\mu) := \sum_{j=1}^{M} W_{(n-q(M))^+,n}^{j,M,N} \Psi_n^{\theta_j,N}(\mu), \quad (11)$$

where $\Psi_n^{\theta_j,N}(\mu)$ is the interacting particle filter corresponding to parameter value θ_j, i.e.

$$\Psi_n^{\theta_j,N}(\mu) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_i^n(\theta_j)}. \quad (12)$$

Once again, we have to consider a new probability space that will also capture the additional randomness in the evolution of the particle filters. We denote by $(\Omega_\xi, \mathcal{F}_n^\xi, \mathbb{P}_{\mu,u}^\xi)$ the canonical space for the particle system $\{(\xi_n^{i,N}(\theta_j), \theta_j)\}_{i=1\ldots,N,j=1,\ldots,M,n>0}$, which is assumed to be independent of the state-observation pair (X, Y). Then, the new probability space will be $(\hat{\Omega}, (\hat{\mathcal{F}}_n)_{n\geq 0}, \hat{\mathbb{P}}_{\mu,\theta,u})$, where $\hat{\Omega} = \Omega \times \Omega_\xi$, $\hat{\mathcal{F}}_n = \mathcal{F}_n \times \mathcal{F}_n^\xi$ and $\hat{\mathbb{P}}_{\mu,\theta,u} = \mathbb{P}_{\mu,\theta} \otimes \mathbb{P}_{\mu,u}^\xi$. Similarly, the expectation with respect to $\hat{\mathbb{P}}_{\mu,\theta,u}$ will be denoted by $\hat{\mathbb{E}}_{\mu,\theta,u}$. To keep notation simpler, we will often drop the u when it is clear what it is.

5 Uniform Convergence of the Full Algorithm

The following lemma describes the conditions for the uniform convergence of the interacting particle filter. The proof can be found in [6] or [7]. We alter the statement, though, to fit it in the setting of this paper.

Lemma 5.1. *Suppose that $\forall \theta \in \Theta$, the following holds $\forall f \in \mathcal{C}_b(E)$:*

$$\lim_{q\to\infty} \sup_{\mu,\mu'\in\mathcal{P}(E)} \sup_{n\geq q} |\Psi_{n-q,n}^\theta(\mu)(f) - \Psi_{n-q,n}^\theta(\mu')(f)| = 0, \quad (13)$$

where $\Psi_{n-q,n}^\theta(\mu) = \mathbb{P}_{\mu,\theta}(X_n|Y_n, \ldots, Y_{n-q+1})$. Then, $\forall f \in \mathcal{C}_b(E)$

$$\lim_{N\to\infty} \sup_{n\geq 0} \hat{\mathbb{E}}_{\mu,\alpha}|\Psi_n^{\theta,N}(\mu)(f) - \Psi_n^\theta(\mu)(f)| = 0,$$

where $\Psi_n^{\theta,N}$ is the interacting particle filter applied to system 1, for fixed parameter value θ.

In fact, if for each $\theta \in \Theta$, system 1 is exponentially asymptotically stable, i.e. $\exists \gamma_\theta > 0$ and $q_{0,\theta} > 0$ such that,

$$\forall \mu, \mu' \in \mathcal{P}(E), \ \forall q \geq q_{0,\theta}, \ \sup_{n \geq q} |\Psi_{n-q,n}^{\theta}(\mu)(f) - \Psi_{n-q,n}^{\theta}(\mu')(f)| \leq e^{-\gamma_\theta q},$$

then $\exists N_{0,\theta} > 0$ such that $\forall N \geq N_{0,\theta}$

$$\sup_{n \geq 0} \hat{\mathbb{E}}_{\mu,\alpha} |\Psi_n^{\theta,N}(\mu)(f) - \Psi_n^{\theta}(\mu)(f)| \leq \frac{C_\theta e^{\gamma'}}{N^{\frac{a_\theta}{2}}},$$

where C_θ is some constant, $\gamma' = 1 - \log(2\pi\sigma^2)$ and $a_\theta = \frac{\gamma_\theta}{\gamma_\theta + \gamma} > 0$.

We will prove the following result, under the general assumption that the particle filter approximating the optimal filter for fixed parameter value converges uniformly with respect to time and parameter, with rate $p(N)$. Such a particle filter could be the interacting particle filter, under the appropriate conditions described above, or some other particle filter (for example, the regularized particle filter described in [12]), if, in addition, the parameter space is compact and the optimal filters uniformly continuous with respect to the parameter.

Theorem 5.2. *Suppose that (4) holds and, in addition, that the interacting particle filter used to approximate the optimal filter for any fixed parameter value converges uniformly with rate $p(N)$, i.e. it satisfies*

$$\sup_{n \geq 0} \hat{\mathbb{E}}_{\mu,\alpha} \|\Psi_n^{\theta,N}(\mu) - \Psi_n^{\theta}(\mu)\|_{tv} \leq p(N), \ \forall \theta \in \Theta, \tag{14}$$

where $p(N)$ is independent of θ and $\lim_{N \uparrow \infty} p(N) = 0$. Then, there exists a function $M = M(N)$ such that $\lim_{N \uparrow \infty} M(N) = \infty$ and the following holds for all $f \in \mathcal{C}_b(E)$:

$$\lim_{N \uparrow \infty} \sup_{n \geq 0} \hat{\mathbb{E}}_{\mu,\alpha} |\Psi_n^{u,M(N),N}(\mu)(f) - \Psi_n^{u}(\mu)(f)| = 0. \tag{15}$$

Proof. For simplicity of notation, we assume that $p = 1$, i.e. the observations take values in $\mathbb{R}$. We first show that $\forall f \in \mathcal{C}_b(E)$,

$$\lim_{N \uparrow \infty} \sup_{M > 0} \sup_{n \geq 0} \hat{\mathbb{E}}_{\mu,\alpha} |\Psi_n^{u,M,N}(\mu)(f) - \sum_{j=1}^{M} W_{(n-q(M))^+,n}^{j,M,N} \Psi_n^{\theta_j}(\mu)(f)| = 0 \tag{16}$$

We note that

$$\hat{\mathbb{E}}_{\mu,\alpha} |\Psi_n^{u,M,N}(\mu)(f) - \sum_{j=1}^{M} W_{(n-q(M))^+,n}^{j,M,N} \Psi_n^{\theta_j}(\mu)(f)| =$$

$$= \hat{\mathbb{E}}_{\mu,\alpha} |\sum_{j=1}^{M} W_{(n-q(M))^+,n}^{j,M,N} \Psi_n^{\theta_j,N}(\mu)(f) - \sum_{j=1}^{M} W_{(n-q(M))^+,n}^{j,M,N} \Psi_n^{\theta_j}(\mu)(f)| \leq$$

$$\leq \hat{\mathbb{E}}_{\mu,\alpha} \sum_{j=1}^{M} (W_{(n-q(M))^+,n}^{j,M,N} |\Psi_n^{\theta_j,N}(\mu)(f) - \Psi_n^{\theta_j}(\mu)(f)|) =$$

$$= \sup_j \hat{\mathbb{E}}_{\mu,\alpha} |\Psi_n^{\theta_j,N}(\mu)(f) - \Psi_n^{\theta_j}(\mu)(f)| \leq \|f\|_\infty p(N). \tag{17}$$

Thus, (16) follows from (14). Note that in order to simplify the notation, the dependence of $M = M(N)$ to N has been suppressed and will only be shown when relevant. Next, we show that

$$\limsup_{N\uparrow\infty} \sup_{n\geq 0} \hat{\mathbb{E}}_{\mu,\alpha} | \sum_{j=1}^{M} W^{j,M,N}_{(n-q(M))^+,n} \Psi^{\theta_j}_n(\mu)(f) - \sum_{j=1}^{M} W^{j,M}_{(n-q(M))^+,n} \Psi^{\theta_j}_n(\mu)(f)| = 0, \quad (18)$$

where $W^{j,M}_{(n-q(M))^+,n}$ is the likelihood function for parameter value θ_j, i.e.

$$W^{j,M}_{(n-q(M))^+,n} = \frac{Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_j}(\mu), Y)}{\sum_{k=1}^{M} Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y)} = \tilde{\mathbb{P}}_{\mu,u^M}(\theta_j | Y_n, \ldots, Y_{(n-q(M))^++1}),$$

where $Z_{(n-q(M))^+,n}$ is defined in (10). Clearly,

$$\hat{\mathbb{E}}_{\mu,\alpha} | \sum_{j=1}^{M} W^{j,M,N}_{(n-q(M))^+,n} \Psi^{\theta_j}_n(\mu)(f) - \sum_{j=1}^{M} W^{j,M}_{(n-q(M))^+,n} \Psi^{\theta_j}_n(\mu)(f)| \leq$$

$$\leq \hat{\mathbb{E}}_{\mu,\alpha} \sum_{j=1}^{M} (|W^{j,M,N}_{(n-q(M))^+,n} - W^{j,M}_{(n-q(M))^+,n}| \cdot |\Psi^{\theta_j}_n(\mu)(f)|) \leq$$

$$\leq \|f\|_\infty \hat{\mathbb{E}}_{\mu,\alpha} \sum_{j=1}^{M} |W^{j,M,N}_{(n-q(M))^+,n} - W^{j,M}_{(n-q(M))^+,n}|.$$

Thus, it is sufficient to show that

$$\limsup_{N\uparrow\infty} \sup_{n\geq 0} \hat{\mathbb{E}}_{\mu,\alpha} \sum_{j=1}^{M} |W^{j,M,N}_{(n-q(M))^+,n} - W^{j,M}_{(n-q(M))^+,n}| = 0.$$

We follow the same approach as in [5], in order to bound the denominator away from zero. We define the following quantities:

$$V^{q(M)}_n(Y,\theta) = \frac{1}{2\sigma^2} \sum_{k=(n-q(M))^++1}^{n} \bar{\mathbb{E}}_{\mu,\theta}[(Y_k - h(X_k))^2] =$$

$$= \frac{1}{2\sigma^2} \sum_{k=(n-q(M))^++1}^{n} [(Y_k - \bar{\mathbb{E}}_{\mu,\theta} h(X_k))^2 + \bar{\mathbb{E}}_{\mu,\theta}(h(X_k) - \bar{\mathbb{E}}_{\mu,\theta} h(X_k))^2],$$

where the expectation $\bar{\mathbb{E}}_{\mu,\theta}$ is with respect to measure $\hat{\Psi}^\theta(\mu) = \otimes_{k=0}^{\infty} \hat{\Psi}^\theta_k(\mu)$. By the boundedness of h we get that $\forall \theta \in \Theta$,

$$\sup_{n\geq 0}[(h(X_n) - \bar{\mathbb{E}}_{\mu,\theta} h(X_n))^2] \leq 4\|h\|^2_\infty < \infty.$$

Then

$$\mathbb{E}_{\mu,\alpha} V^{q(M)}_n(Y,\theta) \leq \frac{1}{2\sigma^2} q(M)[\sup_k \mathbb{E}_{\mu,\alpha}(V_k + h(X_k) - \bar{\mathbb{E}}_{\mu,\theta} h(X_k))^2 + 4\|h\|^2_\infty] \leq$$

$$\leq \frac{1}{2\sigma^2} q(M)(8\|h\|^2_\infty + 4\sigma\|h\|_\infty + \sigma^2) = V(h)q(M) < +\infty,$$

where we used the notation $V(h) = \frac{8\|h\|_\infty^2 + 4\sigma\|h\|_\infty + \sigma^2}{2\sigma^2}$. Now consider the set

$$\Omega_Y(n, r, q(M)) = \{\omega \in \Omega_2 : \sup_{\theta \in \Theta} V_n^{q(M)}(Y(\omega), \theta) \le r(q(M))\},$$

for some non-decreasing function $r : \mathbb{N} \to \mathbb{R}^+$, to be identified later. Then

$$\mathbb{P}_{\mu,\alpha}(\Omega_2 - \Omega_Y(n, \theta, q(M))) \le V(h)\frac{q(M)}{r(q(M))},$$

(see [5]) and consequently

$$\hat{\mathbb{E}}_{\mu,\alpha}\{\sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}|(\mathbf{1}_{\Omega_Y(n,\theta,q(M))} + \mathbf{1}_{\Omega_2 - \Omega_Y(n,\theta,q(M))})\} \le$$

$$\hat{\mathbb{E}}_{\mu,\alpha}\{\sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}|\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} \qquad +2V(h)\frac{q(M)}{r(q(M))}.$$

The second term in the right hand side is going to vanish as $M \uparrow \infty$, for any choice of r, so that $\lim_{q\uparrow\infty} \frac{q}{r(q)} = 0$. It remains to show that

$$\lim_{N\uparrow\infty} \sup_{n \ge 0} \hat{\mathbb{E}}_{\mu,\alpha}\{\sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}|\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} = 0.$$

We define $U_n^{M,N}$ as

$$U_n^{M,N} = \frac{\sum_{i=1}^M Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i,N}(\mu), Y)}{\sum_{k=1}^M Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y)}.$$

Then, we observe the following:

$$\hat{\mathbb{E}}_{\mu,\alpha} \quad \{\sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}|U_n^{M,N}\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} =$$

$$= \quad \hat{\mathbb{E}}_{\mu,\alpha} \quad \{\frac{\sum_{j=1}^M |Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i,N}(\mu), Y) - Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i}(\mu), Y)U_n^{M,N}|}{\sum_{k=1}^M Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y)}\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} \le$$

$$\le \quad \hat{\mathbb{E}}_{\mu,\alpha} \quad \{\frac{\sum_{j=1}^M |Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i,N}(\mu), Y) - Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i}(\mu), Y)|}{\sum_{k=1}^M Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y)}\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} +$$

$$+ \quad \hat{\mathbb{E}}_{\mu,\alpha}\{|1 - U_n^{M,N}|\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} \le$$

$$\le \quad 2\hat{\mathbb{E}}_{\mu,\alpha} \quad \{\frac{\sum_{j=1}^M |Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i,N}(\mu), Y) - Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i}(\mu), Y)|}{\sum_{k=1}^M Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y)}\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\}.$$

Similarly

$$\hat{\mathbb{E}}_{\mu,\alpha} \quad \{\sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}|(1 - U_n^{M,N})\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} =$$

$$\le \quad 2\hat{\mathbb{E}}_{\mu,\alpha} \quad \{\frac{\sum_{j=1}^M |Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i,N}(\mu), Y) - Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i}(\mu), Y)|}{\sum_{k=1}^M Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y)}\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\}.$$

However, $\forall \omega \in \Omega_Y(n, \theta, q(M))$ we see that

$$\tfrac{1}{M} \sum_{k=1}^M Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_k}(\mu), Y(\omega)) =$$

$$= \tfrac{1}{M} \sum_{k=1}^M \bar{\mathbb{E}}_{\theta_k}(\exp\{-\tfrac{1}{2\sigma^2} \sum_{j=(n-q(M))^++1}^n (Y_j(\omega) - h(X_j))^2\}) \geq$$

$$\geq \tfrac{1}{M} \sum_{k=1}^M \exp[-V_n^{q(M)}(Y(\omega), \theta_k)] \geq \exp[-\tfrac{1}{M} \sum_{k=1}^M V_n^{q(M)}(Y(\omega), \theta_k)] \geq \exp(-r(q(M)))$$

by applying Jensen's inequality twice. Putting everything together, we get that

$$\hat{\mathbb{E}}_{\mu,\alpha}\{\sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}|\mathbf{1}_{\Omega_Y(n,\theta,q(M))}\} \leq$$

$$4e^{r(q(M))}\hat{\mathbb{E}}_{\mu,\alpha}\tfrac{1}{M} \sum_{j=1}^M |Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i,N}(\mu), Y) - Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta_i}(\mu), Y)| \leq$$

$$\leq 4e^{r(q(M))} \sup_{\theta\in\Theta} \hat{\mathbb{E}}_{\mu,\alpha}|Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta,N}(\mu), Y) - Z_{(n-q(M))^+,n}(\hat{\Psi}^\theta(\mu), Y)|. \quad (19)$$

We set $g_j(x) = \exp(-\tfrac{1}{2\sigma^2}(Y_j - h(x))^2)$. Then

$$|Z_{(n-q(M))^+,n}(\hat{\Psi}^{\theta,N}(\mu), Y) - Z_{(n-q(M))^+,n}(\hat{\Psi}^\theta(\mu), Y)| =$$

$$= |\prod_{j=(n-q(M))^++1}^n \hat{\Psi}_n^{\theta,N}(\mu)(g_j) - \prod_{j=(n-q(M))^++1}^n \hat{\Psi}_n^\theta(\mu)(g_j)| \leq$$

$$\leq \sum_{j=(n-q(M))^++1}^n |\hat{\Psi}_n^{\theta,N}(\mu)(g_j) - \hat{\Psi}_n^\theta(\mu)(g_j)| \leq q(M)\sup_{n>0} |\hat{\Psi}_n^{\theta,N}(\mu)(g_j) - \hat{\Psi}_n^\theta(\mu)(g_j)|,$$

which follows from the fact that $\|g_j\|_\infty \leq 1$, $\forall j > 0$. Consequently

$$\sup_{n\geq 0} \hat{\mathbb{E}}_{\mu,\alpha} \sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}| \leq 4e^{r(q(M))}q(M)p(N) + 2V(h)\frac{q(M)}{r(q(M))}. \quad (20)$$

The above estimate will converge to zero as $N \uparrow \infty$, for the appropriate choice of $r = r(q)$, $q = q(M)$ and $M = M(N)$. For example, take $r(q) = q^2$, $q(M) = \sqrt{\tfrac{1}{2}\log M}$ and $M(N) = p(N)^{-1}$.

Finally, we have to show that

$$\lim_{M\uparrow\infty} \sup_{n\geq 0} \hat{\mathbb{E}}_{\mu,\alpha}|\sum_{j=1}^M W_{(n-q(M))^+,n}^{j,M}\Psi_n^{\theta_j}(\mu)(f) - \Psi_n^u(\mu)(f)| = 0. \quad (21)$$

But this is exactly the result of theorem 3.1. Thus, the proof is complete.

$\square$

One can also get an estimate for the rate of convergence of (15). However, this will depend on the rate of convergence of (4). Let's say that this is $o(q)$, i.e.

$$\sup_{n\geq q} \mathbb{E}_{\mu,\alpha}\|\tilde{\Psi}_{n-q,n}^u(\mu') - \Psi_n^u(\mu)\|_{tv} \leq o(q),$$

where $\lim_{q\to\infty} o(q) = 0$. Then, one can actually work out the rate of convergence of (8) by similar arguments to these in [5]:

$$\sup_{n\geq 0} \mathbb{E}'_{\mu,\alpha,u}|\sum_{j=1}^M W_{(n-q(M))^+,n}^{j,M}\Psi_n^{\theta_j}(\mu)(f) - \Psi_n^u(\mu)(f)| \leq$$

$$\leq 6\|f\|_\infty \frac{e^{r(q(M))}}{\sqrt{M}} + 2\|f\|_\infty V(h)\frac{q(M)}{r(q(M))} + \|f\|_\infty o(q(M)).$$

The above, together with (17) and (20) give us an estimate for the rate of convergence of the full algorithm, i.e. (15).

We have described an algorithm for the adaptive computation of the optimal filter of a partially observed Markov Chain whose transition kernel depends on an unknown parameter θ. This can be also seen as a uniformly convergent particle filter for an non ergodic system, where the interacting particle filters converge uniformly within every ergodic class.

Acknowledgments

References

[1] A. Budhiraja and H.J. Kushner. Approximation and limit results for nonlinear filters over an infinite time interval. II. Random sampling algorithms. *SIAM J. Control Optim.* 38(6):1874-1908, 2000.

[2] D. Crisan, P. Del Moral and T. Lyons. Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields*, 5(3):293–318, 1999.

[3] C. Andrieu and A. Doucet. Particle Filtering for Partially Observed Gaussian State Space Models. *J. Royal Statist. Soc. B (methodological)*, 64(4):827-836, 2002.

[4] A. Doucet and V.Tadic. Parameter Estimation in General State-Space Models using Particle Methods. *Ann. Inst. Stat. Math.*, 55(2):409-422, 2003.

[5] P. Del Moral. A Uniform Convergence Theorem for Numerical Solving of the Nonlinear Filtering Problem. *Journal of Applied Probability*, 35(4):873-884, 1998.

[6] P. Del Moral and A. Guionnet. On the Stability of Measure Valued Processes with Applications to Filtering. *Comptes Rendus de l'Académie des Science de Paris. Série I. Mathématique*, 329:429-434, 1999.

[7] P. Del Moral and L. Miclo. Branching and Interactive Particle Systems Approximations of Feynman-Kac Formulae with Application to Non-Linear Filtering. In J. Azéma, M. Émery, M. Ledoux, and M. Yor, editors, *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pp.1-145. Springer-Verlag, Berlin, 2000.

[8] N. Gordon, S. Maskell and T. Kirubarajan, Efficient Particle Filters for Joint Tracking and Classification, *Signal and Data Processing of Small Targets 2002*, SPIE vol. 4728, 2002.

[9] N.J. Gordon, D.J. Salmond and A.F.M. Smith. Novel Approach to Nonlinear/non Gaussian Bayesian State Estimation. *IEE Proceedings F*, 140:107-113, 1993.

[10] G. Kitagawa. Monte-Carlo Filter and Smoother for non-Gaussian Nonlinear State Space Models. *Journal on Computational and Graphical Statistics*, 5(1):1-25, 1996.

[11] S.P. Meyn and R.L.Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

[12] F. Le Gland and N. Oudjane. Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters. *Annals of Applied Probability*, 14(1):144-187, 2004.

[13] J. Liu and M. West. Combined Parameter and State Estimation in Simulation-based Filtering. In *Sequential Monte Carlo in Practice*, A. Doucet, N. de Freitas and N. Gordon, Eds., pp.197-223. Springer-Verlag, 2001.

[14] A. Papavasiliou. Parameter Estimation and Asymptotic Stability in Stochastic Filtering. *Submitted*. `http://www.princeton.edu/~ap/stability.pdf`

[15] T. Rydén. On Recursive Estimation for Hidden Markov Models. *Stochastic Processes and their Applications*, 66:79-96, 1997.

[16] G. Storvik. Particle Filters for State Space Models with the Presence of Unknown Static Parameters. *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 281-289, 2002.