Nonlinear filtering in discrete time : A particle convolution approach

Vivien Rossi and Jean-Pierre Vila (`{rossiv,vila}@ensam.inra.fr`)
UMR Analyse des Systmes et de Biomtrie, INRA-ENSAM, 2 Place P. Viala, 34060 Montpellier France

Abstract. In this paper a new generation of particle filters for nonlinear discrete time processes is proposed, based on convolution kernel probability density estimation. The main advantage of this approach is to be free of the limitations encountered by the current particle filters when the likelihood of the observation variable is analytically unknown or when the observation noise is null or too small. To illustrate this convolution kernel approach the counterparts of the well-known sequential importance sampling (SIS) and sequential importance sampling-resampling (SIS-R) filters are considered and their stochastic convergence to the optimal filter under different modes are proved. Their good behaviour with respect to that of these filters is shown on several simulated case studies.

1. Introduction

Nonlinear filtering, *i.e* estimating the distribution of a dynamical system state variables x_t conditional on the observations of state-dependent variables y_t up to time t, is a real problem in many fields of engineering. Since the development of the Kalman filter in 1960 for the linear case, a wide variety of approximate methods were proposed to deal with the nonlinear one. The most famous and the most widely used by the engineers is the extended Kalman filter (EKF). For more details about the EKF see (Jazwinski, 1970). But the EKF lacks of theorical support and often presents troubles in practice. Some improvements, a few of which are presented in (Chen, 1993), were proposed as generalised Kalman filters but without overcoming these drawbacks. Investigation in the field of nonlinear filtering has thus remained very active.

Among the alternatives to the EKF, the Monte Carlo approaches offer practical and theorical results. The corresponding state of the art is reviewed by (Liu & Chen, 1998) and (Doucet, 1998). The Monte Carlo filters are divided into two families : the first one also the oldest, includes the filters based on the sequential importance sampling (SIS)

algorithm and the second the filters based on the sequential importance sampling resampling (SIS-R) algorithm. The principle of Monte Carlo filter based on SIS was developed in the early 70's, by (Handschin, 1970) and by (Akashi & al., 1975) but due to the limitations of the processors of the time remained dormant until the 80's when it was revived by (Davis, 1981) and (Kitagawa, 1987). However, the SIS filters suffered from divergence in long time. The new generation of Monte Carlo SIS-R filters improved on this issue but did not completely overcome it. The idea of introducing a resampling step in the SIS algorithm was independently proposed by (Gordon & al., 1993) with "the boostrap filter", (Del Moral & al., 1992), (Del Moral, 1995) with "the Interacting Particle Filter" (IPF) still one of the most performant filters, and (Kitagawa, 1996). A lot of filters based on SIS-R were then developed. A complete review of these works is presented in (Doucet & al., 2001). The convergence of the IPF to the optimal filter is proved by (Del Moral, 1998), (Del Moral & al., 2001) or (LeGland & Oudjane, 2004). However, in spite of their theorical properties, the filters based on SIS-R still present several drawbacks in practice. Systems with non-noisy observations are not supported by these filters because the density of the noise is used to weight the particles. Even too small observation noise can induce divergence of the filters. Moreover, the problem of divergence in long time has not completely disappeared. According to (Hrzeler & Knsch, 1998) it is caused by the discrete nature of the distribution approximations produced by the SIS-R filters. A step of regularization on the state variable distribution was then added into the SIS-R algorithm successively by (Hrzeler & Knsch, 1998), (Oudjane, 2000), (Warnes, 2001) and (Musso & al., 2001), in order to produce a probability density as approximation of the optimal filter. Actually, the convolution kernels are used to stabilize signal to noise, this idea has been introduced and analyzed in some details in (Del Moral & Miclo, 2000) and in (LeGland & Oudjane, 2003). Practically, the addition of this regularization step improves the behavior of the filters and the theorical properties are kept. See (Oudjane, 2000) or (LeGland & Oudjane, 2004) for the convergence of the resulting Regularized Interacted Particle Filter (RIPF) using convolution kernel regularization.

However these improvements do not relieve definitively from practical difficulties in case of too large (and even too small) signal-to-noise ratio. Moreover, these regularized filters still rely on the analytical availability of the observation likelihood, a classic Monte-Carlo filter assumption not frequently met in real situations. Only (Del Moral & al., 2001), (Del Moral & Jacod, 2001) considered a context in which this likelihood is not accessible and used a regularization of the observation

distribution .

The approach we propose is based on convolution kernel density estimation and implicit regularization of both state and observation variable distributions, and is free of their analytical knowledge. Only the capability of simulating the state and observation noises is required. The problem of null or small observation noise is also overcome. The theoretical properties of these new particle filters rely entirely upon the kernel probability density estimation theory and not upon that of the current particle filter theory, even if from a certain point of view these filters can be interpreted as generalization of that of (Del Moral & al., 2001).

The paper is organized as follows. The next section is devoted to recalling the Monte carlo filtering context. A transition from the particle filters to the convolution filters is presented in Section 3. The algorithmic and theoretical properties of the basic Convolution Filter (CF) are presented in Section 4 and that of the Resampled-Convolution Filter (R-CF) in Section 5. In Section 6 the behaviour of our convolution filters are compared with that of their counterpart interacting particle filters and Monte Carlo filter, on simulated case studies in different noise situations. Finally, results of kernel density estimation theory useful to the study of the convergence properties of our convolution filters are gathered in Appendix A and the proofs of these properties themselves are presented in Appendix B.

2. The filtering context

Consider a general discrete dynamical system

$$\begin{cases} x_t = f_t(x_{t-1}, \varepsilon_t) \\ y_t = h_t(x_t, \eta_t) \end{cases} \tag{1}$$

in which $x_t \in \mathbb{R}^d$ and $y_t \in \mathbb{R}^q$ are the unobserved state variable and observed variable respectively. ε_t and η_t are the independent state noise and observation noise respectively. f_t and h_t are two known Borel measurable functions, possibly time-varying.

The filtering problem is to estimate the distribution π_t of x_t conditional on the past observations $y_1, \ldots, y_t$, the so called optimal filter. When the density function of π_t exists it will be noted $p(x_t | y_1, \ldots, y_t)$ or $p(x_t | y_{1:t})$ or simply p_t. Of course the filtering problem only exists when the functions h_t are not bijective. All the filters are thus developed to deal with such a context, and generally the functions h_t are not even assumed injective.

9

To build up Monte Carlo filters, the following hypotheses are generally assumed:

- the distribution, π_0, of the initial state variable x_0, is known.

- the distributions of the noises ε_t and η_t, are known for all $t \in \mathbb{N}$.

The density of η_t plays a crucial role for Monte Carlo filters since it is used to weight the generated particles. Let us illustrate this point by recalling briefly the Interacting Particle Filter (IPF) algorithm, one of our filters of reference in this paper:

- For $t = 0$
  generation of n particles : $(\tilde{x}_0^1, \cdots, \tilde{x}_0^n)$ $i.i.d.$ $\sim \pi_0$
  Estimation of π_0 : $\pi_0^n = \sum_{i=1}^n \omega_0^i \delta_{\tilde{x}_0^i}$ (where $\omega_0^i = \frac{1}{n}$; $i = 1, \cdots, n$).

- For $t \geq 1$
  (i)   Sampling step : $(\tilde{x}_{t-1}^1, \cdots, \tilde{x}_{t-1}^n) \sim \pi_{t-1}^n$.
  (ii)   Evolving step : $\tilde{x}_{t|t-1}^i \sim f_t(\tilde{x}_{t-1}^i, .)$
  (iii)   Weighting step : $\omega_t^i = \Psi_t(\tilde{x}_{t|t-1}^i)/(\sum_{i=1}^n \Psi_t(\tilde{x}_{t|t-1}^i))$
  (iv)   Approximation step : $\pi_t^n = \sum_{i=1}^n \omega_t^i \delta_{\tilde{x}_{t|t-1}^i}$,

  with $\delta_{\tilde{x}_{t|t-1}^i}$ Dirac measure.

The weight ω_t^i of each particle is given by the likelihood function $\Psi_t(x) = p(y_t|x)$. The observations y_t thus operate the filter through the likelihood function which is assumed to exist and to be known. This assumption is rather restricting in practice. Moreover let us note that it rules out the non-noisy case and can also cause trouble when the noise η_t is too small and also when the noise is non-additive as in the general system (1).

These severe drawbacks are circumvented in the approach proposed in this paper which as will be seen shortly, uses convolution kernels to weight the generated particles, with furthermore interesting theoretical and practical benefits.

As a transition let us see beforehand a possible kernel-based improvement of the particle filters to deal with the unknown likelihood case.

3. From the Particle Filter to the Convolution Filter

We shall illustrate this transition through the well-known IPF particle filter of (Del Moral, 1995) but it could have been performed with any filter built up from the SIS or SIS-R principle.

Suppose that a limited knowledge of the distribution of η_t prevents access to the likelihood function $\Psi_t(x) = p(y_t|x)$ and only permits to simulate observation y_t from a given x_t through model (1).

Let $K_h(.)$ be a Parzen-Rosenblatt kernel (see appendix A). In order to clarify the presentation, we just recall that, for all the paper, a kernel K is a bounded, positive, symmetrical application from $\mathbb{R}^d \to \mathbb{R}$, such that $\int K d\lambda = 1$, where λ is the Lebesgue measure. Using kernel estimation theory we can then consistently approximate Ψ_t, by simulating observations, and use this approximation in place of the true function in the previous IPF algorithm. This amounts to replace steps (iii) and (iv) by

(iii)'   Weighting step

> For (i=1:n)
>> Stage 1
>>> Generation of N observations:
>>> $$\tilde{y}_t^1, \ldots, \tilde{y}_t^N \sim h_t(\tilde{x}_{t|t-1}^i, .)$$
>> Stage 2
>>> Approximation of Ψ_t:
>>> $$\hat{p}^N(y_t|\tilde{x}_{t|t-1}^i) = \frac{1}{Nh^q}\sum_{j=1}^{N} K_h(y_t - \tilde{y}_t^j)$$
>>> Approximation of the weight of $\tilde{x}_{t|t-1}^i$:
>>> $$\omega_t^i = \hat{p}^N(y_t|\tilde{x}_{t|t-1}^i)$$
>
> Normalization of the weights $\hat{\omega}_t^i = \omega_t^i/(\sum_{i=1}^{n} \omega_t^i)$

(iv)   Approximation step $\hat{\pi}_t^n = \sum_{i=1}^{n} \hat{\omega}_t^i \delta_{\tilde{x}_{t|t-1}^i}$.

Theorems A.1 and A.2 of Appendix A give general conditions ensuring convergence of $\hat{p}^N(y_t|\tilde{x}_{t|t-1}^i)$ to $\Psi_t(\tilde{x}_{t|t-1}^i)$ as N tends to ∞. But the interest of this approximation (and the possible recovering of the IPF properties) is impaired by the high computing cost induced at each step ($N \times n$ random variable generations).

The convolution kernel approach we propose to remedy to the theoretical and practical limitations of the particle filters do not suffer from this computing handicap. It uses a joint (x_t, y_t) density estimation

which does not demand extra $N \times n$ observation simulations at each step. Moreover the kernel density estimation theory allows a complete original theoretical study of the convergence properties of these new filters.

We shall introduce first the Convolution Filter, *i.e.* the convolution counterpart of the basic particle filter.

4. The Convolution Filter (CF)

The density of the optimal filter is defined by

$$p(x_t|y_{1:t}) = \frac{p_{XY}(x_t, y_{1:t})}{p_Y(y_{1:t})} \tag{2}$$

where $p_{XY}(x_t, y_{1:t})$ and $p_Y(y_{1:t})$ are the $(x_t, y_{1:t})$ joint density and the marginal density of $y_{1:t}$, respectively. For all the $y_{1:t}$ such that $p_Y(y_{1:t}) = 0$, we use the convention $p(x_t|y_{1:t}) \equiv 0$.

Let $z_t = (x_t, y_{1:t})$.

Assumptions:

- the distribution, π_0, of the initial state variable x_0, is known.

- the simulation of the noise variables ε_t and η_t according to their true distributions is possible for all $t \in \mathbb{N}$.

- there exists a probability measure μ_t such that $z_t \sim \mu_t$, for all t.

- there exists a probability measure ν_t such that $y_{1:t} \sim \nu_t$, for all t.

4.1. Kernel estimation of the optimal filter density

Let $\tilde{x}_0^i$ $(i = 1, \cdots, n)$ be generated according to π_0. For $i = 1, \ldots, n$: starting from $\tilde{x}_0^i$ a recursive simulation of system (1) t times successively, leads to $\tilde{z}_t^i = (\tilde{x}_t^i, \tilde{y}_{1:t}^i) \sim \mu_t$.

Empirical estimates of the measures μ_t and ν_t are given by

$$\mu_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{z}_t^i} \quad \text{and} \quad \nu_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{y}_{1:t}^i}$$

where $\delta_{\tilde{z}_t^i}$ and $\delta_{\tilde{y}_{1:t}^i}$ are Dirac measures. Kernel estimates p_{XY}^n and p_Y^n of the densities p_{XY} and p_Y are then obtained by convolution of these empirical measures with appropriate kernels:

12

$$p^n_{XY}(z_t) = K^z_{h_n} * \mu^n_t(z_t) = \frac{1}{n}\sum_{i=1}^n K^z_{h_n}(z_t - \tilde{z}^i_t)$$
$$= \frac{1}{n}\sum_{i=1}^n K^x_{h_n}(x_t - \tilde{x}^i_t)K^{\bar{y}}_{h_n}(y_{1:t} - \tilde{y}^i_{1:t})$$

and for p_Y

$$p^n_Y(y_{1:t}) = K^{\bar{y}}_{h_n} * \nu^n_t(y_{1:t}) = \frac{1}{n}\sum_{i=1}^n K^{\bar{y}}_{h_n}(y_{1:t} - \tilde{y}^i_{1:t}).$$

in which $K^z_{h_n}$, $K^x_{h_n}$ and $K^y_{h_n}$ are Parzen-Rosenblatt kernels of appropriate dimensions and $K^{\bar{y}}_{h_n}(y_{1:t} - \tilde{y}^i_{1:t}) = \Pi_{j=1}^t K^y_{h_n}(y_j - \tilde{y}^i_j)$.
An estimate of the density of the optimal filter $p_t(x_t|y_{1:t})$ is then given by:

$$
\begin{aligned}
p^n(x_t|y_{1:t}) &= \frac{p^n_{XY}(z_t)}{p^n_Y(y_{1:t})} \\
&= \frac{\sum_{i=1}^n K^z_{h_n}(z_t - \tilde{z}^i_t)}{\sum_{i=1}^n K^{\bar{y}}_{h_n}(y_{1:t} - \tilde{y}^i_{1:t})} \\
&= \frac{\sum_{i=1}^n K^x_{h_n}(x_t - \tilde{x}^i_t)K^{\bar{y}}_{h_n}(y_{1:t} - \tilde{y}^i_{1:t})}{\sum_{i=1}^n K^{\bar{y}}_{h_n}(y_{1:t} - \tilde{y}^i_{1:t})}.
\end{aligned}
\tag{3}
$$

The basic convolution filter (CF) is defined by this density estimator. As for the true densities, we use the convention : $p^n(x_t|y_{1:t}) \equiv 0$ for all the $y_{1:t}$ such that $p^n_Y(y_{1:t}) = 0$. Before giving convergence properties of CF, let us see a simple recursive algorithm for its practical computing.

4.2. CF ALGORITHM

– For $t = 0$

. generation of $\tilde{x}^1_0, \ldots, \tilde{x}^n_0 \sim \pi_0$.
. initialization of the weights : $w^i = 1$ for $(i = 1, \cdots, n)$.

– For $t \geq 1$

(i) evolving step : $\quad \tilde{x}^i_t \sim f_t(\tilde{x}^i_{t-1}, .)$ and $\tilde{y}^i_t \sim h_t(\tilde{x}^i_t, .)$,
for $(i = 1, \cdots, n)$

(ii) weight updating : $\quad w^i = K^y_{h_n}(y_t - \tilde{y}^i_t) \times w^i$,
for $(i = 1, \cdots, n)$

(iii) estimation :

$$p^n_t(x_t|y_{1:t}) = \frac{\sum_{i=1}^n w^i K^x_{h_n}(x_t - \tilde{x}^i_t)}{\sum_{i=1}^n w^i} \tag{4}$$

4.3. CONVERGENCE PROPERTIES OF THE CF FILTER

Several results of stochastic convergence and convergence rate are proposed in the following for the CF filter. Proofs are given in Appendix B.

Let us see first some results of ponctual convergence.

THEOREM 4.1 (quadratic mean convergence).

If K^x and $K^{\bar{y}}$ are Parzen-Rosenblatt kernels, if p_Y is positive and continuous at $y_{1:t}$ and if $p_t(x|y_{1:t})$ is bounded then,

$$
\left.\begin{array}{l}
\lim_{n\to\infty} h_n = 0 \\
\lim_{n\to\infty} nh_n^{tq+d} = \infty
\end{array}\right\} \implies \lim_{n\to\infty} \mathbb{E}[p_t^n(x_t|y_{1:t}) - p_t(x_t|y_{1:t})]^2 = 0
$$

The expectation is defined with respect to all the simulated variables $(\tilde{x}_t^i, \tilde{y}_{1:t}^i)_{i=1,\dots,n}$, conditional on the observed variables $y_{1:t}$.

THEOREM 4.2 (a.s. convergence).

If K^x and $K^{\bar{y}}$ are positive, bounded, Parzen-Rosenblatt kernels, if p_Y is positive and continuous at $y_{1:t}$ and if $p_{XY}(x_t, y_{1:t})$ is continuous at $(x_t, y_{1:t})$ then

$$
\begin{array}{l}
\lim_{n\to\infty} h_n = 0 \\
\lim_{n\to\infty} \frac{nh_n^{tq+d}}{\log n} = \infty
\end{array} \implies \lim_{n\to\infty} p_t^n(x_t|y_{1:t}) = p_t(x_t|y_{1:t}) \quad a.s.
$$

Results of L_1-convergence are now provided which have no equivalent for the usual (not regularized) particle filters since only the optimal filter probability measure is estimated by these filters.

THEOREM 4.3 (a.s. L_1-convergence).

If K^x and $K^{\bar{y}}$ are positive, bounded, Parzen-Rosenblatt kernels, if p_Y is positive and continuous at $y_{1:t}$ and if $x_t \mapsto p_{XY}(x_t, y_{1:t})$ is continuous for almost every x_t, then

$$
\begin{array}{l}
\lim_{n\to\infty} h_n = 0 \\
\lim_{n\to\infty} \frac{nh_n^{tq+d}}{\log n} = \infty
\end{array} \implies \lim_{n\to\infty} \int |p_t^n(x_t|y_{1:t}) - p_t(x_t|y_{1:t})| \mathrm{d}x_t = 0 \quad a.s
$$

Proof: Theorem 4.2 and Glick's theorem (A.3) ensure this result.

Let us complete this L_1-norm result by a convergence rate result, This necessitates to recall the notion of class of a kernel.

DEFINITION 4.1 (kernel of class s).

Let $s \geq 1$. A kernel of class s is a Borel measurable positive function K which satisfies
  (i)   *K is symmetric, $K(-x) = K(x)$, $x \in \mathbb{R}^d$.*
  (ii)   *$\int_{\mathbb{R}^d} K = 1$.*
  (iii)   *$\int_{\mathbb{R}^d} x^\alpha K(x)\mathrm{d}x = 0$ for $1 \leq \alpha \leq s - 1$.*
  (iv)   *$\int_{\mathbb{R}^d} |x^\alpha||K(x)|\mathrm{d}x < \infty$ for $|\alpha| = s$.*
where $\alpha \in \mathbb{N}^d$, $x^\alpha = x_1^{\alpha_1} \ldots x_d^{\alpha_d}$ and $|\alpha| = \alpha_1 + \ldots + \alpha_d$.

Note that, a kernel K of class $s > 2$ must necessary take negative values. Although in this paper we only consider positive kernels, in non-parametric estimation advanced works, (Devroye, 1987) for example, kernels taking negative values are allowed.

Let $W^{s,p}(\Omega) = \{f \in L_p(\Omega) \,|\, D^\alpha f \in L_p(\Omega), \; \forall \alpha : |\alpha| \leq s\}$ the standard Sobolev space.

THEOREM 4.4 (rate of L_1-convergence).
Suppose that $p_Y \in W^{s,1}(\mathbb{R}^{tq})$ and $p_{XY} \in W^{s,1}(\mathbb{R}^{tq+d})$ and that $K^{\bar{y}} \in L_1(\mathbb{R}^{tq})$ and $K^z \in L_1(\mathbb{R}^{tq+d})$ are kernels of class $s \geq 1$. Assume further that for some $\varepsilon > 0$, $K = K^{\bar{y}}, K^z$, $f = p_Y, p_{XY}$, $\delta = tq, tq + d$, we have $\int \|u\|^{\delta+\epsilon} K(u)^2 \mathrm{d}u < \infty$ and $\int (1 + \|u\|^{\delta+\epsilon}) f(u)\mathrm{d}u < \infty$. Then,

$$\mathbb{E}\Big[\int |p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})|\mathrm{d}x_t\Big] = O(h_n^s) + O(1/\sqrt{nh_n^{tq+d}})$$

The expectation is defined with respect to all the simulated random variables $(\tilde{x}_t^i, \tilde{y}_t^i)$ and the observation variables $y_{1:t}$.

4.4. COMMENTS

The assumptions $\lim_{n\to\infty} nh_n^{tq+d} = \infty$ and $\lim_{n\to\infty} h_n = 0$ used in the first three theorems imply that the number n of generated particles must grow with t to ensure convergence. This restricting condition was already required by the Monte Carlo filters built from the SIS algorithm. The same improvements can be considered. The first one is to limit the memory of the filter to T time steps backwards. It is easy to show that convergence results analogous to those just presented are obtained under the more satisfying assumptions $\lim_{n\to\infty} nh_n^{Tq+d} = \infty$ and $\lim_{n\to\infty} h_n = 0$. However, the optimal choice of T is difficult in practice and such a limited memory approximation of the optimal filter is only justified under mixing assumptions on the dynamical system (Del Moral, 2004). Other approaches are possible, as the introduction of a forgetting factor of the past. A more efficient one is the introduction of a trajectory selection step in which only simulated trajectories $(\tilde{y}_{1:t}^i)$

sufficiently close to the observed one $(y_{1:t})$ are kept, with a complete set of theoretical convergence results (Rossi, 2004). In this paper we present another improvement of the basic convolution filter using a resampling step analogous to that of the SIS-R filters, with the effect of relieving from a necessary particle number increase with time to improve the optimal filter estimation.

5. The Resampled-Convolution Filter (R-CF)

A resampling step can take place very easily at the beginning of each time step cycle of the basic CF algorithm, as follows.

5.1. R-CF Algorithm

– For $t = 0$

let p_n^0 be taken as the probability density of the initial state distribution π_0.

– For $t \geq 1$

   (i)   resampling step : $(\tilde{x}_{t-1}^1, \ldots, \tilde{x}_{t-1}^n) \sim p_{t-1}^n$

   (ii) evolving step : $\tilde{x}_t^i \sim f_t(\tilde{x}_{t-1}^i, .)$ and $\tilde{y}_t^i \sim h_t(\tilde{x}_t^i, .),$

                       for $i = 1, \ldots, n.$

   (iii) estimation step :

$$p_t^n(x_t|y_{1:t}) = \frac{\sum_{i=1}^n K_{h_n}^y(y_t - \tilde{y}_t^i) K_{h_n}^x(x_t - \tilde{x}_t^i)}{\sum_{i=1}^n K_{h_n}^y(y_t - \tilde{y}_t^i)} \tag{5}$$

Remark: even though the dependence of the x_t density estimator p_t^n on the past observed values $y_{1:t-1}$ is not explicit in (5), the resampling step makes it effective, as shown by the following convergent properties of the filter.

5.2. Convergence properties of the R-CF filter

All the convergence results proper to the CF filter are kept and even improved since they are now free of any particle number increase with time. We shall just show it through the study of the L_1-convergence of the filter.

THEOREM 5.1 (a.s. L_1-convergence). *If K^x and K^y are positive bounded Parzen-Rosenblatt kernels, if $p(\cdot|y_{1:t-1})$ is positive and continuous at y_t*

*and exists $M > 0$ such that $p(y_t|x_t) \leq M$ for all t, if exits $\alpha \in]0,1[$
such that $nh_n^{2q} = O(n^\alpha)$ then*

$$\begin{cases} \lim_{n\to\infty} \frac{nh_n^{d+q}}{\log n} = \infty \\ \lim_{n\to\infty} h_n = 0 \end{cases} \implies \lim_{n\to\infty} \int |p_t^n(x_t|y_{1:t}) - p_t(x_t|y_{1:t})| \mathrm{d}x_t = 0 \quad a.s$$

THEOREM 5.2 (rate of convergence).
Suppose that $p_Y = p(y_t|y_{1:t-1}) \in W^{s,1}(\mathbb{R}^q)$ and $p_{XY} = p(x_t, y_t|y_{1:t-1}) \in W^{s,1}(\mathbb{R}^{q+d})$ and that the kernels $K^y \in L_1(\mathbb{R}^q)$ and $K^{xy} = K^x K^y \in L_1(\mathbb{R}^{q+d})$ are of class $s \geq 1$. Assume further that for some $\varepsilon > 0$, for $K = K^y, K^{xy}$, $f = p_Y, p_{XY}$, $\delta = q, q+d$, we have $\int \|u\|^{\delta+\epsilon} K(u)^2 \mathrm{d}x < \infty$ and $\int (1 + \|u\|^{\delta+\epsilon}) f(u) \mathrm{d}u < \infty$. Then

$$\mathbb{E}\Big[\int |p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})| \mathrm{d}x_t\Big] = u_t[O(h_n^s) + O(1/\sqrt{nh_n^{q+d}})]$$

with $u_t = 2^t - 1$.

The expectation is defined with respect to all the simulated random variables $(\tilde{x}_t^i, \tilde{y}_t^i)$ and the observation variables $y_{1:t}$.

5.3. COMMENTS

As wanted, by comparison with the CF filter convergence conditions the assumptions $\lim_{n\to\infty} nh_n^{q+d}/\log n = \infty$ and $\lim_{n\to\infty} h_n = 0$ ensuring convergence of the R-CF to the optimal filter, are more satisfactory and the rate of convergence with respect to the particle number is less time dependent.

6. Simulated case studies

The behaviour of the proposed CF and R-CF convolution filters are now compared with that of their SIS and SIS-R counterparts, on a well-known bench-mark nonlinear dynamical system, for different state and observation noise situations.
Let us consider the following system, considered by (Netto & al., 1978), (Kitagawa, 1996; Kitagawa, 1998) and (Doucet, 1998; Doucet & al., 2001) among others.

$$\begin{cases} x_t = \frac{1}{2}x_{t-1} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t) + v_t \\ y_t = \frac{x_t^2}{20} + w_t \end{cases}$$

where $x_0 \sim N(0,5)$.

The competing filters are

MCF	:	Monte Carlo Filter
IPF	:	Interacted particle Filter
IPF-R	:	Interacted particle Filter post-Regularised
CF	:	Convolution Filter
R-CF	:	Resampled Convolution Filter

in the three different situations

Case 1	:	$v_t \sim N(0,1)$ and $w_t \sim N(0,0.1^2)$
Case 2	:	$v_t \sim N(0,1)$ and $w_t \sim N(0,1)$
Case 3	:	$v_t \sim N(0,10)$ and $w_t \sim N(0,1)$

For each of the three noise cases and a given number n of particles, the behaviours of the respective filters are compared through a mean squared error criterion computed for each filter over $N = 100$ trajectories of length $L = 500$ time steps,

$$MSE = \frac{1}{L} \sum_{t=1}^{L} \left(\frac{1}{N} \sum_{j=1}^{N} \parallel x_{j,t} - \hat{x}_{j,t} \parallel^2 \right)$$

in which $x_{j,t}$ is the true value of the state variable for the jth trajectory at time t and $\hat{x}_{j,t}$ is the estimate of $\mathrm{E}[x_{j,t}|y_{j,1}, \cdots, y_{j,t}]$ given by $\frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{j,t}^{i}$ the mean of the n particles generated according to the filter at time t.

A Gaussian kernel is used to build up the CF and R-CF filters and to regularize the IPF filter (*i.e.* to compute the IPF-R). The kernel bandwith is taken as $h_n = \mathrm{std}(\tilde{x}_t^1, \ldots, \tilde{x}_t^n)/n^{1/5}$, as usually recommended.

Table 1 shows that in Case 1 (small observation noise) only the CF and R-CF filters perform safely, whereas the MCF, IPF and IPF-R filters all diverge whatever the number of particles used.

For greater observation noise (Case 2 and 3) the performances of the interacting particle filters and convolution filters are rather close to each other, with a slight advantage for the R-CF filter.

7. Conclusion

The convolution kernel estimation theory offers a set of probability density estimating tools very efficient when large samples of observations distributed according to the distribution to be estimated are available. Following these lines the simulation of the state and observation variables of a dynamical system from a given initial state distribution, through the system noise simulation, allows to build up several types

Table I. Case 1 : $v_t \sim N(0, 1)$ and $w_t \sim N(0, 0.1^2)$

Nb of particles	MCF	IPF	IPF-R	CF	R-CF
N=20	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	**17.39**
N=50	$\emptyset$	$\emptyset$	$\emptyset$	16.80	**10.27**
N=100	$\emptyset$	$\emptyset$	$\emptyset$	14.70	**9.26**
N=200	$\emptyset$	$\emptyset$	$\emptyset$	13.98	**8.93**
N=500	$\emptyset$	$\emptyset$	$\emptyset$	12.92	**8.09**
N=5000	$\emptyset$	$\emptyset$	$\emptyset$	13.26	**7.58**

Table II. Case 2 : $v_t \sim N(0, 1)$ and $w_t \sim N(0, 1)$

Nb of particles	MCF	IPF	IPF-R	CF	R-CF
N=20	27.57	32.52	25.90	24.53	**24.33**
N=50	22.60	19.63	15.84	19.55	**15.70**
N=100	19.73	13.54	**11.67**	16.26	12.43
N=200	17.88	11.77	**10.91**	15.98	11.39
N=500	16.19	10.94	**10.70**	14.76	10.89
N=1000	14.66	**10.60**	10.64	14.29	10.65
N=5000	12.55	10.56	10.54	13.90	**10.46**

Table III. Case 3 : $v_t \sim N(0, 10)$ and $w_t \sim N(0, 1)$

Nb of particles	MCF	IPF	IPF-R	CF	R-CF
N=20	60.76	53.95	50.21	46.69	**38.55**
N=50	53.14	33.60	31.08	41.31	**27.99**
N=100	47.31	25.96	25.55	36.50	**24.75**
N=200	46.30	24.16	**23.70**	37.76	23.89
N=500	44.23	22.96	**22.59**	34.60	23.07
N=1000	41.62	**22.20**	22.26	36.27	22.31
N=5000	37.45	21.71	21.69	36.72	**21.68**

of convergent estimate of the optimal filter of the system, given an observed trajectory. This new convolution-based filtering approach can take place within the particle filter family. However besides the estimator construction itself, it differs from these filters by several important traits such as the required assumptions (no use of the observation likelihood), the structure of the noises (additive or not), the intrinsic regularization of the distribution estimate (a density), the estimation stability (enhanced robustness to small noise) and even no restriction about system without observation noise. Moreover, one can notice the

rather simple form taken by the convolution filter algorithms proposed, in particular that of the R-CF filter which achieved the best performances in all our case studies. The analytical study of the convergent properties of these new filters, quite different from that of the usual particle filters, takes advantage of general results from kernel estimation theory. Finally, no attention was paid in this paper to issues of more practical than theoretical importance as the choice of the type of convolution kernel to use and that of the kernel bandwidth parameter. A deeper recourse to nonparametric density estimation theory could help in improving these choices.

Appendix

A. Elements of kernel estimation theory

DEFINITION A.1. *A kernel K is a bounded, positive, symmetrical application from $\mathbb{R}^d \to \mathbb{R}$, such that $\int K \mathrm{d}\lambda = 1$, where λ is the Lebesgue measure.*

Example: the simple Gaussian kernel $K(x) = \left(\dfrac{1}{\sqrt{2\pi}}\right)^d \exp\left(-\dfrac{||x||^2}{2}\right)$.

DEFINITION A.2. *A Parzen-Rosenblatt kernel is a kernel such that*

$$\lim_{||x|| \to \infty} ||x||^d K(x) = 0$$

DEFINITION A.3. *Let $X_1, \cdots, X_n$ be i.i.d. random variables with common density f. The kernel estimator of f, f_n, associated with the kernel K is given by*

$$f_n(x) = \frac{1}{n h_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) = (K_{h_n} * \mu_n)(x) \qquad x \in \mathbb{R}^d$$

where the bandwidth parameter $h_n > 0$ and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure associated to $X_1, \cdots, X_n$.

We often use the practical notation $K_{h_n}(x) = \frac{1}{h_n^d} K(\frac{x}{h_n}) \qquad x \in \mathbb{R}^d$.

LEMMA A.1 (Bochner). *Let K be a Parzen-Rosenblatt kernel and $g \in L^1(\mathbb{R}^d)$, then for all x where g is continuous*

$$\lim_{h \to 0} (g * K_h)(x) = g(x)$$

20

Proof : see (Bosq & Lecoutre, 1987).

THEOREM A.1. *For any $f \in L_2$, if f_n is associated with a Parzen-Rosenblatt kernel, we have*

$$h_n \to 0, \ nh_n^d \to \infty \quad \Longrightarrow \quad \mathbb{E}[f_n(x) - f(x)]^2 \to 0$$

Proof : Theroem proved by (Parzen, 1962) for $d = 1$ and by (Cacoullos, 1966) for $d > 1$.

THEOREM A.2. *If f_n is associated with a positive bounded Parzen-Rosenblatt kernel, we have*

$$h_n \to 0, \ \frac{nh_n^d}{\log n} \to \infty \quad \Longrightarrow \quad f_n(x) \to f(x) \quad a.s.$$

whenever f is continuous at x.

Proof : see (Rao, 1983).

THEOREM A.3 (Glick). *If $\{f_n\}$ is a sequence of density estimates converging almost everywhere to a density in probability (or almost surely), then $\int |f_n - f| \to 0$ in probability (or almost surely).*

Proof : see (Glick, 1974) or (Devroye, 1987)

THEOREM A.4 (rate of convergence). *Let $f \in W^{s,1}$ be a probability density. Suppose that $K \in L_1(\mathbb{R}^d)$ is a kernel of class $s \geq 1$. Assume further that for some $\varepsilon > 0$ we have $\int \|x\|^{d+\epsilon} K(x)^2 dx < \infty$ and $\int (1 + \|x\|^{d+\epsilon}) f(x) dx < \infty$. Then,*

$$\mathbb{E}[\int |f_n - f|] = O(h_n^s) + O(1/\sqrt{nh_n^d})$$

Proof : see (Holmstrm & Klemel, 1992).

B. Proofs

B.1. Proof of Theorem 4.1 (quadratic mean convergence of the CF filter)

The expectations are defined with respect to all the simulated $(\tilde{x}_t^i, \tilde{y}_t^i)$, conditional on the observed $y_{1:t}$.

21

$$p^n(x_t|y_{1:t}) - p(x_t|y_{1:t}) = \frac{p^n(x_t|y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]}\Big(\mathbb{E}[p_Y^n(y_{1:t})] - p_Y^n(y_{1:t})\Big)$$

$$+ \frac{p_{XY}^n(x_t, y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]} - p(x_t|y_{1:t})$$

$$= \frac{p^n(x_t|y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]}\Big(\mathbb{E}[p_Y^n(y_{1:t})] - p_Y^n(y_{1:t})\Big)$$

$$+ \frac{p_{XY}^n(x_t, y_{1:t}) - p_{XY}(x_t, y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]}$$

$$+ \Big(\frac{p_{XY}(x_t, y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]} - \frac{p_{XY}(x_t, y_{1:t})}{p_Y(y_{1:t})}\Big)$$

According to Lemma A.1 (Bochner) one has

$$\lim_{n\to\infty} h_n \to 0, \ \lim_{n\to\infty} nh_n^{tq} \to \infty \Rightarrow \lim_{n\to\infty} \mathbb{E}[p_Y^n(y_{1:t})] = p_Y(y_{1:t})$$

which implies

$$\lim_{n\to\infty} \Big(\frac{p_{XY}(x_t, y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]} - \frac{p_{XY}(x_t, y_{1:t})}{p_Y(y_{1:t})}\Big) = 0.$$

In addition, Theorem A.1 ensures that

$$\lim_{n\to\infty} h_n \to 0, \ \lim_{n\to\infty} nh_n^{tq+d} \to \infty \Rightarrow \ \mathbb{E}[p_{XY}^n(x_t, y_{1:t}) - p_{XY}(x_t, y_{1:t})]^2 = 0$$

which implies

$$\lim_{n\to\infty} \mathbb{E}[\frac{p_{XY}^n(x_t, y_{1:t}) - p_{XY}(x_t, y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]}]^2 \to 0.$$

It only remains to study the behavior of

$$\mathbb{E}\Big[\frac{p^n(x_t|y_{1:t})}{\mathbb{E}[p_Y^n(y_{1:t})]}(\mathbb{E}[p_Y^n(y_{1:t})] - p_Y^n(y_{1:t}))\Big]^2.$$

By construction

$$p^n(x_t|y_{1:t}) = \frac{\sum_{i=1}^n K_h^x(x_t - \tilde{x}_t^i) K_h^{\bar{y}}(y_{1:t} - \tilde{y}_{1:t}^i)}{\sum_{i=1}^n K_h^{\bar{y}}(y_{1:t} - \tilde{y}_{1:t}^i)}$$
$$= \sum_{i=1}^n W_i K_h^x(x_t - \tilde{x}_t^i)$$

22

where
$$W_i = \frac{K_h^{\bar{y}}(y_{1:t} - \tilde{y}_{1:t}^i)}{\sum_{i=1}^n K_h^{\bar{y}}(y_{1:t} - \tilde{y}_{1:t}^i)} \quad i = 1, \cdots, n.$$

and by applying Jensen inequality

$$p^n(x_t|y_{1:t})^2 \leq \sum_{i=1}^n W_i K_h^x(x_t - \tilde{x}_t^i)^2$$

then

$$\mathbb{E}\left[p^n(x_t|y_{1:t})(\mathbb{E}[p_Y^n] - p_Y^n)|\tilde{y}_{1:t}^1, \ldots, \tilde{y}_{1:t}^n\right]^2$$

$$\leq \sum_{i=1}^n W_i \left(\mathbb{E}[p_Y^n] - p_Y^n\right)^2 \mathbb{E}\left[K_h^x(x_t, \tilde{x}_t^i)^2|\tilde{y}_{1:t}^i\right]$$

$$= \sum_{i=1}^n W_i \left(\mathbb{E}[p_Y^n] - p_Y^n\right)^2 \int K_h^x(x_t - \tilde{x}_t^i)^2 p(\tilde{x}_t^i|\tilde{y}_{1:t}^i) \mathrm{d}\tilde{x}_t^i$$

$$= \frac{1}{h_n^d} \sum_{i=1}^n W_i \left(\mathbb{E}[p_Y^n] - p_Y^n\right)^2 \int K^x(u)^2 p(x_t - uh|\tilde{y}_{1:t}^i) \mathrm{d}u$$

$$\leq \frac{M}{h_n^d} \left(\mathbb{E}[p_Y^n] - p_Y^n\right)^2 \int K^x(u)^2 \mathrm{d}u$$

where $M = \sup_{x_t, y_{1:t}} p(x_t|y_{1:t})$.
In addition

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[p_Y^n] - p_Y^n\right]^2 &= V\left[p_Y^n\right] \\
&= \frac{1}{n} V[K_h^{\bar{y}}(y_{1:t} - \tilde{y}_{1:t})] \\
&= \frac{1}{n}[E[K_h^{\bar{y}}(y_{1:t} - \tilde{y}_{1:t})^2] - E[K_h^{\bar{y}}(y_{1:t} - \tilde{y}_{1:t})]^2] \\
&= \frac{1}{n h_n^{tq}}\left((K_h^{\bar{y}})^2 * p_Y\right) - \frac{1}{n}\mathbb{E}[(p_Y^n)]^2
\end{aligned}$$

and according to Bochner's lemma (A.1) applied to p_Y with the kernel $(K^{\bar{y}})^2 / \int ((K^{\bar{y}})^2)$,

$$\lim_{h \to 0} \left((K_h^{\bar{y}})^2 * p_Y\right)(y_{1:t}) = p_Y(y_{1:t}) \int (K^{\bar{y}}(v))^2 \mathrm{d}v$$

then

$$n h_n^{tq} \mathbb{E}\left[\mathbb{E}[p_Y^n(y_{1:t})] - p_Y^n(y_{1:t})\right]^2 \to p_Y(y_{1:t}) \int (K^{\bar{y}}(v))^2 \mathrm{d}v.$$

23

Finally

$$nh_n^{tq+d}\mathbb{E}\Big[\frac{p_{X|Y}^n}{\mathbb{E}[p_Y^n(y_{1:t})]}\Big(\mathbb{E}[p_Y^n]-p_Y^n\Big)\Big]^2$$

$$\leq \frac{Mnh_n^{tq}\mathbb{E}\Big[\mathbb{E}[p_Y^n(y_{1:t})]-p_Y^n(y_{1:t})\Big]^2 \int K^x(u)^2\mathrm{d}u}{\mathbb{E}[p_Y^n(y_{1:t})]^2}$$

$$\to \frac{M\int K^x(u)^2\mathrm{d}u \int K^{\bar{y}}(v)^2\mathrm{d}v}{p_Y(y_{1:t})}$$

ensuring that

$$\lim_{n\to\infty}\mathbb{E}[p_t^n(x_t|y_{1:t})-p(x_t|y_{1:t})]^2 = 0.$$

$\square$

B.2. Proof of Theorem 4.2 (a.s. convergence of the CF filter)

By construction

$$p^n(x_t|y_{1:t}) = \frac{p^n(x_t,y_{1:t})}{p^n(y_{1:t})}$$

Theorem A.2 ensures that

$$p^n(x_t,y_{1:t}) \to p(x_t,y_{1:t}) \quad a.s.$$
$$p^n(y_{1:t}) \to p(y_{1:t}) \quad a.s.$$

As $p(y_{1:t})$ is assumed positive, the result is proved.

$\square$

B.3. Proof of Theorem 4.4 (L_1-convergence rate of the CF filter)

Unless specified the expectations are defined with respect to all the simulated $(\tilde{x}_t^i, \tilde{y}_t^i)$ and observed $y_{1:t}$ variables.

Let $B_+ = \{y_{1:t} : p_Y(y_{1:t}) > 0\}$, we have

$$\mathbb{E}\Big[\int |p^n(x_t|y_{1:t})-p(x_t|y_{1:t})|\mathrm{d}x_t\Big] = \mathbb{E}_{(\tilde{x}_t^i,\tilde{y}_t^i),B_+}\Big[\int |p^n(x_t|y_{1:t})-p(x_t|y_{1:t})|\mathrm{d}x_t\Big].$$

24

We shall then suppose $y_{1:t} \in B_+$ in the following equations :

$$p^n(x_t|y_{1:t}) - p(x_t|y_{1:t}) = \frac{p^n_{XY}(z_t)}{p^n_Y(y_{t:t})} - \frac{p_{XY}(z_t)}{p_Y(y_{t:t})}$$

$$= \frac{p^n_{XY}}{p^n_Y} - \frac{p_{XY}}{p_Y}$$

$$= \frac{p^n_{XY}p_Y - p_{XY}p^n_Y}{p^n_Y p_Y}$$

$$= \frac{p^n_{XY}p_Y - p^n_{XY}p^n_Y + p^n_{XY}p^n_Y - p_{XY}p^n_Y}{p^n_Y p_Y}$$

$$= \frac{p^n_{XY}(p_Y - p^n_Y) + p^n_Y(p^n_{XY} - p_{XY})}{p^n_Y p_Y}$$

$$= \frac{1}{p_Y}\left[p^n_{XY} - p_{XY} + (p_Y - p^n_Y)p^n_{X|Y}\right]$$

then

$$|p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})| \leq \frac{1}{p_Y(y_{1:t})}\Big[|p^n_{XY}(x_t, y_{1:t}) - p_{XY}(x_t, y_{1:t})|$$
$$+ |p_Y(y_{1:t}) - p^n_Y(y_{1:t})| p^n_{X|Y}(x_t|y_{1:t})\Big]$$

$$\int |p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})|\mathrm{d}x_t \leq \frac{1}{p_Y(y_{1:t})}\Big[\int |p^n_{XY}(x_t, y_{1:t}) - p_{XY}(x_t, y_{1:t})|\mathrm{d}x_t$$
$$+ |p_Y(y_{1:t}) - p^n_Y(y_{1:t})| \int p^n_{X|Y}(x_t|y_{1:t})\mathrm{d}x_t\Big]$$

$$= \frac{1}{p_Y(y_{1:t})}\Big[\int |p^n_{XY}(x_t, y_{1:t}) - p_{XY}(x_t, y_{1:t})|\mathrm{d}x_t$$
$$+ |p_Y(y_{1:t}) - p^n_Y(y_{1:t})|\Big]$$

Now

$$\mathbb{E}_{y_{1:t}}\left[\int |p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})|\mathrm{d}x_t\right] = \iint |p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})|\mathrm{d}x_t p_Y(y_{1:t})\mathrm{d}y_{1:t}$$

$$\leq \iint |p^n_{XY}(x_t, y_{1:t}) - p_{XY}(x_t, y_{1:t})|\mathrm{d}x_t \mathrm{d}y_{1:t}$$
$$+ \int |p_Y(y_{1:t}) - p^n_Y(y_{1:t})|\mathrm{d}y_{1:t}$$

or

$$\mathbb{E}_{y_{1:t}}[\|p^n_{X|Y} - p_{X|Y}\|_{L1}] \leq \|p^n_{XY} - p_{XY}\|_{L1} + \|p^n_Y - p_Y\|_{L1}$$

25

then

$$\mathbb{E}[\|p^n_{X|Y} - p_{X|Y}\|_{L1}] \leq \mathbb{E}[\|p^n_{XY} - p_{XY}\|_{L1}] + \mathbb{E}[\|p^n_Y - p_Y\|_{L1}].$$

Finally, according to Theorem A.4

$$\begin{aligned}
\mathbb{E}[\|p^n_{X|Y} - p_{X|Y}\|_{L1}] &= O(h^s_n) + O(1/\sqrt{nh^{tq+d}_n}) + O(h^s_n) + O(1/\sqrt{nh^{tq}_n}) \\
&= O(h^s_n) + O(1/\sqrt{nh^{tq+d}_n})
\end{aligned}$$

$\square$

B.4. Proof of Theorem 5.1

For $t = 1$ the result is true according to Theorem 4.3. Let us assume it is true until time t and let us show it is still true at time $t + 1$.
The three following lemmas will be useful to ensure this result.

LEMMA B.1. *Let $u(x)$ and $v(x)$ be two probability densities on $\mathbb{R}^d$ and $f = \min(u, v)$. Let U and V be respectively the subsets of $\mathbb{R}^d$ on which $\min(u, v) = u$ and $\min(u, v) = v$. Let $I = U \cap V$. Then $\frac{f}{1 - \frac{1}{2} \int |u-v| \mathrm{d}x}$ is a probability density.*

Proof of the lemma

$$\begin{aligned}
\int f &= \int_U u + \int_V v - \int_I u \\
&= 1 - \int_{U^c} u + \int_G v - \int_I u \\
&= 1 - [\int_V u - \int_I u] + \int_V v - \int_I u \\
&= 1 - \int_V u + \int_V v \\
&= 1 - \int_V (u - v) \\
&= 1 - \frac{1}{2} \int |u - v| \quad \text{according to Scheffe's lemma.}
\end{aligned}$$

$\square$

Now let
$$\Delta_n = \frac{1}{2}\|p^n_t(x_t|y_{1:t}) - p_t(x_t, |y_{1:t})\|_{L1} \tag{6}$$

By assumption
$$\lim_{n \to \infty} \Delta_n = 0 \quad a.s$$

Let $S^n_t = (\bar{x}^1_t, \ldots, \bar{x}^n_t)$ be sampled from p^n_t. We shall show that there exists a subsample of S^n_t, $\bar{x}^{i_1}_t, \ldots, \bar{x}^{i_{Mn}}_t$ and a new sample $\dot{x}^{i_1}_t, \ldots, \dot{x}^{i_{Nn}}_t$, which together can be considered as sampled from p_t. Such a device was used by (Devroye, 1987) to study the robustness of kernel estimates.

26

Let us define the three following functions:

$$
\begin{array}{rcl}
f_n &=& \frac{\min(p_t^n, p_t)}{1 - \Delta_n} \\
g_n &=& \frac{p_t^n - \min(p_t^n, p_t)}{\Delta_n} \\
h_n &=& \frac{p_t - \min(p_t^n, p_t)}{\Delta_n}
\end{array}
$$

Let us note that by Lemma B.1, f_n, g_n and h_n are density functions and that

$$
\begin{array}{rcl}
p_t^n &=& \Delta_n \cdot g_n + (1 - \Delta_n) \cdot f_n \\
p_t &=& \Delta_n \cdot h_n + (1 - \Delta_n) \cdot f_n
\end{array}
$$

This shows that each $\bar{x}_t^i$ sampled according to p_t^n is with probability Δ_n sampled from g_n. Let $Z_1, \ldots, Z_n$ be random variables such that $Z_i = 1$ if $\bar{x}_t^i \sim g_n$ and $Z_i = 0$ if $\bar{x}_t^i \sim f_n$. The Z_i are thus Bernoulli variables with parameter Δ_n. $N_n = \sum Z_i$ is then a binomial variable: $N_n \sim \mathcal{B}(n, \Delta_n)$.

$M_n = n - N_n$ is the random number of $\bar{x}_t^i$ sampled from f_n. Let $\bar{x}_t^{i_1}, \ldots, \bar{x}_t^{i_{M_n}}$ be this subsample, $1 \leq i_1 < \ldots < i_{M_n} \leq n$. Let $I_M = \{i_1, \ldots, i_{M_n}\}$ and $I_N = \{1, \ldots, n\} - I_{M_n}$.

Consider now the new following random sample

$$
\tilde{x}_t^i = \left\{ \begin{array}{ll} \bar{x}_t^i & \text{if } i \in I_M \\ \dot{x}_t^i & \text{with } \dot{x}_t^i \sim h_n, \text{ si } i \in I_N \end{array} \right. \quad \text{for } i = 1, \ldots, n \quad (7)
$$

$\tilde{x}_t^1, \ldots, \tilde{x}_t^n$ can then be considered as a virtual random sample from the unknown p_t which holds M_n elements common with the previous sample S_t^n drawn from p_t^n.

By applying system (1) to the actual $\bar{x}_t^1, \ldots, \bar{x}_t^n$ and then to the virtual $\tilde{x}_t^i$ for $i \in I_N$, one gets $(x_{t+1}^1, y_{t+1}^1), \ldots, (x_{t+1}^n, y_{t+1}^n)$ and $(\tilde{x}_{t+1}^i, \tilde{y}_{t+1}^i)$ for $i \in I_N$. Thus one gets new actual and virtual samples, $(x_{t+1}^1, y_{t+1}^1), \ldots \ldots, (x_{t+1}^n, y_{t+1}^n)$ and $(\tilde{x}_{t+1}^1, \tilde{y}_{t+1}^1), \ldots, (\tilde{x}_{t+1}^n, \tilde{y}_{t+1}^n)$ respectively, with M_n common pairs.

With the first sample, let us build the R-CF estimate of the optimal filter:

$$
p_{t+1}^n(x_{t+1}|y_{1:t}) = \frac{\sum_{i=1}^n K_h^x(x_{t+1} - x_{t+1}^i) K_h^y(y_{t+1} - y_{t+1}^i)}{\sum_{i=1}^n K_h^y(y_{t+1} - y_{t+1}^i)} \quad (8)
$$

With the second sample, we can consider the virtual kernel estimate of the optimal filter:

$$
\tilde{p}_{t+1}^n(x_{t+1}|y_{1:t}) = \frac{\sum_{i=1}^n K_h^x(x_{t+1} - \tilde{x}_{t+1}^i) K_h^y(y_{t+1} - \tilde{y}_{t+1}^i)}{\sum_{i=1}^n K_h^y(y_{t+1} - \tilde{y}_{t+1}^i)} \quad (9)
$$

built from particles $\tilde{x}_{t+1}^i$ themselves born from particles $\tilde{x}_t^i$ sampled from the true unknown $p_t(x_t|y_{1:t})$. This virtual estimate can then be considered as a simple convolution filter for one time step ahead.

Now we can write

$$\|p_{t+1}^n(x_{t+1}|y_{1:t}) - p_{t+1}(x_{t+1}|y_{1:t})\|_{L_1} \leq \|p_{t+1}^n(x_{t+1}|y_{1:t}) - \tilde{p}_{t+1}^n(x_{t+1}|y_{1:t})\|_{L_1} + \|\tilde{p}_{t+1}^n(x_{t+1}|y_{1:t}) - p_{t+1}(x_{t+1}|y_{1:t})\|_{L_1}$$

and according to Theorem 4.3 for CF filters, we have

$$\begin{cases} \lim_{n\to\infty} \frac{nh_n^{d+q}}{\log n} = \infty \\ \lim_{n\to\infty} h_n = 0 \end{cases} \implies \lim_{n\to\infty} \|\tilde{p}_{t+1}^n(x_{t+1}|y_{1:t}) - p_{t+1}(x_{t+1}|y_{1:t})\|_{L1} = 0 \quad a.s.$$

It thus remains to study the behaviour of

$$\|p_{t+1}^n(x_{t+1}|y_{1:t}) - \tilde{p}_{t+1}^n(x_{t+1}|y_{1:t})\|_{L_1} = \int |p_{t+1}^n(x_{t+1}|y_{1:t}) - \tilde{p}_{t+1}^n(x_{t+1}|y_{1:t})| \mathrm{d}x_{t+1}$$

Let

$$D_n(x_{t+1}) = p_{t+1}^n(x_{t+1}|y_{1:t}) - \tilde{p}_{t+1}^n(x_{t+1}|y_{1:t})$$

then by definition one has

$$\begin{aligned}
D_n(x_{t+1}) &= \frac{\sum_{i=1}^n K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i)}{\sum_{i=1}^n K_{h_n}(y_{t+1} - y_{t+1}^i)} \\
&\quad - \frac{\sum_{i=1}^n K_{h_n}(x_{t+1} - \tilde{x}_{t+1}^i)K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}{\sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)} \\
&= \frac{\sum_{i=1}^n K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i)}{\sum_{i=1}^n K_{h_n}(y_{t+1} - y_{t+1}^i)} \\
&\quad - \frac{\sum_{i=1}^n K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i)}{\sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)} \\
&\quad + \frac{\sum_{i=1}^n K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i)}{\sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)} \\
&\quad - \frac{\sum_{i=1}^n K_{h_n}(x_{t+1} - \tilde{x}_{t+1}^i)K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}{\sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}
\end{aligned}$$

28

it implies

$$|D_n(x_{t+1})| \leq \sum_{i=1}^{n} K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i)$$

$$\times \left| \frac{1}{\sum_{i=1}^{n} K_{h_n}(y_{t+1} - y_{t+1}^i)} - \frac{1}{\sum_{i=1}^{n} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)} \right|$$

$$+ \frac{1}{\sum_{i=1}^{n} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)} \left| \sum_{i=1}^{n} K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i) \right.$$

$$\left. - \sum_{i=1}^{n} K_{h_n}(x_{t+1} - \tilde{x}_{t+1}^i)K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i) \right|$$

As $(\tilde{x}_{t+1}^i, \tilde{y}_{t+1}^i) = (x_{t+1}^i, y_{t+1}^i)$ for $i \in I_M$, one gets

$$|D_n(x_{t+1})| \leq \frac{\sum_{i=1}^{n} K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i)}{\sum_{i=1}^{n} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)\sum_{i=1}^{n} K_{h_n}(y_{t+1} - y_{t+1}^i)}$$

$$\times \left| \sum_{i \in I_N} K_{h_n}(y_{t+1} - y_{t+1}^i) - \sum_{i \in I_N} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i) \right|$$

$$+ \frac{1}{\sum_{i=1}^{n} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)} \left| \sum_{i \in I_N} K_{h_n}(x_{t+1} - x_{t+1}^i)K_{h_n}(y_{t+1} - y_{t+1}^i) \right.$$

$$\left. - \sum_{i \in I_N} K_{h_n}(x_{t+1} - \tilde{x}_{t+1}^i)K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i) \right|$$

By assumption, the kernel K is a density, then

$$\int |D_n(x_{t+1})| \mathrm{d}x_{t+1} \leq \frac{\left| \sum_{i \in I_N} K_{h_n}(y_{t+1} - y_{t+1}^i) - \sum_{i \in I_N} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i) \right|}{\sum_{i=1}^{n} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}$$

$$+ \frac{\sum_{i \in I_N} K_{h_n}(y_{t+1} - y_{t+1}^i) + \sum_{i \in I_N} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}{\sum_{i=1}^{n} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}$$

Finally we get

29

$$\int |D_n(x_{t+1})| dx_{t+1} \leq 2 \frac{\sum_{i \in I_N} K_{h_n}(y_{t+1} - y_{t+1}^i) + \sum_{i \in I_N} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}{\sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}$$

$$\leq \frac{2N_n}{n} \frac{\frac{1}{N_n} \sum_{i \in I_N} K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i) + \frac{1}{N_n} \sum_{i \in I_N} K_{h_n}(y_{t+1} - y_{t+1}^i)}{\frac{1}{n} \sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}$$

$$\leq \frac{2N_n h_{N_n}^q}{n h_n^q} \frac{\frac{1}{N_n} \sum_{i \in I_N} \frac{K(\frac{y_{t+1} - \tilde{y}_{t+1}^i}{h_n})}{h_{N_n}^q} + \frac{1}{N_n} \sum_{i \in I_N} \frac{K(\frac{y_{t+1} - y_{t+1}^i}{h_n})}{h_{N_n}^q}}{\frac{1}{n} \sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}$$

According to theorem 4.2, it holds

$$\frac{1}{n} \sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i) \to p(y_{t+1}|y_{1:t}) \quad a.s.$$

and $p(y_{t+1}|y_{1:t})$ is positive by assumption.

Let us show that $\frac{1}{N_n} \sum_{i \in I_N} \frac{K(\frac{y_{t+1} - \tilde{y}_{t+1}^i}{h_n})}{h_{N_n}^q}$ and $\frac{1}{N_n} \sum_{i \in I_N} \frac{K(\frac{y_{t+1} - y_{t+1}^i}{h_n})}{h_{N_n}^q}$
are *a.s.* asymptotically bounded.
Let us consider the first term and to alleviate notations let

$$X_i = \frac{K(\frac{y_{t+1} - \tilde{y}_{t+1}^i}{h_n})}{h_{N_n}^q}, \quad \text{for } i \in I_N.$$

Let $M_n = \mathbb{E}[X_i|\mathcal{F}_t]$ and $Z_i = X_i - M_n$ with $\mathcal{F}_t$ the set of all generated variables through the instant t. The variables $Z_1, \ldots, Z_n$ are identically distributed and independant conditionally to $\mathcal{F}_t$.
We have

$$\mathbb{E}[(\sum_{i=1}^{N_n} Z_i)^4] = N_n \mathbb{E}[Z_1^4] + \frac{N_n(N_n - 1)}{2} \mathbb{E}[Z_1^2]^2$$

because $\mathbb{E}[Z_i Z_j Z_k Z_l] = \mathbb{E}[\mathbb{E}[Z_i|\mathcal{F}_t]\mathbb{E}[Z_j Z_k Z_l|\mathcal{F}_t]] = 0$ for all (j, k, l) among $\{1, \ldots, n\} - i$. According to Markov-Tchebychev inequality we deduce

$$P(|\frac{1}{N_n} \sum_{i=1}^{N_n} Z_i| > \varepsilon) \leq \frac{|\frac{1}{N_n} \sum_{i=1}^{N_n} Z_i|^4}{\varepsilon^4}$$

$$\leq \frac{\mathbb{E}[Z_1^4]}{N_n^3 \varepsilon^4} + \frac{\mathbb{E}[Z_1^2]^2}{2N_n^2 \varepsilon^4} \tag{10}$$

30

Let us study the terms $\mathbb{E}[Z_1^2]$ et $\mathbb{E}[Z_1^4]$:

$$
\begin{aligned}
\mathbb{E}[Z_1^2] &= \mathbb{E}[\mathbb{E}[Z_1^2|\mathcal{F}_t]] \\
&= \mathbb{E}[\mathbb{E}[X_1^2|\mathcal{F}_t] - \mathbb{E}[X_i|\mathcal{F}_t]^2] \\
&\leq \mathbb{E}[\mathbb{E}[X_1^2|\mathcal{F}_t]].
\end{aligned}
$$

However

$$
\begin{aligned}
\mathbb{E}[X_1^2|\mathcal{F}_t] &= \int \frac{K(\frac{y_{t+1}-\tilde{y}_{t+1}^i}{h_n})^2}{h_{N_n}^{2q}}\tilde{p}(\tilde{y}_{t+1}|y_{1:t})\mathrm{d}\tilde{y}_{t+1} \\[2mm]
&= \int \frac{h_n K(u)^2}{h_{N_n}^{2q}}\tilde{p}(y_{t+1}-hu|y_{1:t})\mathrm{d}u \\[2mm]
&\leq \int \frac{K(u)^2}{h_{N_n}^q}M\mathrm{d}u \\[2mm]
&\leq \frac{M_1}{h_{N_n}^q}
\end{aligned}
$$

with $M_1 = M\int K^2$ and $p(y_t|x_t) \leq M$, M exists by assumption. The inequality $\tilde{p}(\tilde{y}_{t+1}|y_{1:t}) \leq M$ rises from the joint-density $\tilde{p}(\tilde{x}_{t+1},\tilde{y}_{t+1}|y_{1:t})$ and that the link between $\tilde{x}_{t+1}$ and $\tilde{y}_{t+1}$ is in conformity with the model. Indeed, $\tilde{y}_{t+1}$ is obtained by applying the equation of observation of the system to $\tilde{x}_{t+1}$. Thus we can write

$$
\begin{aligned}
\tilde{p}(\tilde{y}_{t+1}|y_{1:t}) &= \int \tilde{p}(\tilde{x}_{t+1},\tilde{y}_{t+1}|y_{1:t})\mathrm{d}\tilde{x}_{t+1} \\
&= \int p(\tilde{y}_{t+1}|\tilde{x}_{t+1})\tilde{p}(\tilde{x}_{t+1}|y_{1:t})\mathrm{d}\tilde{x}_{t+1} \\
&\leq \int M\tilde{p}(\tilde{x}_{t+1}|y_{1:t})\mathrm{d}\tilde{x}_{t+1} \\
&\leq M
\end{aligned}
$$

Finally we get

$$
\mathbb{E}[Z_1^2] \leq \frac{M_1}{h_{N_n}^q}. \tag{11}
$$

Let us study now $\mathbb{E}[Z_1^4]$:

$$
\begin{aligned}
\mathbb{E}[Z_1^4] &= \mathbb{E}[\mathbb{E}[Z_1^4|\mathcal{F}_t]] \\
&= \mathbb{E}\Big[\mathbb{E}[X_1^4|\mathcal{F}_t] - 4\mathbb{E}[X_i^3|\mathcal{F}_t]\mathbb{E}[X_i|\mathcal{F}_t] - 4\mathbb{E}[X_i|\mathcal{F}_t]\mathbb{E}[X_i|\mathcal{F}_t]^3 \\
&\quad + 6\mathbb{E}[X_i^2|\mathcal{F}_t]\mathbb{E}[X_i|\mathcal{F}_t]^2 + \mathbb{E}[X_1|\mathcal{F}_t]^4\Big] \\
&\leq \mathbb{E}\Big[\mathbb{E}[X_1^4|\mathcal{F}_t] + 6\mathbb{E}[X_i^2|\mathcal{F}_t]\mathbb{E}[X_i|\mathcal{F}_t]^2\Big]
\end{aligned}
$$

31

However

$$\mathbb{E}[X_1^2|\mathcal{F}_t] \leq \frac{M_1}{h_{N_n}^q}$$

$$\mathbb{E}[X_1|\mathcal{F}_t] \leq \frac{Mh_n^q}{h_{N_n}^q}$$

and

$$\mathbb{E}[X_1^4|\mathcal{F}_t] = \int \frac{K(\frac{y_{t+1}-\tilde{y}_{t+1}^i}{h_n})^4}{h_{N_n}^{4q}} \tilde{p}(\tilde{y}_{t+1}|y_{1:t}) \mathrm{d}\tilde{y}_{t+1}$$

$$= \int \frac{h_n^q K(u)^4}{h_{N_n}^{4q}} \tilde{p}(y_{t+1} - hu|y_{1:t}) \mathrm{d}u$$

$$\leq \int \frac{K(u)^4}{h_{N_n}^{3q}} M \mathrm{d}u$$

$$\leq \frac{M_2}{h_{N_n}^{3q}}$$

with $M_2 = M \int K^4$. We get

$$\mathbb{E}[Z_1^4] \leq \frac{M_2}{h_{N_n}^{3q}} + 6 \frac{M_1}{h_{N_n}^q} \frac{M^2 h_n^{2q}}{h_{N_n}^{2q}} \tag{12}$$

The application of (11) and (12) to (10) leads to

$$P(|\frac{1}{N_n} \sum_{i=1}^{N_n} Z_i| > \varepsilon) \leq \frac{M_2}{N_n^3 h_{N_n}^{3q} \varepsilon^4} + \frac{6M_1 M^2}{N_n^3 h_{N_n}^q \varepsilon^4} + \frac{M_1^2}{2N_n^2 h_{N_n}^{2q} \varepsilon^4}$$

By assumption, exists $\alpha > 0$ such that $Nh_N^{2q} = O(N^\alpha)$, then the serie of general term $\frac{M_2}{N_n^3 h_{N_n}^{3q} \varepsilon^4} + \frac{6M_1 M^2}{N_n^3 h_{N_n}^q \varepsilon^4} + \frac{M_1^2}{2N_n^2 h_{N_n}^{2q} \varepsilon^4}$ converge. According to Borel Cantelli lemma, we deduce $\frac{1}{N_n} \sum_{i=1}^{N_n} Z_i$ converge *a.s.* to zero as N_n tends to infinity. Let

$$\lim_{N_n \to \infty} \frac{1}{N_n} \sum_{i \in I_N} \frac{K(\frac{y_{t+1}-\tilde{y}_{t+1}^i}{h_n})}{h_{N_n}^q} - \mathbb{E}[\frac{K(\frac{y_{t+1}-\tilde{y}_{t+1}^i}{h_n})}{h_{N_n}^q}|\mathcal{F}_t] = 0 \quad a.s.$$

as

$$0 \leq \mathbb{E}[\frac{K(\frac{y_{t+1}-\tilde{y}_{t+1}^i}{h_n})}{h_{N_n}^q}|\mathcal{F}_t] \leq \frac{Mh_n^q}{h_{N_n}^q}$$

where $h_n \le h_{N_n}$. Hence $\frac{1}{N_n}\sum_{i \in I_N} \frac{K(\frac{y_{t+1} - \tilde{y}_{t+1}^i}{h_n})}{h_{N_n}^q}$ is $a.s.$ asymptotically bounded.

The application of the same reasoning to $\frac{1}{N_n}\sum_{i \in I_N} \frac{K(\frac{y_{t+1} - y_{t+1}^i}{h_n})}{h_{N_n}^q}$ leads to the same result.

Thus for large N_n we have

$$\int |D_n(x_{t+1})|\mathrm{d}x_{t+1} \le \frac{2N_n h_{N_n}^q}{n h_n^q}\frac{\frac{Mh_n^q}{h_{N_n}^q} + \frac{Mh_n^q}{h_{N_n}^q}}{\frac{1}{n}\sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}$$

$$\le \frac{4N_n}{n}\frac{M}{\frac{1}{n}\sum_{i=1}^n K_{h_n}(y_{t+1} - \tilde{y}_{t+1}^i)}$$

And finally for large n and N_n

$$\int |D_n(x_{t+1})|\mathrm{d}x_{t+1} = O(\frac{N_n}{n}) \quad a.s. .$$

However $\frac{N_n}{n}$ is the empirical estimate of $\Delta_n \in [0, 1]$, by Hoeffding's Inequality (Hoeffding, 1963), for any Δ_n,

$$P(|\frac{N_n}{n} - \Delta_n| \ge \varepsilon) \le 2\exp\{-2n\varepsilon^2\}. \tag{13}$$

As $\lim_{n\to\infty} \Delta_n \to 0$ $a.s.$, (13) implies $\lim_{n\to\infty} \Delta_n = 0$ $a.s.$. This completes the proof of the theorem.

$\square$

B.5. Proof of Theorem 5.2

To alleviate notations let us denote transitorily
$p_Y = p(y_t|y_{1:t-1}), \quad p_{XY} = p(x_t, y_t|y_{1:t-1}), \quad p_{X|Y} = p(x_t|y_{1:t}).$
$p_Y^n = p^n(y_t|y_{1:t-1}), \quad p_{XY}^n = p^n(x_t, y_t|y_{1:t-1}), \quad p_{X|Y}^n = p^n(x_t|y_{1:t}),$
estimated from couples (x_t^i, y_t^i) born from particles $\bar{x}_{t-1}^i$ generated from $p^n(x_{t-1}|y_{1:t-1})$.
$\tilde{p}_Y^n = \tilde{p}(y_t|y_{1:t-1})$, the virtual estimate of p_Y from particles $\tilde{x}_{t-1}^i$ sampled from $p(x_{t-1}|y_{1:t-1})$ the true optimal filter distribution, as given by (7).
$\tilde{p}_{XY}^n = \tilde{p}^n(x_t, y_t|y_{1:t-1})$, the virtual estimate of p_{XY} from the same particles $\tilde{x}_t^i$.
Then by definition

33

$$p_{X|Y}^n - p_{X|Y} \quad = \quad \frac{p_{XY}^n}{p_Y^n} - \frac{p_{XY}}{p_Y}$$

$$= \quad \frac{p_{XY}^n p_Y - p_{XY} p_Y^n}{p_Y^n p_Y}$$

$$= \quad \frac{p_{XY}^n p_Y - p_{XY}^n p_Y^n + p_{XY}^n p_Y^n - p_{XY} p_Y^n}{p_Y^n p_Y}$$

$$= \quad \frac{p_{XY}^n (p_Y - p_Y^n) + p_Y^n (p_{XY}^n - p_{XY})}{p_Y^n p_Y}$$

$$= \quad \frac{1}{p_Y}\Big[p_{XY}^n - p_{XY} + (p_Y - p_Y^n) p_{X|Y}^n \Big].$$

$$|p_{X|Y}^n - p_{X|Y}| \quad \leq \quad \frac{1}{p_Y}\Big[|p_{XY}^n - p_{XY}| + |p_Y - p_Y^n| p_{X|Y}^n \Big].$$

We deduce

$$|p_{X|Y}^n - p_{X|Y}|\, p_Y \leq |p_{XY}^n - p_{XY}| + |p_Y - p_Y^n|\, p_{X|Y}^n.$$

We easily have

$$\mathbb{E}_{y_t|y_{1:t-1}}[\|p_{X|Y}^n - p_{X|Y}\|_{L1}] \quad \leq \quad \|p_{XY}^n - p_{XY}\|_{L1} + \|p_Y^n - p_Y\|_{L1}$$

and

$$\mathbb{E}[\|p_{X|Y}^n - p_{X|Y}\|_{L1}] \quad \leq \quad \mathbb{E}[\|p_{XY}^n - p_{XY}\|_{L1}] + \mathbb{E}[\|p_Y^n - p_Y\|_{L1}].$$

Now

$$\mathbb{E}[\|p_{XY}^n - p_{XY}\|_{L1}] \quad \leq \quad \mathbb{E}[\|p_{XY}^n - \tilde{p}_{XY}^n\|_{L1}] + \mathbb{E}[\|\tilde{p}_{XY}^n - p_{XY}\|_{L1}]$$

and

$$\mathbb{E}[\|p_Y^n - p_Y\|_{L1}] \quad \leq \quad \mathbb{E}[\|p_Y^n - \tilde{p}_Y^n\|_{L1}] + \mathbb{E}[\|\tilde{p}_Y^n - p_Y\|_{L1}].$$

As the virtual $\tilde{x}_t^i$ are generated according to the optimal filter $p_{X|Y}$, by Theorem A.4 it holds

$$\mathbb{E}[\|\tilde{p}^n_{XY} - p_{XY}\|_{L1}] = O(h^s_n) + O(1/\sqrt{nh^{q+d}_n})$$

and also

$$\mathbb{E}[\|\tilde{p}^n_Y - p_Y\|_{L1}] = O(h^s_n) + O(1/\sqrt{nh^q_n})$$

Now, let us turn to the term $\mathbb{E}[\|p^n_{XY} - \tilde{p}^n_{XY}\|_{L1}]$ and consider again the quantity Δ_n (6) previously introduced:

$$\Delta_n = \frac{1}{2}\|p^n_{t-1}(x_{t-1}|y_{1:t-1}) - p_{t-1}(x_{t-1},|y_{1:t-1})\|_{L1}$$

Let us notice that p^n_{XY} and $\tilde{p}^n_{XY}$ are built from $M_n = n - N_n$ common couples (x^i_t, y^i_t) born from M_n common particles $\bar{x}^i_{t-1}$, with $N_n \sim \mathcal{B}(n, \Delta_n)$. Then

$$
\begin{aligned}
\|p^n_{XY} - \tilde{p}^n_{XY}\|_{L1} &= \frac{1}{n}\int |\sum_{i \in I_N} K^x_{h_n}(x_t - x^i_t)K^y_{h_n}(y_t - y^i_t) \\
&\quad - K^x_{h_n}(x_t - \tilde{x}^i_t)K^y_{h_n}(y_t - \tilde{y}^i_t)|\mathrm{d}x_t\mathrm{d}y_t \\
&\le \frac{2N_n}{n}.
\end{aligned}
$$

As $\mathbb{E}[\frac{N_n}{n}|\Delta_n] = \Delta_n$, it holds

$$\mathbb{E}[\|p^n_{XY} - \tilde{p}^n_{XY})\|_{L_1}] \le 2\mathbb{E}[\Delta_n]$$

and by the same arguments

$$\mathbb{E}[\|p^n_Y - \tilde{p}^n_Y\|_{L_1}] \le 2\mathbb{E}[\Delta_n].$$

Finally

$$\mathbb{E}[\|p^n_{X|Y} - p_{Y|Y}\|] \le 4\mathbb{E}[\Delta_n] + O(h^s_n) + O(1/\sqrt{nh^{q+d}_n})$$

namely

$$
\begin{aligned}
\mathbb{E}\Big[\|p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})\|_{L_1}\Big] &\le 2\mathbb{E}[\|p^n_{t-1}(x_{t-1}|y_{1:t-1}) - p_{t-1}(x_{t-1},|y_{1:t-1})\|_{L1} \\
&\quad + O(h^s_n) + O(1/\sqrt{nh^{q+d}_n}).
\end{aligned}
$$

Furthermore for $t = 1$, according to Theorem A.4, it holds

$$\mathbb{E}\Big[\|p^n(x_1|y_1) - p(x_1|y_1)\|_{L_1}\Big] \le O(h^s_n) + O(1/\sqrt{nh^{q+d}_n})$$

which by ascending recursion leads to

$$\mathbb{E}\Big[\|p^n(x_t|y_{1:t}) - p(x_t|y_{1:t})\|_{L_1}\Big] \le (2^t - 1)\Big(O(h^s_n) + O(1/\sqrt{nh^{q+d}_n})\Big)_\square$$

References

Akashi H., Kumamoto, H. and Nose, K. Application of Monte Carlo Methods to Optimal Control for Linear Systems under Measurement Noise with Markov Dependent Statistical Property. *International Journal on Control*, vol.22, no. 6, p821-836, 1975.

Bosq D. and Lecoutre, J.P. *Thorie de l'estimation fonctionnelle.* Economica, Paris, 1987.

Cacoullos T. Estimation of a multivariate density. *Ann. Inst. Statist. Math.*, 18, p179-189, 1966.

Davis M. New Approach to Filtering nonlinear Systems. *IEEE Proceedings*, Part D, vol. 128, no. 5, p166-172, 1981.

Chen G. *Approximate Kalman Filtering.* World Scientific, Approximations and Decompositions vol. 2, 1993.

Del Moral, P., Rigal G., Salut G. Estimation et commande optimale non linaire : un cadre unifi pour la rsolution particulaire. Technical report 2, LAAS/CNRS, contrat DRET-DIGILOG, Toulouse, 1992.

Del Moral P. Nonlinear filtering using random particles. *Theory Probab. Appl* Vol40 No4, 1995.

Del Moral P. A uniform convergence theorem for the numerical solving of the nonlinear filtering problem. *Journal of Applied Probability*, 35(4):873-884, 1998.

Del Moral P. Measure-valued processes and interacting particule systems. Application to nonlinear filtering problems. *The Annals of Applied Probability*, vol. 8, No2, 438-495, 1998.

Del Moral P. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications.* Probability and its Apllications series, Springer, 2004.

Del Moral, P., Miclo, L. Branching and Interacting Particles systems Approximations of Feynman-Kac Formulae with applications to Non-Linear Filtering. In *Sminaire de Probabilit XXXIV*, Ed. J. Azéma, M. Emery, M. Ledoux & M. Yor, Lecture Notes in Mathematics, Springer-Verlag Berlin, vol 1729, p1-145, 2000.

Del Moral, P., Jacod, J. and Protter, P. The Monte-Carlo method for filtering with discrete-time observations. *Probab Theory Relat. Fields* 120, 346-368, 2001.

Del Moral, P., Jacod, J. Interacting Particle Filtering With Discrete Observation. In *Sequential Monte Carlo Methods in Practice* Ed. Doucet A., de Freitas, N., Gordon., N., Statistics for Engeering and Information Science, Springer, p43-75, 2001.

Devroye, L. *A Course in Density Estimation.* Birkhuser Boston, 1987.

Devroye, L. The equivalence of weak, strong and complete convergence in L1 for kernel density estimates. *Ann. Statist.*, vol 11, 896-904, 1983.

Doucet, A. On Sequential Simulation-Based Methods for Bayseian Filtering. Technical report CUED/F-INFENG/TR.310, University of Cambridge, 1998.

Doucet, A., De Freitas, N. and Gordon, N. *Sequential Monte Carlo Methods in Practice.* Statistics for Engeering and Information Science, Springer, 2001.

Glick, N. Consistency conditions for probablity estimators and integrals of density estimators. *Utilitas Mathematica*, vol. 6, p61-74, 1974.

Gordon, N.J., Salmond, D.J. and Smith, A.F.M Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, vol. 140, no. 2, 1993.

Handschin, J.E. Monte Carlo Techniques for Prediction and Filtering of Non-Linear Stochastic Processes. *Automatica*, no. 6, p555-563, 1970.

Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, vol. 58, p13-30

Hrzeler M. and Knsch, H.R. Monte Carlo Approximations for General State-Space Models. *Journal of Computational and Graphical Statistics*, vol.7, no.2, p175-193, 1998.

Holmstrm L, and Klemel, J. Asymptotic Bounds for the Excepted L^1 Error of a Multivariate Kernel Density Estimator. *Joural of Multivariate Analysis*, 42, p245-266, 1992.

Jazwinski, A.H. *Stochastic processes and filtering theory* . Academic Press, London, 1970.

Kitagawa, G. Non-Gaussian State-Space Modeling of Nonstationary Time Series. *Journal of the American Statistical Association*, 82(400), p1032-1041, 1987.

Kitagawa, G. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, vol.5, no.1, p1-25, 1996.

Kitagawa, G. A self-organisation state-space model. *Journal of the American Statistical Association*, 93(443), p1203-1215, 1998.

LeGLand F., Oudjane, N. A Robustification Approach to Stability, and to Uniform Particle Approximation of Nonlinear Filters. *Stochastic Process. and Appl.*, vol. 106, 2, pp. 279-316, 2003.

LeGLand F., Oudjane, N. Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters. *The Annals of Applied Probability*, 14, 1, pp. 144-187, 2004.

Liu J.S. and Chen, R. Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*, 93(443), p1032-1044, 1998.

Musso C., Oudjane, N., LeGland, F. Improving Regularised Particle Filters In *Sequential Monte Carlo Methods in Practice*, Eds : Doucet A., de Freitas N., Gordon N., Statistics for Engeering and Information Science, Springer-Verlag, New York, p247-271, 2001.

Netto, M.L.A., Gimeno, L. and Mendes, M.J. Nonlinear filtering of discrete time systems. *Proceedings of the 4th IFAC Symposium on Identification and System Parameter Estimation*, Tbilisi, USSR, p2123-2130, 1978.

Oudjane, N. *Stabilit et approximations particulaires en filtrage non linaires, application au pistage* PhD thesis, Universit de Rennes I, France, 2000. Statistics PhD.

Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, p1065-1076, 1962.

Rao, B.L.S. Prakasa *Non Parametric functionnal estimation.* Probability and Mathematical Statistics, Academic Press, Orlando, 1983.

Rossi, V. *Filtrage Non Linaire par Noyaux de Convolution. Application un Procd de Dpollution Biologique.* PhD thesis, Ecole Nationale Supérieure d'Agronomie de Montpellier, France, 2004. Statistics PhD.

Warnes, G.R. The Normal Kernel Coupler : An adaptave Markov Chain Monte Carlo method for efficiently sampling from multi-modal distributions. Technical Report 39, Departement of Statistics, University of Washington, 2001.