

# A UNIFORMLY CONVERGENT ADAPTIVE PARTICLE FILTER

ANASTASIA PAPAVALIOU,\* *Princeton University*

## 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 in which 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.

*Keywords:* Nonlinear filtering; particle filter; Bayesian estimation

2000 Mathematics Subject Classification: Primary 60G35

Secondary 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 to dealing with unknown parameters in the system is to treat them as part of the state variable (see [11] for a historical perspective). In this case, the Bayesian posterior distribution of the parameters is a marginal of the optimal filter. 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 in which 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, it will eventually converge to the true optimal filter as long as the prior distribution satisfies some condition; namely that 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 [12].

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 must resort to numerical methods. One of the most efficient schemes for the recursive computation of the

---

Received 29 March 2004; revision received 9 February 2005.

\* Current address: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.

Email address: a.papavasiliou@warwick.ac.uk

optimal filter is the interacting particle filter (or sequential Monte Carlo method), first suggested (independently) in [8] and [9]. 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 of these particles converges to the optimal filter at every time instant (see, for example, [3], [6], or [2]). Under suitable conditions, the uniform convergence of the particle filter has also been shown (see [5]). However, the uniform convergence results assume some ergodicity conditions on the transition kernel of the Markov chain that – it is clear – will not hold for the augmented system. To the best of the author's knowledge, the only particle filter 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 [4]. 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, [1], [7], or [13]). These algorithms try to avoid this degeneracy by combining the particle filters with other methods of parameter estimation, or by introducing some possibility of the 'rebirth' of a particle. As yet, there is no proof of convergence for any of these filters. Here we study a variation of the particle filter suggested in [7] 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 results of the previous layer of computations, 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 the posterior distribution of the parameters (see [7]).

Note, however, 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 lead us to choose 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 [4] (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 the paper is as follows. We first describe our main assumptions and some previous results, regarding the asymptotic stability of the optimal filter, that we will be using. In

Section 3, we study the case in which the system is linear and Gaussian but 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 filters – 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  $\sigma$ -field. Below we construct 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  $\theta$  from a parameter space  $\Theta$ . More specifically, we want to approximate the optimal filter of the following system.

**Definition 1.** (*System 1.*) Let  $\{X_n\}$  be a homogeneous Markov chain taking values in  $(E, \mathcal{B}(E))$ . Let  $\mu$  be its initial distribution and  $K_\theta$  its transition probability kernel, which depends on a parameter  $\theta \in \Theta$ . Furthermore, we assume that, for each  $\theta \in \Theta$ ,  $K_\theta$  is Feller and mixing, i.e. there exists a constant  $\varepsilon_\theta$ ,  $0 < \varepsilon_\theta \leq 1$ , and a nonnegative measure  $\lambda_\theta \in \mathcal{M}^+(E)$  ( $\mathcal{M}^+(E)$  being the set of finite nonnegative measures on  $E$ ) such that

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

The observation process is defined by  $Y_n = h(X_n) + V_n$ , where  $V_n$  are independent, identically Gaussian( $0, \sigma^2$ )-distributed,  $\mathbb{R}^p$ -valued random variables independent of  $X$ , and  $h : E \rightarrow \mathbb{R}^p$  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  $\Theta$  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.

**Definition 2.** (*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  $K_\theta$  is Feller and mixing in the sense of (1). As usual,  $\delta_{\theta'}(\cdot)$  denotes the Dirac measure on  $\theta'$ . 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  $K_\theta$  and initial distribution  $\mu$  is denoted by  $(\Omega_1 = E^{\mathbb{N}}, (\mathcal{F}_n^{(X)})_{n \geq 0}, P_{\mu, \theta})$ , where  $\mathcal{F}_n^{(X)} = \sigma(X_0, X_1, \dots, X_n)$  is the  $\sigma$ -algebra constructed from the random variables  $X_0, X_1, \dots, X_n$ . 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  $\mathcal{F}_n^{(Y)} = \sigma(Y_1, \dots, Y_n)$ . The law of the observation process,  $Q_{\mu, \theta}$ , is given by

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

for any  $n \geq 0$  and  $k_1, \dots, k_n \in \mathbb{R}_+$ , where  $E^{\otimes n} = E \times \dots \times E$  is the product space of  $n$  copies of  $E$ . By  $Q_{\mu, \theta}^n$  we denote the restriction of  $Q_{\mu, \theta}$  to the  $\sigma$ -algebra  $\mathcal{F}_n^{(Y)}$ . Then we can define the pair process  $(X, Y)$  on the space

$$(\Omega = \Omega_1 \times \Omega_2, (\mathcal{F}_n = \mathcal{F}_n^{(X)} \times \mathcal{F}_n^{(Y)})_{n \geq 0}, P_{\mu, \theta}),$$

where the measure  $P_{\mu, \theta}$  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, [4]). We will denote the expectation with respect to  $P_{\mu, \theta}$  by  $E_{\mu, \theta}$ .

Similarly, we define the triplet  $(X, Y, \theta)$  on the space

$$(\tilde{\Omega} = \Omega \times \Omega_3, (\tilde{\mathcal{F}}_n = \mathcal{F}_n \times \sigma(\theta))_{n \geq 0}, \tilde{P}_{\mu, u}),$$

where  $\theta$  is a  $\Theta$ -valued random variable defined on  $(\Omega_3, \sigma(\theta), u)$  and  $\Omega_3$  is some set that is homeomorphic to  $\Theta$ . The marginals of  $\tilde{P}_{\mu, u}$  on  $(X, Y)$  and  $\theta$  are  $\int_{\Theta} P_{\mu, \theta} u(d\theta)$  and  $u$ , respectively. We will denote the expectation with respect to  $\tilde{P}_{\mu, u}$  by  $\tilde{E}_{\mu, u}$ .

Furthermore, we denote by  $\Psi_n^\theta(\mu)$  and  $\Phi_n(\mu \otimes u)$  the optimal filters for systems 1 and 2, with respective initial distributions  $\mu$  and  $\mu \otimes u$  defined as the posterior distributions of the state variable, given the observations. The name ‘optimal filter’ is due to the fact that they are the best estimators adapted to the available information (the  $\sigma$ -algebra constructed by the observations), with respect to the  $L_2$ -norm. They are random measures on the spaces  $E$  and  $E \times \Theta$ , respectively, defined as follows. For every  $f \in \mathcal{C}_b(E \times \Theta)$ , i.e. every real-valued continuous bounded function defined on  $E \times \Theta$ ,

$$\begin{aligned} \Phi_n(\mu \otimes u)(f) &= \tilde{E}_{\mu, u}[f(X_n, \theta) \mid Y_n, \dots, Y_1] \\ &= \frac{\int_{\Theta} \int_{E^{\otimes n}} f(x_n, \theta) \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu, \theta}(dx_1, \dots, dx_n) u(d\theta)}{\int_{\Theta} \int_{E^{\otimes n}} \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu, \theta}(dx_1, \dots, dx_n) u(d\theta)}. \end{aligned}$$

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

$$\begin{aligned} \Psi_n^\theta(\mu)(f') &= E_{\mu, \theta}[f'(X_n) \mid Y_n, \dots, Y_1] \\ &= \frac{\int_{E^{\otimes n}} f'(x_n) \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu, \theta}(dx_1, \dots, dx_n)}{\int_{E^{\otimes n}} \prod_{k=1}^n g(y_k - h(x_k)) P_{\mu, \theta}(dx_1, \dots, dx_n)}. \end{aligned}$$

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  $f_\alpha(x) = f(x, \alpha)$ ). Similarly, we define  $\Psi_n^\mu(\mu)$  as the marginal of  $\Phi_n(\mu \otimes u)$  with respect to  $X$ , i.e.

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

The mixing condition (1) implies ergodicity of the signal. We denote by  $\mu_\theta$  the limiting distribution of the Markov chain whose transition kernel is  $K_\theta$ . The measure  $\mu_\theta$  is uniquely defined in this way as a result of the ergodicity 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 [12]):

$$\lim_{n \rightarrow \infty} E_{\mu, \alpha}[\|\Psi_n^\mu(\mu') - \Psi_n^\alpha(\mu)\|_{TV}] = 0.$$

Here,  $\mu$  and  $\mu'$  are arbitrary initial distributions and, as usual, we denote by  $\|\cdot\|_{TV}$  the total variation norm.

In this paper, we need the slightly stronger result

$$\lim_{q \rightarrow \infty} \sup_{n \geq q} E_{\mu, \alpha}[\|\tilde{\Psi}_{n-q, n}^\mu(\mu') - \Psi_n^\alpha(\mu)\|_{TV}] = 0, \tag{2}$$

where

$$\tilde{\Psi}_{n-q, n}^\mu(\mu')(dx) = \int_{\Theta} \Psi_n^\theta(\mu')(dx) P_{\Psi_{n-q}^\theta(\mu'), u}(d\theta \mid Y_n, \dots, Y_{n-q+1}).$$

This definition is consistent with the way we approximate the posterior distribution of the parameter  $\theta$  in (5) and (9), below.

**Lemma 1.** *Let  $Y$  be as in system 1. We make the following assumptions.*

- *The prior distribution  $u$  on the parameter space is such that there exist a sequence  $\varepsilon_n \downarrow 0$  and a function  $p: \mathbb{N} \rightarrow [1, \infty)$  with the following properties:*

$$\frac{p(n)}{n} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

$$\sup_{n \uparrow \infty} E_{\mu, \alpha} \left[ \left( \frac{\sup_{\theta \in \mathcal{N}_{\varepsilon_n}(\alpha)} dQ_{\mu, \alpha}^n(Y_n, \dots, Y_1) / dQ_{\mu', \theta}^n}{u(\mathcal{N}_{\varepsilon_n}(\alpha))} \right)^{1/p(n)} \right] < \infty.$$

- *For each  $\eta > 0$ , there exist an  $\varepsilon > 0$  and an  $I_\eta > 0$  such that*

$$\limsup_{n \rightarrow \infty} \sup_{k \geq 0} \frac{1}{n} \log \int_{\mathcal{N}_\eta(\alpha)^c} E_{\mu, \alpha} [P_{\mu'_{k, \theta}}(L_n(Y) \in \mathcal{B}_\varepsilon(\nu_\alpha))] u(d\theta) \leq -I_\eta < 0, \tag{3}$$

where  $L_n(Y)$  is the empirical measure of the observation process up to time  $n$ , i.e.  $L_n(Y) = (1/n) \sum_{k=1}^n \delta_{Y_k}$ , and  $\mu'_{k, \theta} = P_{\mu', \theta}(X_0 \in dx \mid Y_0, \dots, Y_{-k+1})$ .

- *The optimal filters are uniformly continuous, i.e.*

$$\lim_{\theta \rightarrow \alpha} \sup_{n > 0} E_{\mu', \alpha}[\|\Psi_n^\theta(\mu) - \Psi_n^\alpha(\mu)\|_{TV}] = 0. \tag{4}$$

Then

$$\lim_{q \rightarrow \infty} \sup_{n \geq q} E_{\mu, \alpha} [\|\tilde{\Psi}_{n-q, n}^u(\mu') - \Psi_n^\alpha(\mu)\|_{TV}] = 0$$

for any initial distributions  $\mu, \mu' \in \mathcal{P}(E)$ , where  $\mathcal{P}(E)$  is the set of probability measures on  $E$ . Note that we have used the notation  $\mathcal{N}_\rho(\alpha)$  to denote the ball of radius  $\rho$  and center  $\alpha$  with respect to the metric  $d_\Theta$ , for any  $\rho > 0$ . Similarly, by  $\mathcal{B}_\varepsilon(v_\alpha)$  we denote the ball of radius  $\varepsilon$  and center  $v_\alpha$  with respect to the Lévy–Prohorov metric.

The proof is exactly the same as in [12], except that we now need the large deviation estimate (3) to hold uniformly. In [12], there is a lengthy discussion on how to verify assumptions (3) and (4) using only the properties of the kernel. The assumption on the prior quantifies how much mass the prior should have on the true parameter, so that it is possible to recover it. In the case of a finite parameter space, for example, it is sufficient that the prior assigns a positive probability to all the parameters (i.e.  $\varepsilon_n \equiv 0$  and  $p(n) \equiv 1$ ).

Below we construct a particle filter that approximates  $\Psi_n^u(\mu)$  uniformly with respect to time. With this, it is an obvious corollary of Lemma 1 that the particle filter will converge to the true one as time  $n$  and the number of particles  $N$  go to infinity, i.e.  $(n, N) \rightarrow (\infty, \infty)$ .

### 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

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

where, for every  $\theta \in \Theta$ , the coefficients  $F(\theta)$ ,  $c(\theta)$ ,  $G(\theta)$ ,  $H$ , and  $d$  are deterministic and belong to the Euclidean spaces  $\mathbb{R}^{p \times p}$ ,  $\mathbb{R}^p$ ,  $\mathbb{R}^{p \times p}$ ,  $\mathbb{R}^{q \times p}$ , and  $\mathbb{R}^q$ , 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, independent, identically Gaussian-distributed sequences, with covariance  $\text{cov}(W_n) = Q$  and  $\text{cov}(V_n) = \Sigma$ , independent of each other and of the initial distribution  $\mathcal{L}(X_0)$  of  $X_0$ . Moreover, we assume that the initial distribution of  $X_0$  is also Gaussian, with mean  $E(X_0) = \bar{X}_0$  and covariance  $Q_0$ .

Then, for a fixed  $\theta$ , the optimal filter of this system is the Kalman–Bucy filter, which can be computed exactly. However, once we introduce the parameter to the system (as in system 2), 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 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  $\theta$ , the optimal filter can be computed exactly.

#### 3.1. The algorithm

The following algorithm is based on the Monte Carlo particle filter presented in [4], where the role of the state variable is played by the optimal filters.

*Time  $n = 0$  (initialization).* At time  $n = 0$ , the particle system consists of  $M$  independent particles  $\{\theta_j\}_{j=1, \dots, M}$  with common law  $u$ . The weights  $W_{0,0}^{j,M}$  of the parameters are set to be equal to  $1/M$ , i.e.  $W_{0,0}^{j,M} = 1/M$ . 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  $\mu$  is the initial distribution of the Markov chain (which is Gaussian in the case that  $\Psi_n^\theta$  is the Kalman–Bucy filter).

Time  $n \geq 0$  (evolution). At time  $n \geq 0$ , we compute the new weights  $W_{(n-q(M))^+,n}^{j,M}$ , for a nondecreasing function  $q$  such that  $\lim_{M \uparrow \infty} q(M) = \infty$ , as follows:

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

Here,

$$\hat{\Psi}_n^\theta(\mu)(dx) := P_{\mu,\theta}(X_n \in dx \mid y_{n-1}, \dots, y_1)$$

is the optimal one-step predictor for a fixed  $\theta$  (the Kalman–Bucy predictor, in the case of the linear Gaussian system) and, for each  $j = 1, \dots, M$ ,

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

We use the notation  $a^+ = \max\{a, 0\}$  for all  $a \in \mathbb{R}$ . Also,  $\sigma$  is the standard deviation of the Gaussian distribution (note that, for the sake of simplicity, we set the dimension of the observation variables,  $p$ , equal to 1). 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 0 very quickly since it is a product of numbers less than 1. Note that multiplying  $g$  by  $\sqrt{2\pi} \sigma$  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 0 (see also [4]).

In practice, things are worse: numbers below some threshold are treated as equal to 0 in computer arithmetic and, consequently, only a series of ‘unlikely’ observations with respect to some  $\theta$  can set the corresponding weight equal to 0. This weight will of course remain 0 for ever, so we are back to the problem of ‘losing’ particles. To remedy this, we can set the weights to be  $\max\{W_{(n-q(M))^+,n}^{j,M}, \varepsilon\}$ , for some very small  $\varepsilon$ .

In order to account for the additional randomness in the choice of  $\bar{\theta}^M = (\theta_1, \dots, \theta_M)$ , we define a new probability space

$$(\tilde{\Omega}' = \tilde{\Omega} \times \Omega_3^M, (\tilde{\mathcal{F}}'_n = \tilde{\mathcal{F}}_n \times \sigma(\theta_1, \dots, \theta_M))_{n \geq 0}, \tilde{P}'_{\mu,u} = \tilde{P}_{\mu,u} \otimes u^M)$$

and the corresponding expectation  $\tilde{E}'_{\mu,u}$ . Similarly, we define the probability space

$$(\Omega' = \Omega \times \Omega_3^M, (\mathcal{F}'_n = \mathcal{F}_n \times \sigma(\theta_1, \dots, \theta_M))_{n \geq 0}, P'_{\mu,\theta,u} = P_{\mu,\theta} \otimes u^M)$$

and the corresponding expectation  $E'_{\mu,\theta,u}$ .

The following theorem is an extension of the main theorem of [4].

**Theorem 1.** *Suppose that (2) holds and that*

$$\begin{aligned} \sup_n E_{\mu,\alpha}[h(X_n)]^2 &< \infty, \\ \sup_{n,\theta} E_{\mu,\alpha}[E_{\mu,\theta}[h(X_n) \mid Y_n, \dots, Y_1]^2] &< \infty. \end{aligned}$$

(This holds trivially in the case that  $h$  is bounded, but we include this condition to also cover linear Gaussian systems with unknown parameters.) Then there exists a nondecreasing function  $q: \mathbb{N} \rightarrow \mathbb{N}$  (for example  $q(M) = \frac{1}{2}(\log M)^{1/2}$ ) for which the error made by the particle filter  $\Psi_n^{\mu,M}(\mu)$  given by (7) will converge to 0 uniformly with respect to time, i.e.

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

*Proof.* The theorem can be proved by following the arguments of the proof of uniform convergence for the Monte Carlo particle filter (see [4]).

#### 4. The full algorithm

In [5], Del Moral and Guionnet demonstrated the uniform convergence of the interacting particle filter, but under some strong ergodicity conditions that are not satisfied by our system 2. In [4], Del Moral proved 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, requires a dense sampling of the path space.

Below we present 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  $K_\theta$  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 1, that the Monte Carlo particle filter for 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’, i.e. those that are close to the correct value of the parameter. Instead, the Monte Carlo particle filter computes the weighted average. So, no particles are lost and the weight depends on several past observations rather than on just the last one.

Below we give a detailed description of the algorithm (by  $q$  we denote a nondecreasing function such that  $\lim_{N \uparrow \infty} q(N) = \infty$ ).



Time  $n = 0$  (initialization). At time  $n = 0$ , the particle system consists of  $M$  independent particles  $\{\theta_j\}_{j=1,\dots,M}$  with common law  $u$ , and  $N * M$  independent particles

$$\{\xi_i^0(\theta_j)\}_{i=1,\dots,N, j=1,\dots,M}$$

with common law  $\mu$ . The weights of the parameters are set to be equal, i.e.  $W_{0,0}^{j,M,N} = 1/M$ . 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  $\theta_j$ , i.e.

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

Time  $n \geq 0$  (evolution). 1. Each particle  $\xi_i^n(\theta_j)$ ,  $i = 1, \dots, N$  evolves according to the kernel  $K_{\theta_j}$ , for  $j = 1, \dots, M$ , i.e.  $\hat{\xi}_i^n(\theta_j) \sim K_{\theta_j}(\xi_i^n(\theta_j), \cdot)$ . We set

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

Note that the particles  $(\hat{\xi}_i^n(\theta_j))_{i,j}$  are such that  $\hat{\Psi}_n^{\theta_j,N}$  approximates the one-step predictor  $P_{\mu,\theta_j}(X_{n+1} | Y_n, \dots, Y_1)$  for every  $j$ .

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

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

Here

$$\begin{aligned} Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta_j,N}(\mu))_{\ell=1}^n, Y) &:= \prod_{k=(n-q(M))^++1}^n \int_E g(Y_k - h(z_k)) \hat{\Psi}_k^{\theta_j,N}(\mu)(dz_k) \\ &= \prod_{k=(n-q(M))^++1}^n \left( \frac{1}{N} \sum_{i=1}^N g(Y_k - h(\hat{\xi}_i^k(\theta_j))) \right) \end{aligned} \tag{10}$$

for each  $j = 1, \dots, M$ .

3. For each  $j = 1, \dots, M$ , we resample the particles  $\{\hat{\xi}_i^n(\theta_j)\}_{i=1,\dots,N}$  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 obtain  $\{\xi_i^{n+1}(\theta_j)\}_{i=1,\dots,N}$ .

The particle filter is then defined by

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

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

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

Once again, we must consider a new probability space that will also capture the additional randomness in the evolution of the particle filters. We denote by  $(\Omega_\xi, \mathcal{F}_n^\xi, P_{\mu,u}^\xi)$  the canonical space for the particle system

$$\{(\xi_n^{i,N}(\theta_j), \theta_j)\}_{i=1,\dots,N, j=1,\dots,M, n>0},$$

which is assumed to be independent of the state observation pair  $(X, Y)$ . Then the new probability space will be

$$(\hat{\Omega} = \Omega \times \Omega_\xi, (\hat{\mathcal{F}}_n = \mathcal{F}_n \times \mathcal{F}_n^\xi)_{n \geq 0}, \hat{P}_{\mu,\theta,u} = P_{\mu,\theta} \otimes P_{\mu,u}^\xi).$$

The expectation with respect to  $\hat{P}_{\mu,\theta,u}$  will be denoted by  $\hat{E}_{\mu,\theta,u}$ . To simplify our notation, we will often drop the  $u$  when its meaning is clear.

### 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 [5] or [6]. We alter the statement, however, to better fit the present context.

**Lemma 2.** *Suppose that*

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

for all  $\theta \in \Theta$  and all  $f \in \mathcal{C}_b(E)$ , where  $\Psi_{n-q,n}^\theta(\mu) = P_{\mu,\theta}(X_n | Y_n, \dots, Y_{n-q+1})$ . Then, for all  $f \in \mathcal{C}_b(E)$ ,

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

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

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

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

then there exists an  $N_{0,\theta} > 0$  such that

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

for all  $N \geq N_{0,\theta}$ , where  $C_\theta$  is some constant,  $\gamma' = 1 - \log(2\pi\sigma^2)$ , and  $a_\theta = \gamma_\theta / (\gamma_\theta + \gamma) > 0$ .

We will prove the following result under the general assumption that the particle filter approximating the optimal filter for a 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 [10]) if, in addition, the parameter space is compact and the optimal filters uniformly continuous with respect to the parameter.

**Theorem 2.** *Suppose that (2) 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. that it satisfies*

$$\sup_{n \geq 0} \hat{E}_{\mu, \alpha} [\|\Psi_n^{\theta, N}(\mu) - \Psi_n^\theta(\mu)\|_{TV}] \leq p(N) \quad \text{for all } \theta \in \Theta, \tag{11}$$

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

$$\lim_{N \uparrow \infty} \sup_{n \geq 0} \hat{E}_{\mu, \alpha} [\|\Psi_n^{\mu, M(N), N}(\mu)(f) - \Psi_n^\mu(\mu)(f)\|] = 0. \tag{12}$$

*Proof.* For simplicity of notation, we assume that  $p = 1$ , i.e. the observations take values in  $\mathbb{R}$ . We first show that

$$\lim_{N \uparrow \infty} \sup_{M > 0} \sup_{n \geq 0} \hat{E}_{\mu, \alpha} \left[ \left\| \Psi_n^{\mu, M, N}(\mu)(f) - \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M, N} \Psi_n^{\theta_j}(\mu)(f) \right\| \right] = 0 \tag{13}$$

for all  $f \in \mathcal{C}_b(E)$ . We note that

$$\begin{aligned} & \hat{E}_{\mu, \alpha} \left[ \left\| \Psi_n^{\mu, M, N}(\mu)(f) - \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M, N} \Psi_n^{\theta_j}(\mu)(f) \right\| \right] \\ &= \hat{E}_{\mu, \alpha} \left[ \left\| \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M, N} \Psi_n^{\theta_j, N}(\mu)(f) - \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M, N} \Psi_n^{\theta_j}(\mu)(f) \right\| \right] \\ &\leq \hat{E}_{\mu, \alpha} \left[ \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M, N} |\Psi_n^{\theta_j, N}(\mu)(f) - \Psi_n^{\theta_j}(\mu)(f)| \right] \\ &\leq \sup_j \hat{E}_{\mu, \alpha} [\|\Psi_n^{\theta_j, N}(\mu)(f) - \Psi_n^{\theta_j}(\mu)(f)\|] \\ &\leq \|f\|_\infty p(N), \end{aligned} \tag{14}$$

where  $\|\cdot\|_\infty$  denotes the  $L_\infty$ -norm.

Thus, (13) follows from (11). Note that, in order to simplify our notation, we have suppressed the dependence of  $M$  on  $N$ ; it will only be shown when relevant. We next show that

$$\lim_{N \uparrow \infty} \sup_{n \geq 0} \hat{E}_{\mu, \alpha} \left[ \left\| \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M, N} \Psi_n^{\theta_j}(\mu)(f) - \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M} \Psi_n^{\theta_j}(\mu)(f) \right\| \right] = 0,$$

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

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

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

$$\begin{aligned} \hat{E}_{\mu,\alpha} &\left[ \left| \sum_{j=1}^M W_{(n-q(M))^+,n}^{j,M,N} \Psi_n^{\theta_j}(\mu)(f) - \sum_{j=1}^M W_{(n-q(M))^+,n}^{j,M} \Psi_n^{\theta_j}(\mu)(f) \right| \right] \\ &\leq \hat{E}_{\mu,\alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}| |\Psi_n^{\theta_j}(\mu)(f)| \right] \\ &\leq \|f\|_\infty \hat{E}_{\mu,\alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}| \right]. \end{aligned}$$

Thus, it is sufficient to show that

$$\lim_{N \uparrow \infty} \sup_{n \geq 0} \hat{E}_{\mu,\alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}| \right] = 0.$$

We follow the approach used in [4] to bound the denominator away from 0. We define the following quantity for all  $n > 0$ , where the expectation  $\bar{E}_{\mu,\theta}$  is with respect to the measure  $\hat{\Psi}^\theta(\mu) = \otimes_{k=0}^\infty \hat{\Psi}_k^\theta(\mu)$ :

$$\begin{aligned} V_n^{q(M)}(Y, \theta) &= \frac{1}{2\sigma^2} \sum_{k=(n-q(M))^+ + 1}^n \bar{E}_{\mu,\theta}[(Y_k - h(X_k))^2] \\ &= \frac{1}{2\sigma^2} \sum_{k=(n-q(M))^+ + 1}^n ((Y_k - \bar{E}_{\mu,\theta}[h(X_k)])^2 + \bar{E}_{\mu,\theta}[h(X_k) - \bar{E}_{\mu,\theta}[h(X_k)]]^2). \end{aligned}$$

By the boundedness of  $h$ , we obtain

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

for all  $\theta \in \Theta$ . Then

$$\begin{aligned} E_{\mu,\alpha}[V_n^{q(M)}(Y, \theta)] &\leq \frac{1}{2\sigma^2} q(M) \left[ \sup_k E_{\mu,\alpha}[V_k + h(X_k) - \bar{E}_{\mu,\theta}[h(X_k)]]^2 + 4\|h\|_\infty^2 \right] \\ &\leq \frac{1}{2\sigma^2} q(M) (8\|h\|_\infty^2 + 4\sigma\|h\|_\infty + \sigma^2) \\ &= V(h)q(M) \\ &< \infty, \end{aligned}$$

where we have used the notation

$$V(h) = \frac{8\|h\|_\infty^2 + 4\sigma\|h\|_\infty + \sigma^2}{2\sigma^2}.$$

Now consider the set

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

for some nondecreasing function  $r : \mathbb{N} \rightarrow \mathbb{R}^+$  to be identified later. It follows that

$$P_{\mu, \alpha}(\Omega_2 - \Omega_Y(n, \theta, q(M))) \leq V(h) \frac{q(M)}{r(q(M))}$$

(see [4]) and, consequently,

$$\begin{aligned} & \hat{E}_{\mu, \alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+, n}^{j, M, N} - W_{(n-q(M))^+, n}^{j, M}| (\mathbf{1}_{\Omega_Y(n, \theta, q(M))} + \mathbf{1}_{\Omega_2 - \Omega_Y(n, \theta, q(M))}) \right] \\ & \leq \hat{E}_{\mu, \alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+, n}^{j, M, N} - W_{(n-q(M))^+, n}^{j, M}| \mathbf{1}_{\Omega_Y(n, \theta, q(M))} \right] + 2V(h) \frac{q(M)}{r(q(M))}, \end{aligned}$$

where  $\mathbf{1}_A$  denotes the indicator function of the set  $A$ . The second term on the right-hand side will vanish as  $M \uparrow \infty$ , for any choice of  $r$  satisfying  $\lim_{q \uparrow \infty} q/r(q) = 0$ . It remains to show that

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

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

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

Then we observe that

$$\begin{aligned} & \hat{E}_{\mu, \alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+, n}^{j, M, N} - W_{(n-q(M))^+, n}^{j, M}| U_n^{M, N} \mathbf{1}_{\Omega_Y(n, \theta, q(M))} \right] \\ & = \hat{E}_{\mu, \alpha} \left[ \frac{\sum_{j=1}^M |Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_j, N}(\mu))_{\ell=1}^n, Y) - Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_j}(\mu))_{\ell=1}^n, Y) U_n^{M, N}|}{\sum_{k=1}^M Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_k}(\mu))_{\ell=1}^n, Y)} \right. \\ & \quad \left. \times \mathbf{1}_{\Omega_Y(n, \theta, q(M))} \right] \\ & \leq \hat{E}_{\mu, \alpha} \left[ \frac{\sum_{j=1}^M |Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_j, N}(\mu))_{\ell=1}^n, Y) - Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_j}(\mu))_{\ell=1}^n, Y)|}{\sum_{k=1}^M Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_k}(\mu))_{\ell=1}^n, Y)} \right. \\ & \quad \left. \times \mathbf{1}_{\Omega_Y(n, \theta, q(M))} \right] \\ & \quad + \hat{E}_{\mu, \alpha} [ |1 - U_n^{M, N}| \mathbf{1}_{\Omega_Y(n, \theta, q(M))} ] \\ & \leq 2 \hat{E}_{\mu, \alpha} \left[ \frac{\sum_{j=1}^M |Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_j, N}(\mu))_{\ell=1}^n, Y) - Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_j}(\mu))_{\ell=1}^n, Y)|}{\sum_{k=1}^M Z_{(n-q(M))^+, n}((\hat{\Psi}_\ell^{\theta_k}(\mu))_{\ell=1}^n, Y)} \right. \\ & \quad \left. \times \mathbf{1}_{\Omega_Y(n, \theta, q(M))} \right]. \end{aligned}$$

Similarly,

$$\begin{aligned} & \hat{E}_{\mu,\alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}| (1 - U_n^{M,N}) \mathbf{1}_{\Omega_Y(n,\theta,q(M))} \right] \\ & \leq 2 \hat{E}_{\mu,\alpha} \left[ \frac{\sum_{j=1}^M |Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta_i,N}(\mu))_{\ell=1}^n, Y) - Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta_i}(\mu))_{\ell=1}^n, Y)|}{\sum_{k=1}^M Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta_k}(\mu))_{\ell=1}^n, Y)} \right. \\ & \quad \left. \times \mathbf{1}_{\Omega_Y(n,\theta,q(M))} \right]. \end{aligned}$$

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

$$\begin{aligned} & \frac{1}{M} \sum_{k=1}^M Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta_k}(\mu))_{\ell=1}^n, Y(\omega)) \\ & = \frac{1}{M} \sum_{k=1}^M \bar{E}_{\theta_k} \left[ \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=(n-q(M))^++1}^n (Y_j(\omega) - h(X_j))^2 \right\} \right] \\ & \geq \frac{1}{M} \sum_{k=1}^M \exp \{ -V_n^{q(M)}(Y(\omega), \theta_k) \} \\ & \geq \exp \left\{ -\frac{1}{M} \sum_{k=1}^M V_n^{q(M)}(Y(\omega), \theta_k) \right\} \\ & \geq \exp \{ -r(q(M)) \}, \end{aligned}$$

by applying Jensen’s inequality twice. By putting everything together, we find that

$$\begin{aligned} & \hat{E}_{\mu,\alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+,n}^{j,M,N} - W_{(n-q(M))^+,n}^{j,M}| \mathbf{1}_{\Omega_Y(n,\theta,q(M))} \right] \\ & \leq \frac{4e^{r(q(M))}}{M} \hat{E}_{\mu,\alpha} \left[ \sum_{j=1}^M |Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta_i,N}(\mu))_{\ell=1}^n, Y) - Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta_i}(\mu))_{\ell=1}^n, Y)| \right] \\ & \leq 4e^{r(q(M))} \sup_{\theta \in \Theta} \hat{E}_{\mu,\alpha} [|Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta,N}(\mu))_{\ell=1}^n, Y) - Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta}(\mu))_{\ell=1}^n, Y)|]. \end{aligned}$$

If we set  $g_j(x) = \exp\{-(Y_j - h(x))^2/(2\sigma^2)\}$ , then

$$\begin{aligned} & |Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta,N}(\mu))_{\ell=1}^n, Y) - Z_{(n-q(M))^+,n}((\hat{\Psi}_\ell^{\theta}(\mu))_{\ell=1}^n, Y)| \\ & = \left| \prod_{j=(n-q(M))^++1}^n \hat{\Psi}_n^{\theta,N}(\mu)(g_j) - \prod_{j=(n-q(M))^++1}^n \hat{\Psi}_n^{\theta}(\mu)(g_j) \right| \\ & \leq \sum_{j=(n-q(M))^++1}^n |\hat{\Psi}_n^{\theta,N}(\mu)(g_j) - \hat{\Psi}_n^{\theta}(\mu)(g_j)| \\ & \leq q(M) \sup_{n>0} |\hat{\Psi}_n^{\theta,N}(\mu)(g_j) - \hat{\Psi}_n^{\theta}(\mu)(g_j)|, \end{aligned}$$

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

$$\sup_{n \geq 0} \hat{E}_{\mu, \alpha} \left[ \sum_{j=1}^M |W_{(n-q(M))^+, n}^{j, M} - W_{(n-q(M))^+, n}^{j, M}| \right] \leq 4e^{r(q(M))} q(M) p(N) + 2V(h) \frac{q(M)}{r(q(M))}. \tag{15}$$

The above estimate will converge to 0 as  $N \uparrow \infty$ , for the appropriate choice of  $r = r(q)$ ,  $q = q(M)$ , and  $M = M(N)$  (for example, take  $r(q) = q^2$ ,  $q(M) = (\frac{1}{2} \log M)^{1/2}$ , and  $M(N) = p(N)^{-1}$ ).

Finally, we must show that

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

However, this is exactly the result of Theorem 1. Thus, the proof is complete.

We can also obtain an estimate for the rate of convergence of (12). However, this will depend on the rate of convergence of (2). Let us say that this is  $o(q)$ , i.e.

$$\sup_{n \geq q} E_{\mu, \alpha} [\|\tilde{\Psi}_{n-q, n}^u(\mu') - \Psi_n^u(\mu)\|_{TV}] \leq o(q),$$

where  $\lim_{q \rightarrow \infty} o(q) = 0$ . Then we can work out the rate of convergence of (8) by arguments similar to those in [4]:

$$\begin{aligned} \sup_{n \geq 0} E'_{\mu, \alpha, u} \left[ \left| \sum_{j=1}^M W_{(n-q(M))^+, n}^{j, M} \Psi_n^{\theta_j}(\mu)(f) - \Psi_n^u(\mu)(f) \right| \right] \\ \leq 6\|f\|_\infty \frac{e^{r(q(M))}}{\sqrt{M}} + 2\|f\|_\infty V(h) \frac{q(M)}{r(q(M))} + \|f\|_\infty o(q(M)). \end{aligned}$$

The above, together with (14) and (15), gives us an estimate for the rate of convergence of (12), i.e. of the full algorithm.

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  $\theta$ . This can also be seen as a uniformly convergent particle filter for a nonergodic system in which the interacting particle filters converge uniformly within every ergodic class.

### Acknowledgements

I would like to thank René Carmona and Pierre Del Moral for stimulating discussions on the subject considered here.

### References

- [1] ANDRIEU, C. AND DOUCET, A. (2002). Particle filtering for partially observed Gaussian state space models. *J. R. Statist. Soc. B* **64**, 827–836.
- [2] BUDHIRAJA, A. AND KUSHNER, H. J. (2000). Approximation and limit results for nonlinear filters over an infinite time interval. II. Random sampling algorithms. *SIAM J. Control Optimization* **38**, 1874–1908.
- [3] CRISAN, D., DEL MORAL, P. AND LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Proc. Relat. Fields* **5**, 293–318.

- [4] DEL MORAL, P. (1998). A uniform convergence theorem for numerical solving of the nonlinear filtering problem. *J. Appl. Prob.* **35**, 873–884.
- [5] DEL MORAL, P. AND GUIONNET, A. (1999). On the stability of measure valued processes with applications to filtering. *C. R. Acad. Sci. Paris* **329**, 429–434.
- [6] DEL MORAL, P. AND MICLO, L. (2000). Branching and interactive particle systems approximations of Feynman–Kac formulae with application to non-linear filtering. In *Séminaire de Probabilités XXXIV* (Lecture Notes Maths. **1729**), eds J. Azéma, M. Émery, M. Ledoux and M. Yor, Springer, Berlin, pp. 1–145.
- [7] GORDON, N., MASKELL, S. AND KIRUBARAJAN, T. (2002). Efficient particle filters for joint tracking and classification. In *Proc. SPIE Signal Data Process. Small Targets*, SPIE, Vol. 4728, pp. 439–449.
- [8] GORDON, N. J., SALMOND, D. J. AND SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F* **140**, 107–113.
- [9] KITAGAWA, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graphical Statist.* **5**, 1–25.
- [10] LE GLAND, F. AND OUDJANE, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters. *Ann. Appl. Prob.* **14**, 144–187.
- [11] LIU, J. AND WEST, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, eds A. Doucet, N. de Freitas and N. Gordon, Springer, New York, pp. 197–223.
- [12] PAPAVALIIOU, A. (2005). Parameter estimation and asymptotic stability in stochastic filtering. To appear in *Stoch. Process. Appl.*
- [13] STORVIK, G. (2002). Particle filters for state space models with the presence of unknown static parameters. *IEEE Trans. Signal Process.* **50**, 281–289.