

Stability and Approximation of Nonlinear Filters in the Hilbert Metric, and Application to Particle Filters*

François LeGland
IRISA / INRIA
Campus de Beaulieu
35042 RENNES Cédex, France
e-mail : leglan@irisa.fr

Nadia Oudjane
ONERA, DTIM / MCT
29 avenue de la division Leclerc
92320 CHATILLON, France
e-mail : oudjane@onera.fr

Abstract

In this paper, the stability of the optimal filter w.r.t. its initial condition and w.r.t. the model, is studied in a general HMM using the Hilbert projective metric. These stability results are then used to prove the uniform convergence of the interacting particle filter to the optimal filter, as the number of particles goes to infinity.

Keywords : hidden Markov model, nonlinear filter, stability, Hilbert metric, total variation distance, particle filter.

1 Introduction

The stability of the optimal filter has been the subject of many works. Ocone and Pardoux proved in [10], that the filter forgets its initial condition in the L^p sense without stating any rate of convergence. Recently, a new approach using the projective Hilbert metric has been proposed by Da Prato, Fuhman and Malliavin in [3]. Some results of stability w.r.t. the initial condition are proved by this way in Atar and Zeitouni [1]. Independently, Del Moral and Guionnet proposed in [4] another approach based on semi-group techniques and on the Dobrushin ergodic coefficient to derive some stability results w.r.t. the initial condition, which are used therein to prove the uniform convergence of the interacting particle filter to the optimal filter with a rate $(1/\sqrt{N})^\alpha$, for any $\alpha < 1$. In this article, the approach using the Hilbert metric is used to study the asymptotic behavior of the optimal filter and as in [4] uniform convergence results are derived for the interacting particle filter.

In the next section, the nonlinear filtering problem is defined, and some notations are introduced. In Section 3, some properties of the Hilbert metric are stated,

*This work was partially supported by the CNRS programme *Modélisation et Simulation Numérique*.

which are used in Section 4 to prove the stability of the optimal filter w.r.t. its initial condition : the total variation error at time n is related to the initial error in the Hilbert metric or in total variation. Then stability of the optimal filter w.r.t. the model is proved : the error in the weak sense at time n is related to the local error (committed at each time step) in the weak sense. In Section 5, these stability results are used to prove the uniform convergence in the weak sense for the interacting particle filter, with a rate $1/\sqrt{N}$.

2 Optimal filter for general HMM

Consider the following model, with two sequences $\{X_n, n \geq 0\}$ and $\{Y_n, n \geq 0\}$, taking values in \mathbb{R}^m and \mathbb{R}^d respectively:

- The hidden state sequence $\{X_n, n \geq 0\}$ is an inhomogeneous Markov chain, with transition probability kernel Q_n , i.e.

$$\mathbb{P}[X_n \in dx' \mid X_{n-1} = x] = Q_n(x, dx'),$$

for all $n \geq 1$, and with initial probability distribution μ_0 .

- The observation sequence $\{Y_n, n \geq 0\}$ is related to the state sequence $\{X_n, n \geq 0\}$ by

$$Y_n = h_n(X_n, V_n),$$

for all $n \geq 0$, where $\{V_n, n \geq 0\}$ is a sequence of independent random variables, not necessarily Gaussian, independent of the state sequence $\{X_n, n \geq 0\}$. It is assumed that the collection of probability distributions $\mathbb{P}[Y_n \in dy \mid X_n = x]$ on \mathbb{R}^d , parametrized by $x \in \mathbb{R}^m$, is dominated, i.e.

$$\mathbb{P}[Y_n \in dy \mid X_n = x] = g_n(x, y) \lambda_n(dy),$$

for some nonnegative measure λ_n on \mathbb{R}^d . The corresponding likelihood function is defined by $\Psi_n(x) = g_n(x, Y_n)$, and depends implicitly on the observation Y_n .

The following notations and definitions are used throughout the paper

- The set of probability distributions on \mathbb{R}^m , and the set of nonnegative measures on \mathbb{R}^m , are denoted by \mathcal{M}^+ and \mathcal{P} respectively.
- With any nonnegative kernel K defined on \mathbb{R}^m , is associated a nonnegative linear operator acting on \mathcal{M}^+ , still denoted by K , and defined by

$$K \mu(dx') = \int_{\mathbb{R}^m} \mu(dx) K(x, dx'),$$

for any $\mu \in \mathcal{M}^+$. Consequently, the adjoint nonnegative linear operator acting on functions, and denoted by K^* , is defined by

$$K^* \phi(x) = \int_{\mathbb{R}^m} K(x, dx') \phi(x'),$$

for any measurable function ϕ defined on \mathbb{R}^m .

- With any nonzero $\mu \in \mathcal{M}^+$, i.e. such that $\langle \mu, 1 \rangle \neq 0$, is associated the normalized nonnegative measure (i.e. the probability distribution) $\bar{\mu} = \mu / \langle \mu, 1 \rangle \in \mathcal{P}$.
- With any nonnegative kernel K defined on \mathbb{R}^m , is associated the normalized nonnegative nonlinear operator \bar{K} on \mathcal{M}^+ and taking values in \mathcal{P} , defined by $\bar{K}(\mu) = K \mu / \langle K \mu, 1 \rangle = K \bar{\mu} / \langle K \bar{\mu}, 1 \rangle = \bar{K}(\bar{\mu})$, for any $\mu \in \mathcal{M}^+$ such that $\langle K \mu, 1 \rangle \neq 0$, and $\bar{K}(\mu) = 0$ otherwise.

The problem of nonlinear filtering is to compute at each time n , the conditional probability distribution μ_n of the state X_n given the observation sequence $Y_{0:n} = (Y_0, \dots, Y_n)$ up to time n . The transition from μ_{n-1} to μ_n is described by the following two steps :

$$\mu_{n-1} \xrightarrow{\text{prediction}} \mu_{n|n-1} = Q_n \mu_{n-1}$$

$$\xrightarrow{\text{correction}} \mu_n = \Psi_n \cdot \mu_{n|n-1},$$

where \cdot denotes the projective product. In the correction step, μ_n is given by the Bayes rule

$$\mu_n = \Psi_n \cdot \mu_{n|n-1} = \frac{\Psi_n \mu_{n|n-1}}{\langle \mu_{n|n-1}, \Psi_n \rangle}.$$

For any $n \geq 0$, introduce the nonnegative kernel

$$R_n(x, dx') = Q_n(x, dx') \Psi_n(x'),$$

and the associated nonnegative linear operator R_n acting on \mathcal{M}^+ , defined by

$$R_n \mu(dx') = \int_{\mathbb{R}^m} \mu(dx) Q_n(x, dx') \Psi_n(x'),$$

for any $\mu \in \mathcal{M}^+$. Notice that R_n depends on the observation Y_n through the likelihood function Ψ_n , hence is random. With this definition, the evolution of the optimal filter can be written as follows

$$\mu_n = \frac{R_n \mu_{n-1}}{\langle R_n \mu_{n-1}, 1 \rangle} = \bar{R}_n(\mu_{n-1}) = \Psi_n \cdot (Q_n \mu_{n-1}),$$

and iteration yields

$$\begin{aligned} \mu_n &= \bar{R}_n(\mu_{n-1}) = \bar{R}_n \circ \dots \circ \bar{R}_m(\mu_{m-1}) \\ &= \bar{R}_{n:m}(\mu_{m-1}). \end{aligned}$$

This equation shows clearly that the evolution of the optimal filter is nonlinear only because of the normalization term coming from the Bayes rule. In the following section a projective metric is introduced precisely to get rid of the normalization and to come down to the analysis of a linear evolution.

Throughout the paper, the notation $\|\cdot\|$ is used for the total variation norm on the set of finite signed measures on \mathbb{R}^m , and for the supremum norm on the set of bounded measurable functions on \mathbb{R}^m , depending on the context.

3 Hilbert metric on the set of finite nonnegative measures

In this section, the Hilbert metric is introduced, and some of its properties are stated, which are used below.

Definition 3.1 *Two nonnegative measures $\mu, \mu' \in \mathcal{M}^+$ are comparable, if there exist positive constants $0 < a \leq b$, such that*

$$a \mu'(dx) \leq \mu(dx) \leq b \mu'(dx).$$

Notice that the two nonnegative measures μ and μ' are comparable if and only if μ and μ' are equivalent with Radon–Nikodym derivatives $\frac{d\mu}{d\mu'}$ and $\frac{d\mu'}{d\mu}$ bounded and bounded away from zero.

Definition 3.2 (Hilbert metric) *The Hilbert metric on \mathcal{M}^+ is defined by*

$$\begin{aligned} h(\mu, \mu') &= \log \frac{\sup_{A: \mu'(A) > 0} \frac{\mu(A)}{\mu'(A)}}{\inf_{A: \mu'(A) > 0} \frac{\mu(A)}{\mu'(A)}} \\ &= \log \left(\left\| \frac{d\mu}{d\mu'} \right\| \left\| \frac{d\mu'}{d\mu} \right\| \right), \end{aligned}$$

if $\mu, \mu' \in \mathcal{M}^+$ are nonzero and comparable, and $h(\mu, \mu') = +\infty$ otherwise.

The Hilbert metric h is a projective distance, i.e. invariant under multiplication by positive scalars, hence $h(\mu, \mu') = h(\bar{\mu}, \bar{\mu}')$, for any $\mu, \mu' \in \mathcal{M}^+$. In the nonlinear filtering context, this property allows to consider the linear transformation $\mu \mapsto R_n \mu$ instead of the nonlinear transformation $\mu \mapsto \bar{R}_n(\mu) = R_n \mu / \langle R_n \mu, 1 \rangle$.

Definition 3.3 (Mixing property) *The nonnegative kernel K defined on \mathbb{R}^m is mixing, if there exist a constant $0 < \varepsilon \leq 1$, and a nonnegative measure $\lambda \in \mathcal{M}^+$, such that*

$$\varepsilon \lambda(dx') \leq K(x, dx') \leq \frac{1}{\varepsilon} \lambda(dx') ,$$

for any $x \in \mathbb{R}^m$.

The following lemma relates the total variation norm and the Hilbert metric.

Lemma 3.4 *For any nonzero $\mu, \mu' \in \mathcal{M}^+$*

$$\|\bar{\mu} - \bar{\mu}'\| \leq \frac{2}{\log 3} h(\mu, \mu') . \quad (1)$$

If in addition the nonnegative kernel K defined on \mathbb{R}^m is mixing, then

$$h(K \mu, K \mu') \leq \frac{1}{\varepsilon^2} \|\bar{\mu} - \bar{\mu}'\| . \quad (2)$$

The first inequality is proved in Atar and Zeitouni [1], and the second inequality follows from the estimate $\log r \leq |r - 1|$, which holds for any $r > 0$, and from the Scheffe theorem.

Theorem 3.5 (Birkhoff [2], Hopf [8]) *Let K be the nonnegative linear operator on \mathcal{M}^+ associated with the nonnegative kernel K defined on \mathbb{R}^m . The contraction coefficient*

$$\tau(K) = \sup_{0 < h(\mu, \mu') < \infty} \frac{h(K \mu, K \mu')}{h(\mu, \mu')} , \quad (3)$$

associated with the Hilbert metric, is called the Birkhoff contraction coefficients, and satisfies

$$\tau(K) = \tanh\left[\frac{1}{4} H(K)\right] ,$$

where $H(K)$ is the following diameter

$$H(K) = \sup_{\mu, \mu' \in \mathcal{M}^+} h(K \mu, K \mu') .$$

Notice that $H(K) < \infty$ implies that $\tau(K) < 1$.

The stability results stated below require in general that for any $n \geq 0$, the nonnegative kernel R_n is mixing, i.e.

there exist a constant $0 < \varepsilon_n \leq 1$, and a nonnegative measure $\lambda_n \in \mathcal{M}^+$, such that

$$\varepsilon_n \lambda_n(dx') \leq R_n(x, dx') \leq \frac{1}{\varepsilon_n} \lambda_n(dx') ,$$

for any $x \in \mathbb{R}^m$. Notice that in full generality ε_n and λ_n depend on the observation Y_n , hence are random variables.

Lemma 3.6 *The nonnegative linear operator R_n defined on \mathcal{M}^+ is a contraction under the Hilbert metric, with Birkhoff contraction coefficient $\tau_n = \tau(R_n) \leq 1$. Moreover*

(i) *If R_n is mixing, with the possibly random constant ε_n , then*

$$\tau_n \leq \frac{1 - \varepsilon_n^2}{1 + \varepsilon_n^2} < 1 .$$

(ii) *If Q_n is mixing, with the nonrandom constant ε_n , then R_n is also mixing, with the same constant ε_n , and*

$$\tau_n \leq \tau(Q_n) \leq \frac{1 - \varepsilon_n^2}{1 + \varepsilon_n^2} < 1 .$$

Throughout the paper, for any integers $m \leq n$, the contraction coefficient of the product $R_{n:m} = R_n \cdots R_m$ is denoted by $\tau_{n:m} = \tau(R_{n:m}) \leq \tau_n \cdots \tau_m$ and by convention $\tau_{n:n+1} = \tau_{m-1:m} = 1$.

4 Stability of nonlinear filters

In practice, the initial distribution of the hidden state is often unknown. Hence from a practical point of view, the stability of the filter w.r.t. its initial condition is a desirable property. Moreover, this property is useful to prove the stability w.r.t. the model.

Consider the filter μ_n correctly initialized with μ_0 , and the filter μ'_n wrongly initialized with μ'_0 , i.e. $\mu_n = \bar{R}_{n:1}(\mu_0)$ and $\mu'_n = \bar{R}_{n:1}(\mu'_0)$. An estimate of the total variation error at time n induced by the initial error, is given by the following.

Theorem 4.1 *Without any assumption on the nonnegative kernels, the following inequality holds*

$$\|\mu_n - \mu'_n\| \leq \frac{2}{\log 3} \tau_{n:m} h(\mu_{m-1}, \mu'_{m-1}) .$$

If in addition the nonnegative kernel R_m is mixing, then

$$\|\mu_n - \mu'_n\| \leq \frac{2}{\log 3} \tau_{n:m+1} \frac{1}{\varepsilon_m^2} \|\mu_{m-1} - \mu'_{m-1}\| .$$

Corollary 4.2 *If for any $k \geq 0$, the nonnegative kernel R_k is mixing with $\varepsilon_k \geq \varepsilon > 0$, then*

$$\|\mu_n - \mu'_n\| \leq \frac{2}{\varepsilon^2 \log 3} \tau^{n-m} \|\mu_{m-1} - \mu'_{m-1}\| ,$$

$$\text{with } \tau = \frac{1 - \varepsilon^2}{1 + \varepsilon^2} .$$

PROOF. Using (1) and the definition (3) of the Birkhoff contraction coefficient, yields

$$\begin{aligned} \|\bar{R}_{n:m}(\mu) - \bar{R}_{n:m}(\mu')\| &\leq \frac{2}{\log 3} h(R_{n:m} \mu, R_{n:m} \mu') \\ &\leq \frac{2}{\log 3} \tau_{n:m} h(\mu, \mu') , \end{aligned}$$

for any $\mu, \mu' \in \mathcal{P}$. If the nonnegative kernel R_m is mixing, then using (2) yields

$$\begin{aligned} &\|\bar{R}_{n:m}(\mu) - \bar{R}_{n:m}(\mu')\| \\ &\leq \frac{2}{\log 3} h(R_{n:m+1} R_m \mu, R_{n:m+1} R_m \mu') \\ &\leq \frac{2}{\log 3} \tau_{n:m+1} h(R_m \mu, R_m \mu') \\ &\leq \frac{2}{\log 3} \tau_{n:m+1} \frac{1}{\varepsilon_m^2} \|\mu - \mu'\| . \end{aligned} \quad (4)$$

Taking $\mu = \mu_m$ and $\mu' = \mu'_m$ finishes the proof. \square

This stability result w.r.t. the initial condition is used below to prove the stability of the filter w.r.t. the model. In practice, the model is often unknown, and instead of using the true model, it is common to use a wrong model, based on a wrong transition kernel Q'_n and a wrong likelihood function Ψ'_n , which define the evolution operator R'_n for a wrong filter μ'_n . Another situation is when the evolution operator R_n is known, but difficult to compute. For the purpose of practical implementation, an approximate filter μ'_n is designed such that the evolution $\mu'_{n-1} \mapsto \mu'_n$ is easy to compute and close to the true evolution $\mu'_{n-1} \mapsto \bar{R}_n(\mu'_{n-1})$.

The purpose of the following results is to bound the *global error* between μ'_n and μ_n induced by the *local errors* committed at each time step. Without loss of generality, it is assumed here that $\mu_0 = \mu'_0$, since the problem of a wrong initialization has been studied above. In full generality, it is assumed that $\{\mu'_n, n \geq 0\}$ is a random sequence with values in \mathcal{P} , satisfying the following property : for any $n \geq k \geq 0$ and for any bounded measurable function F defined on \mathcal{P}

$$\mathbb{E}[F(\mu'_k) \mid Y_{0:n}] = \mathbb{E}[F(\mu'_k) \mid Y_{0:k}] . \quad (5)$$

The results stated below are based on the following decomposition of the global error into sums of local errors

transported by a sequence of normalized evolution operators \bar{R}_n ,

$$\begin{aligned} \mu'_n - \mu_n &= \sum_{k=1}^n [\bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k}(\mu'_{k-1})] \\ &= \sum_{k=1}^n [\bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})] . \end{aligned} \quad (6)$$

This equation shows the close relation between the stability w.r.t. the initial condition and the stability w.r.t. the model.

Assumption W :

$$\delta_k = \sup_{\phi : \|\phi\|=1} \mathbb{E}[|\langle \mu'_k - \bar{R}_k(\mu'_{k-1}), \phi \rangle| \mid Y_{0:k}] < \infty .$$

If the approximation μ'_k is an empirical measure associated with $\bar{R}_k(\mu'_{k-1})$, then bounding the local error requires to use the law of large numbers which can only provide a bound in the weak sense of Assumption W. However, the following lemma shows that the error transported by one evolution operator can still be bounded in total variation, if the operator is smooth enough.

Lemma 4.3 *If the nonnegative kernel K defined on \mathbb{R}^m is dominated, i.e. if there exist a constant $c > 0$, and a nonnegative measure $\lambda \in \mathcal{M}^+$, such that*

$$K(x, dx') \leq \frac{1}{c} \lambda(dx') ,$$

for any $x \in \mathbb{R}^m$, then

$$\mathbb{E} \|K \mu - K \mu'\| \leq \frac{\langle \lambda, 1 \rangle}{c} \sup_{\phi : \|\phi\|=1} \mathbb{E} |\langle \mu - \mu', \phi \rangle| , \quad (7)$$

for any $\mu, \mu' \in \mathcal{M}^+$, possibly random.

The following estimates are useful to bound the error between two normalized measures in terms of the error between the two corresponding unnormalized measures. For any $\mu, \mu' \in \mathcal{M}^+$

$$|\langle \bar{\mu} - \bar{\mu}', \phi \rangle| \leq \frac{|\langle \mu - \mu', \phi \rangle|}{\langle \mu, 1 \rangle} + \frac{|\langle \mu - \mu', 1 \rangle|}{\langle \mu, 1 \rangle} \|\phi\| , \quad (8)$$

and

$$\|\bar{\mu} - \bar{\mu}'\| \leq \frac{\|\mu - \mu'\|}{\langle \mu, 1 \rangle} + \frac{|\langle \mu - \mu', 1 \rangle|}{\langle \mu, 1 \rangle} . \quad (9)$$

Theorem 4.4 *If for any $k \geq 0$, Assumption W holds, and the nonnegative operator R_k is mixing, then*

$$\begin{aligned} &\sup_{\phi : \|\phi\|=1} \mathbb{E}[|\langle \mu_n - \mu'_n, \phi \rangle| \mid Y_{0:n}] \\ &\leq \delta_n + 2 \frac{\delta_{n-1}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} \tau_{n:k+3} \frac{\delta_k}{\varepsilon_{k+2}^2 \varepsilon_{k+1}^2} . \end{aligned}$$

Corollary 4.5 *If for any $k \geq 0$, Assumption W holds with $\delta_k \leq \delta$, and the nonnegative operator R_k is mixing with $\varepsilon_k \geq \varepsilon > 0$, then*

$$\begin{aligned} & \sup_{\phi: \|\phi\|=1} \mathbb{E}[|\langle \mu_n - \mu'_n, \phi \rangle| \mid Y_{0:n}] \\ & \leq \left(1 + \frac{2}{\varepsilon^2} + \frac{4}{\varepsilon^6 \log 3}\right) \delta. \end{aligned}$$

PROOF. Using the decomposition (6) and the triangle inequality, yields

$$\begin{aligned} & |\langle \mu'_n - \mu_n, \phi \rangle| \leq |\langle \mu'_n - \bar{R}_n(\mu'_{n-1}), \phi \rangle| \\ & + \sum_{k=1}^{n-1} \|\bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\| \|\phi\|. \end{aligned}$$

For any $1 \leq k \leq n-2$, the estimate (4) yields

$$\begin{aligned} & \|\bar{R}_{n:k+1}(\mu'_k) - \bar{R}_{n:k+1} \circ \bar{R}_k(\mu'_{k-1})\| \\ & = \|\bar{R}_{n:k+2} \circ \bar{R}_{k+1}(\mu'_k) - \bar{R}_{n:k+2} \circ \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1})\| \\ & \leq \frac{2}{\log 3} \tau_{n:k+3} \frac{1}{\varepsilon_{k+2}^2} \|\bar{R}_{k+1}(\mu'_k) - \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1})\|. \end{aligned}$$

For any $1 \leq k \leq n-1$, taking conditional expectation w.r.t. the observations, and using estimate (9), yields

$$\begin{aligned} & \mathbb{E}[\|\bar{R}_{k+1}(\mu'_k) - \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1})\| \mid Y_{0:n}] \\ & \leq 2 \mathbb{E}\left[\frac{\|R_{k+1}(\mu'_k - \bar{R}_k(\mu'_{k-1}))\|}{\langle R_{k+1} \mu'_k, 1 \rangle} \mid Y_{0:n}\right]. \end{aligned}$$

The mixing property yields

$$\langle R_{k+1} \mu'_k, 1 \rangle \geq \varepsilon_{k+1} \langle \lambda_{k+1}, 1 \rangle,$$

and the estimate (7) yields

$$\begin{aligned} & \mathbb{E}[\|R_{k+1}(\mu'_k - \bar{R}_k(\mu'_{k-1}))\| \mid Y_{0:n}] \\ & \leq \frac{\langle \lambda_{k+1}, 1 \rangle}{\varepsilon_{k+1}} \sup_{\phi: \|\phi\|=1} \mathbb{E}[|\langle \mu'_k - \bar{R}_k(\mu'_{k-1}), \phi \rangle| \mid Y_{0:n}] \\ & \leq \frac{\langle \lambda_{k+1}, 1 \rangle}{\varepsilon_{k+1}} \delta_k. \end{aligned}$$

Combining these estimates yields

$$\mathbb{E}[\|\bar{R}_{k+1}(\mu'_k) - \bar{R}_{k+1} \circ \bar{R}_k(\mu'_{k-1})\| \mid Y_{0:n}] \leq 2 \frac{\delta_k}{\varepsilon_{k+1}^2}.$$

Finally, taking conditional expectation w.r.t. the observations, yields

$$\begin{aligned} & \sup_{\phi: \|\phi\|=1} \mathbb{E}[|\langle \mu'_n - \mu_n, \phi \rangle| \mid Y_{0:n}] \\ & \leq \delta_n + 2 \frac{\delta_{n-1}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} \tau_{n:k+3} \frac{\delta_k}{\varepsilon_{k+2}^2 \varepsilon_{k+1}^2}. \quad \square \end{aligned}$$

5 Uniform convergence of interacting particle filters

In this section the wrong model is chosen deliberately, such that the wrong filter can easily be computed, and remains close to the optimal filter. Here, particle methods are considered to approximate the optimal filter. Roughly speaking, particle methods are essentially based on the Monte Carlo principle, which allows to approximate a probability distribution μ when a sample (ξ^1, \dots, ξ^N) of i.i.d. random variables with the probability distribution μ is given. Throughout the paper, $S^N(\mu)$ is a shorthand notation for the empirical distribution of an N -sample with probability distribution μ

$$S^N(\mu) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i}.$$

Lemma 5.1 *For any $\mu \in \mathcal{P}$*

$$\sup_{\phi: \|\phi\|=1} \mathbb{E}[|\langle S^N(\mu) - \mu, \phi \rangle|] \leq \frac{1}{\sqrt{N}}.$$

The method is very easy to implement, even in high dimensional problems, since it is sufficient in principle to simulate independent sample paths of the hidden dynamical system. A major and earliest contribution in this field was made by Gordon, Salmond and Smith [7], which proposed to use sampling / importance resampling (SIR) techniques in the correction step : the positive effect of the resampling step is to automatically concentrate particles in regions of interest of the state space. A very complete account of the currently available mathematical results can be found in the survey paper by Del Moral and Miclo [5]. Theoretical and practical aspects can be found in the volume edited by Doucet, de Freitas and Gordon [6].

The transition from μ_{n-1}^N to μ_n^N is described by the following two steps

$$\begin{aligned} \mu_{n-1}^N & \xrightarrow[\text{prediction}]{\text{sampling}} \mu_{n|n-1}^N = S^N(Q_n \mu_{n-1}^N) \\ & \xrightarrow{\text{correction}} \mu_n^N = \Psi_n \cdot \mu_{n|n-1}^N. \end{aligned}$$

Without loss of generality, it is assumed that the likelihood function is bounded, and the following assumption is introduced.

Assumption L :

$$\rho_k = \frac{\sup_{x \in \mathbb{R}^m} \Psi_k(x)}{\inf_{\mu \in \mathcal{P}} \langle Q_k \mu, \Psi_k \rangle} < \infty.$$

Theorem 5.2 *If for any $k \geq 0$, Assumption L holds, and the nonnegative operator R_k is mixing, then the particle filter μ_n^N satisfies*

$$\begin{aligned} & \sup_{\phi : \|\phi\|=1} \mathbb{E}[|\langle \mu_n - \mu_n^N, \phi \rangle| \mid Y_{0:n}] \\ & \leq \delta_n + 2 \frac{\delta_{n-1}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} \tau_{n:k+3} \frac{\delta_k}{\varepsilon_{k+2}^2 \varepsilon_{k+1}^2}, \end{aligned}$$

where for any $k \geq 0$

$$\delta_k \leq \frac{1}{\sqrt{N}} 2 \rho_k .$$

Corollary 5.3 *If for any $k \geq 0$, the nonnegative operator R_k is mixing with $\varepsilon_k \geq \varepsilon > 0$, and $\rho_k \leq \rho$ a.s., then the convergence is uniform in time, with rate $1/\sqrt{N}$, i.e.*

$$\begin{aligned} & \sup_{\phi : \|\phi\|=1} \mathbb{E}[|\langle \mu_n - \mu_n^N, \phi \rangle| \mid Y_{0:n}] \\ & \leq \frac{1}{\sqrt{N}} \left(1 + \frac{2}{\varepsilon^2} + \frac{4}{\varepsilon^6 \log 3} \right) 2 \rho . \end{aligned}$$

PROOF. It is sufficient here to check that Assumption W is satisfied, and to apply Theorem 4.4. Using estimate (8) yields

$$\begin{aligned} & |\langle \mu_k^N - \bar{R}_k(\mu_{k-1}^N), \phi \rangle| \\ & = |\langle \Psi_k \cdot (S^N(Q_k \mu_{k-1}^N)) - \Psi_k \cdot (Q_k \mu_{k-1}^N), \phi \rangle| \\ & \leq \frac{|\langle S^N(Q_k \mu_{k-1}^N) - Q_k \mu_{k-1}^N, \Psi_k \phi \rangle|}{\langle Q_k \mu_{k-1}^N, \Psi_k \rangle} \\ & \quad + \frac{|\langle S^N(Q_k \mu_{k-1}^N) - Q_k \mu_{k-1}^N, \Psi_k \rangle|}{\langle Q_k \mu_{k-1}^N, \Psi_k \rangle} \|\phi\| , \end{aligned}$$

for any bounded measurable test function ϕ defined on \mathbb{R}^m . By definition

$$\langle Q_k \mu_{k-1}^N, \Psi_k \rangle \geq \inf_{\mu \in \mathcal{P}} \langle Q_k \mu, \Psi_k \rangle .$$

Using Lemma 5.1 with $\mu = Q_k \mu_{k-1}^N$, and $\Psi_k \phi$ instead of ϕ , yields

$$\begin{aligned} & \mathbb{E}[|\langle S^N(Q_k \mu_{k-1}^N) - Q_k \mu_{k-1}^N, \Psi_k \phi \rangle| \mid Y_{0:k}, \mu_{k-1}^N] \\ & \leq \frac{1}{\sqrt{N}} \|\phi\| \sup_{x \in \mathbb{R}^m} \Psi_k(x) , \end{aligned}$$

hence

$$\sup_{\phi : \|\phi\|=1} \mathbb{E}[|\langle \mu_k^N - \bar{R}_k(\mu_{k-1}^N), \phi \rangle| \mid Y_{0:k}] \leq \frac{1}{\sqrt{N}} 2 \rho_k .$$

which concludes the proof. \square

Using the same techniques, it is possible to prove the stability of the optimal filter w.r.t. the model in some stronger sense : for instance, the error in total variation at time n can be related to the local error in total variation. This allows to prove the uniform convergence in total variation for the regularized particle filters introduced in Musso, Oudjane and LeGland [9].

References

- [1] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 33(6):697–725, 1997.
- [2] G. Birkhoff. *Lattice Theory*, volume XXV of *Colloquium Publications*. AMS, Providence, RI, 3rd edition, 1967.
- [3] G. Da Prato, M. Fuhrman, and P. Malliavin. Asymptotic ergodicity of the process of conditional law in some problem of nonlinear filtering. *Journal of Functional Analysis*, 164(2):356–377, 1999.
- [4] P. Del Moral and A. Guionnet. On the stability of measure valued processes. Applications to nonlinear filtering and interacting particle systems. Publication du Laboratoire de Statistique et Probabilités 3–98, Université Paul Sabatier, Toulouse, 1998.
- [5] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering. In J. Azéma, M. Émery, M. Ledoux, and M. Yor, editors, *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pages 1–145. Springer–Verlag, Berlin, 2000.
- [6] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer–Verlag, New York, 2001.
- [7] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear / non–Gaussian Bayesian state estimation. *IEE Proceedings, Part F*, 140(2):107–113, Apr. 1993.
- [8] E. Hopf. An inequality for positive integral linear operators. *Journal of Mathematics & Mechanics*, 12(5):683–692, 1963.
- [9] C. Musso, N. Oudjane, and F. Le Gland. Improving regularized particle filters. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*. Springer–Verlag, New York, 2001.
- [10] D. L. Ocone and E. Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM Journal on Control and Optimization*, 34(1):226–243, Jan. 1996.