On the asymptotic variance in the Central Limit Theorem for particle filters

Benjamin Favetto

On the asymptotic variance in the Central Limit Theorem for particle filters

Benjamin Favetto

July 3, 2009

Laboratoire MAP5, Université Paris Descartes, U.F.R. de Mathématique et Informatique, CNRS UMR 8145
45, rue des Saints-Pères, 75270 Paris Cedex 06, France.
e-mail: Benjamin.Favetto@mi.parisdescartes.fr

Abstract

Particle filters algorithms approximate a sequence of distributions by a sequence of empirical measures generated by a population of simulated particles. In the context of Hidden Markov Models (HMM), they provide approximations of the distribution of optimal filters associated to these models. Given a set of observations, the asymptotic behaviour of particle filters, as the number of particles tends to infinity, has been studied: a central limit theorem holds with an asymptotic variance depending on the fixed set of observations. In this paper we establish, under general assumptions on the hidden Markov model, the tightness of the sequence of asymptotic variances when considered as functions of the random observations as the number of observations tends to infinity. We discuss our assumptions on examples and provide numerical simulations. The case of the Kalman filter is treated separately.

Key Words: Hidden Markov Model, Particle filter, Central Limit Theorem, Asymptotic variance, Tightness, Kalman model, Sequential Monte-Carlo

Short title: Asymptotic variances in particle filters approximation.

1 Introduction

Hidden Markov models (or state-space models) form a class of stochastic models which are used in numerous fields of applications. In these models, a discrete time process $(Y_n, n \geq 0)$ – the signal – is observed while the process of interest $(X_n, n \geq 0)$ – the state process – is not observed. The standard assumptions for the joint-process $(X_n, Y_n)_{n\geq 0}$ are that (X_n) is a Markov chain, that, given $(X_n, n \geq 0)$ the random variables (Y_n) are conditionally independent and the conditional distribution of Y_i only depends on the corresponding state variable X_i. (For general references, see *e.g.* Künsch (2001) or Cappé *et al.* (2005)).

Nonlinear filtering is concerned with the estimation of X_k or the prediction of X_{k+1} given the observations $(Y_0, \ldots, Y_k) := Y_{0:k}$. For this, one has to compute the conditional distributions $\pi_{k|k:0} = \mathcal{L}(X_k|Y_k, \ldots, Y_0)$ or $\eta_{k+1|k:0} = \mathcal{L}(X_{k+1}|Y_k, \ldots, Y_0)$ which are derived recursively by a sequence of measure-valued operators depending on the observations

$$\pi_{k|k:0} = \Psi_{Y_k}(\pi_{k-1|k-1:0}) \text{ and } \eta_{k+1|k:0} = \Phi_{Y_k}(\eta_{k|k-1:0})$$

(for more details, see *e.g.* Del Moral (2004) or Del Moral and Jacod (2001a)). Unfortunately, except for very few models, such as the Kalman filter or some other models (for instance, those presented in Chaleyat-Maurel and Genon-Catalot (2006)), these recursions rapidly lead to intractable computations and exact formulae are out of reach. Moreover, the standard Monte-Carlo methods fail to provide good approximations of these distributions (see *e.g.* the introduction in Van Handel (2008)). This justifies the huge popularity of sequential Monte-Carlo methods which are generally the only possible computing approach to solve these problems (see Doucet *et al.* (2001) or Robert and Casella (2004)). Sequential Monte-Carlo methods (or particle filters, or Interacting Particle Systems) are iterative algorithms based on simulated "particles" which provide approximations of the conditional distributions involved in prediction and filtering.

Denoting by $\pi^N_{k|k:0}$ (resp. $\eta^N_{k+1|k:0}$) the particle filter approximations of $\pi_{k|k:0}$ (resp. $\eta_{k+1|k:0}$) based on N particles, several recent contributions have been concerned with the evaluation of errors between the approximate and the exact filter as N grows to infinity, for a given (fixed) set of data $(Y_k, \ldots, Y_0)$ (see *e.g.* Douc *et al.* (2005)). In particular, for the bootstrap particle filter, Del Moral and Jacod (2001a) prove that, for a wide class of real-valued functions f, $\sqrt{N}(\pi^N_{k|k:0}(f) - \pi_{k|k:0}(f))$ (resp. $\sqrt{N}(\eta^N_{k+1|k:0}(f) - \eta_{k+1|k:0}(f))$) converges in distribution to $\mathcal{N}(0, \Gamma_{k|k:0}(f))$ (resp. $\mathcal{N}(0, \Delta_{k+1|k:0}(f))$). Central limit theorems for an exhaustive class of sequential Monte-Carlo methods are also proved in Chopin (2004).

To our knowledge, still little attention has been paid to the time behaviour (with respect to k) of the approximations. Recently, Van Handel (2008) has studied a uniform time average consistency of Monte-Carlo particle filters.

In this paper, we are concerned with the tightness of the asymptotic variances $\Gamma_{k|k:0}(f)$, $\Delta_{k+1|k:0}(f)$ in the central limit theorem for the bootstrap particle filter, when considered as random variables functions of $Y_0, \ldots, Y_k$ as $k \to \infty$. This is an important issue since these asymptotic variances measure the accuracy of the numerical method and provide confidence intervals. In Chopin (2004), for the case of the bootstrap filter, the asymptotic variance $\Gamma_{k|k:0}(f)$ is proved to be bounded from above by a constant, under stringent

assumptions on the conditional distribution of Y_i given X_i and on the transition densities of the unobserved Markov chain. In Del Moral and Jacod (2001b) the asymptotic variance $\Gamma_{k|k:0}(f)$ is proved to be tight (in k) in the case of the Kalman filter. The proof is based on explicit computations which are possible in this model. Below, we consider a general model and prove the tightness of both $\Gamma_{k|k:0}(f)$ and $\Delta_{k+1|k:0}(f)$ for f a bounded function under a set of assumptions which are milder than those in Chopin (2004) but which do not include the Kalman filter. In general, authors concentrate on filtering rather than on prediction as filtering in more important for applications. However, from the theoretical point of view, we stress the fact that prediction is simpler to study. First we prove the tightness of the asymptotic variances $\Delta_{k+1|k:0}(f)$ obtained in the central limit theorem for prediction, and then we are able to deduce the analogous result for $\Gamma_{k|k:0}(f)$. For the transition kernel of the Markov chain, we rely on a strong assumption, which mainly holds when the state space of the hidden chain is compact (Assumption **(A)**). Nevertheless, such an assumption is of common use in this kind of studies (see *e.g.* Oudjane and Rubenthaler (2005) , Atar and Zeitouni (1997), Douc *et al.* (2005)). In the sense of Douc *et al.* (2005), it means that the whole state space of the hidden chain is “small” (see Douc *et al.* (2005)). On the other hand, our assumptions on the conditional distributions of Y_i given X_i are very weak (**(B1)-(B2)**). The Kalman filter model which does not satisfy our assumption **(A2)** is treated separately.
The paper is organized as follows. In Section 2, we present our notations and assumptions, and give the formulae for $\Gamma_{k|k:0}(f)$ and $\Delta_{k+1|k:0}(f)$ and some preliminary propositions in order to obtain formulae as simple as possible for the asymptotic variances. Section 3 is devoted to the proof of the tightness of $\Delta_{k+1|k:0}(f)$ from which we deduce the tightness of $\Gamma_{k|k:0}(f)$. In Section 5, we look at the Kalman filter model as in Del Moral and Jacod (2001b) and propose simplifications for the computation of $\Delta_{k+1|k:0}(f)$ and for proving its tightness. Moreover, we illustrate our assumptions on examples and provide some numerical simulation results.

2 Notations, assumptions and preliminary results

Let (X_k) be the time-homogeneous hidden Markov chain, with state space $\mathcal{X}$ and transition kernel $Q(x, dx')$. The observed random variables (Y_k) take values in another space $\mathcal{Y}$ and are conditionally independent given $(X_k)_{k\geq 0}$ with $\mathcal{L}(Y_i|(X_k)_{k\geq 0}) = F(X_i, dy)$. For $0 \leq i \leq k$, denote by $Y_{i:k}$ the vector $(Y_i, Y_{i+1} \dots Y_k)$.
Denote by $\pi^{\eta_0}_{k|k:0} = \mathcal{L}_{\eta_0}(X_k|Y_{k:0})$ (resp. $\eta^{\eta_0}_{k|k-1:0} = \mathcal{L}_{\eta_0}(X_k|Y_{k-1:0})$) the filtering distribution (resp. the predictive distribution) at step k when η_0 is the initial distribution of the chain (distribution of X_0). By convention, $\eta^{\eta_0}_{0|-1:0} = \eta_0$.

Let us now introduce our assumptions. Assumptions **A** concern the hidden chain, Assumptions **B** concern the conditional distribution of Y_i given X_i.

(A0) $\mathcal{X}$ is an interval of $\mathbb{R}$. The transition operator Q admits transition densities with respect to the Lebesgue measure on $\mathcal{X}$ denoted by dx' : $Q(x, dx') = p(x, x')dx'$. The transition densities are positive and continuous on $\mathcal{X} \times \mathcal{X}$. For φ bounded and continuous on $\mathcal{X}$, $Q\varphi$ is bounded and continuous on $\mathcal{X}$ (Q is Feller).

(A1) The transition operator Q admits a stationary distribution $\pi(dx)$ having a density h with respect to dx which is continuous and positive on $\mathcal{X}$.

(A2) There exists a probability measure μ and two positive numbers $\epsilon_- \leq \epsilon_+$ such that

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}(\mathcal{X}) \quad \epsilon_- \mu(B) \leq Q(x, B) \leq \epsilon_+ \mu(B).$$

Moreover, for all f continuous and positive on $\mathcal{X}$, $\mu(f) > 0$.

(B1) $\mathcal{Y} = \mathbb{R}$, the conditional distribution of Y_k given X_k has density $f(y|x)$ with respect to a dominating measure $\kappa(dy)$, and $(x, y) \longmapsto f(y|x)$ is measurable and positive.

(B2) $x \longmapsto f(y|x)$ is continuous and bounded from above for all y κ a.e.

Under **(B2)**, $q(y) = \sup_{x \in \mathcal{X}} f(y|x)$ is well defined and positive. Up to changing $\kappa(dy)$ into $\frac{1}{q(y)}\kappa(dy)$, we can assume without loss of generality that

$$\forall x \in \mathcal{X}, \quad f(y|x) \leq 1. \tag{1}$$

Except **(A2)** these assumptions are weak and standard. For instance, **(A0)-(A1)** easily hold for discretized one-dimensional diffusions with constant discretization step. Assumption **(A2)**, which is the most stringent, is nevertheless classical and is verified when $\mathcal{X}$ is compact. (see Atar and Zeitouni (1997) and the chronological discussion in Douc *et al.* (2009)). Assumptions **B** are mild. Note that they are much weaker than the corresponding ones in Chopin (2004) and the same as in Van Handel (2008). By **(A0)**, for φ non null, non negative and continuous on $\mathcal{X}$, $Q\varphi > 0$. With **(B2)**, for all y κ a.e., $Q(f(y|.))$ is positive, continuous and bounded (by 1).

Note that in **(A0)** we could replace $\mathcal{X}$ an interval of $\mathbb{R}$ by a convex subset of $\mathbb{R}^d$, and $\mathbb{R}$ by $\mathbb{R}^d$ in **(B1)**.

Some more notations are needed for the sequel. Define the family of operators : for $f : \mathcal{X} \longrightarrow \mathbb{R}$ measurable and bounded, and $k \geq 0$,

$$L_k f(x) = g_k(x) Q f(x) \text{ where } g_k(x) := f(Y_k|x). \tag{2}$$

For $0 \leq i \leq j$, let $L_{i,j} := L_i \dots L_j$ denote the compound operator. For η a probability measure on $\mathcal{X}$, set

$$\Phi_{Y_k}(\eta)(f) = \frac{\eta L_k f}{\eta L_k \mathbf{1}}.$$

Then the predictive distributions satisfy $\eta^{\eta_0}_{0|-1:0} = \eta_0$ and for $k \geq 1$,

$$\eta^{\eta_0}_{k|k-1:0} f = \frac{\eta^{\eta_0}_{k-1|k-2:0} L_{k-1} f}{\eta^{\eta_0}_{k-1|k-2:0} L_{k-1} \mathbf{1}} = \Phi_{Y_{k-1}}(\eta^{\eta_0}_{k-1|k-2:0})(f) \tag{3}$$

By iteration,

$$\mathbb{E}_{\eta_0}\left(f(X_k)|Y_{0:k-1}\right) = \eta^{\eta_0}_{k|k-1:0}(f) = \Phi_{Y_{k-1}} \circ \cdots \circ \Phi_{Y_0}(\eta_0)(f) = \frac{\eta_0 L_{0,k-1} f}{\eta_0 L_{0,k-1} \mathbf{1}} \tag{4}$$

For δ_x the Dirac mass at x, we have

$$\eta^{\delta_x}_{k|k-1:i} f = \frac{\delta_x L_{i,k-1} f}{\delta_x L_{i,k-1} \mathbf{1}} = \Phi_{Y_{k-1}} \circ \cdots \circ \Phi_{Y_i}(\delta_x) f. \tag{5}$$

We will simply set $\eta_{k|k-1:i} f(x) := \eta^{\delta_x}_{k|k-1:i} f$. Moreover we have the relations

$$\begin{aligned} \mathbb{E}_{\eta_0}\left(f(X_k)|Y_{0:k}\right) &= \pi^{\eta_0}_{k|k:0} f = \frac{\eta^{\eta_0}_{k|k-1:0}(g_k f)}{\eta^{\eta_0}_{k|k-1:0}(g_k)} && (6) \\ \text{and } \eta^{\eta_0}_{k+1|k:0}(f) &= \pi^{\eta_0}_{k|k:0}(Qf). && (7) \end{aligned}$$

Note that for all y, $\Phi_y(\delta_x)(dx') = p(x,x')dx'$. For $\eta_0(dx) = h_0(x)dx$, with h_0 positive and continuous on $\mathcal{X}$,

$$\Phi_y(\eta_0)(dx') = \frac{\int_{\mathcal{X}} dx f(y|x) h_0(x) p(x,x')}{\int_{\mathcal{X}} dx f(y|x) p(x,x')} dx',$$

where the denominator is positive. Hence $\Phi_y(\eta)$ has a positive and continuous density when η is a Dirac mass or has a positive and continuous density. For these reasons and assumption **(A0)**, all denominators appearing in our formulae are positive.

Below, for simplicity, when no confusion is possible, we omit the sub- or superscript η_0 in the distributions. Denote the number of interacting particles by N. The distribution of the bootstrap particle filter for the prediction is denoted by $\eta^N_{k|k-1:0}(f)$ and the distribution of the bootstrap particle filter for the filter is denoted by $\pi^N_{k|k:0}(f)$. The following theorem central limit theorem is proved in Del Moral and Jacod (2001a).

Theorem. *For f a bounded measurable function and a given sequence $(Y_{0:k})$ of observations, the following convergences in distribution hold*

$$\sqrt{N}(\eta^N_{k|k-1:0}(f) - \eta_{k|k-1:0}(f)) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \Delta_{k|k-1:0}(f))$$

where

$$\Delta_{k|k-1:0}(f) = \sum_{i=0}^{k} \frac{\eta_{i|i-1:0}\left(\left(L_{i,k-1}\left(f - \eta_{k|k-1:0}f\right)\right)^2\right)}{(\eta_{i|i-1:0}L_{i,k-1}\mathbf{1})^2}, \tag{8}$$

and

$$\sqrt{N}(\pi^N_{k|k:0}(f) - \pi_{k|k:0}(f)) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{k|k:0}(f))$$

where

$$\Gamma_{k|k:0}(f) = \sum_{i=0}^{k} \frac{\eta_{i|i-1:0}\left(\left(L_{i,k-1}\left(g_k\left(f - \pi_{k|k:0}f\right)\right)\right)^2\right)}{(\eta_{k|k-1:0}(g_k))^2(\eta_{i|i-1:0}L_{i,k-1}\mathbf{1})^2}. \tag{9}$$

Note that, in Del Moral and Jacod (2001a), the above theorem is proved for a wider class of functions, including functions with polynomial growth.

In the sequel, we focus on the two asymptotic variances $\Delta_{k|k-1:0}(f)$ and $\Gamma_{k|k:0}(f)$ for f bounded, when considered as functions of $Y_{0:k}$. Proposition 1 gives the link between the two quantities. Recall that the initial distribution is fixed equal to η_0.

Proposition 1. *For f a bounded function, and k a non negative integer,*

$$\Gamma_{k|k:0}(f) = \Delta_{k|k-1:0}\left(\frac{g_k}{\eta_{k|k-1:0}(g_k)}\left(f - \pi_{k|k:0}f\right)\right) \tag{10}$$

and

$$\Delta_{k+1|k:0}(f) = \eta_{k+1|k:0}\left(\left(f - \eta_{k+1|k:0}f\right)^2\right) + \Gamma_{k|k:0}\left(Qf - \pi_{k|k:0}Qf\right). \tag{11}$$

Proof. The first formula is immediate from (8) and (9). Using (3)-(7), we get

$$\Delta_{k|k-1:0}(f) = \sum_{i=0}^{k} \eta_{i|i-1:0}\left(\left(\frac{L_{i,k-1}\mathbf{1}(.)}{\eta_{i|i-1:0}L_{i,k-1}\mathbf{1}}\right)^2\left(\eta_{k|k-1:i}f(.) - \eta_{k|k-1:0}f\right)^2\right) \tag{12}$$

$$\Gamma_{k|k:0}(f) = \sum_{i=0}^{k} \frac{\eta_{i|i-1:0}\left(\left(\frac{L_{i,k-1}\mathbf{1}(.)}{\eta_{i|i-1:0}L_{i,k-1}\mathbf{1}}\right)^2\left(\eta_{k|k-1:i}(g_kf)(.) - \pi_{k|k:0}f\right)^2\right)}{(\eta_{k|k-1:0}(g_k))^2} \tag{13}$$

Noting that

$$\eta_{i|i-1:0}\left(L_{i,k-1}\mathbf{1}\right)\,\eta_{k|k-1:0}\left(g_k\right) = \eta_{i|i-1:0}\left(L_{i,k}\mathbf{1}\right),$$

we derive

$$\begin{aligned}\Delta_{k+1|k:0}(f) &= \eta_{k+1|k:0}\left(\left(f-\eta_{k+1|k:0}f\right)^2\right) \\ &\quad +\Delta_{k|k-1:0}\left(\frac{g_k}{\eta_{k|k-1:0}(g_k)}\left(Qf-\eta_{k+1|k:0}f\right)\right) \\ &= \eta_{k+1|k:0}\left(\left(f-\eta_{k+1|k:0}f\right)^2\right)+\Gamma_{k|k:0}\left(Qf-\eta_{k+1|k:0}f\right).\end{aligned}$$

□

Note that Chopin (2004) derives some related formulae in a slightly different context. Due to all the above relations, it appears that the predictive distributions and the asymptotic variance $\Delta_{k+1|k:0}(f)$ are simpler to study.

3 Tightness of the asymptotic variances

To stress the dependence on the observations (Y_k), we introduce another notation for $\eta^{\nu}_{k|k-1:0}$. For ν a probability measure, A a borelian set, $y_{0:k-1}$ a set of fixed real values, let us introduce

$$\begin{aligned}\eta_{\nu,k}[y_{0:k-1}](A) &= \frac{\mathbb{E}_{\nu}(\prod_{i=0}^{k-1} f(y_i|X_i)\mathbf{1}_A(X_k))}{\mathbb{E}_{\nu}(\prod_{i=0}^{n-1} f(y_i|X_i))} \\ &= \frac{\nu L_{0,k-1}\mathbf{1}_A}{\nu L_{0,k-1}\mathbf{1}} = \frac{\nu g_0 Q g_1 Q \dots g_{k-1} Q \mathbf{1}_A}{\nu g_0 Q g_1 Q \dots g_{k-1}}\end{aligned}$$

Here, $\mathbb{E}_{\nu}$ denotes the expectation with respect to the distribution of the chain (X_k) with initial distribution ν and $y_{0:k-1}$ are fixed values not involved in the expectation. To ensure that all expressions are well defined, we consider probability measures ν equal to either Dirac masses or probabilities with positive continuous densities on $\mathcal{X}$. In the second line, we have set $g_i(x) = f(y_i|x)$ and the formula explains the backward iterations of the operators (2) with $Y_k = y_k$.

The following proposition proves the exponential forgetting of the initial distribution for the predictive distributions.

Proposition 2. *Assume* **(A2)** *and* **(B1)** *and set* $\rho = 1 - \frac{\epsilon_-^2}{\epsilon_+^2}$. *Then for all non negative integer* k, *all probability distributions* ν *and* ν' *on* $\mathcal{X}$ *and all set* $y_{0:k-1}$ *of real values*

$$\|\eta_{\nu,k}[y_{0:k-1}] - \eta_{\nu',k}[y_{0:k-1}]\|_{TV} \le \rho^k,$$

where $\|.\|_{TV}$ *denotes the total variation distance.*

Proof. The above result is generally proved for the filtering distribution (see *e.g.* Atar and Zeitouni (1997), Del Moral and Guionnet (2001) and Douc *et al.* (2009)). To prove it for the predictive distributions, we follow the scheme of Douc *et al.* (2009). Let $\bar{\mathcal{X}} = \mathcal{X} \times \mathcal{X}$ and denote by $\bar{Q}$ the Markov kernel on $\bar{\mathcal{X}}$ given by

$$\bar{Q}((x,x'), A \times A') = Q(x,A)Q(x',A').$$

Set $\bar{g}_i(x,x') = g_i(x)g_i(x')$. For two probability measures ν and ν', notice that

$$\eta_{\nu,k}[y_{0:k-1}](A) - \eta_{\nu',k}[y_{0:k-1}](A) = \Phi_{y_{k-1}} \circ \cdots \circ \Phi_{y_0}(\nu)(\mathbf{1}_A) - \Phi_{y_{k-1}} \circ \cdots \circ \Phi_{y_0}(\nu')(\mathbf{1}_A)$$

$$= \frac{\mathbb{E}_{\nu\otimes\nu'}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i)\mathbf{1}_A(X_k)) - \mathbb{E}_{\nu'\otimes\nu}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i)\mathbf{1}_A(X_k))}{\mathbb{E}_{\nu}(\prod_{i=0}^{k-1} g_i(X_i))\mathbb{E}_{\nu'}(\prod_{i=0}^{k-1} g_i(X_i))}$$

where (X_i) and (X'_i) are two independent copies of the hidden Markov chain and $\mathbb{E}_{\nu\otimes\nu'}$ denotes the expectation with respect to the distribution of the chain (X_i, X'_i) with kernel $\bar{Q}$ and initial distribution $\nu \otimes \nu'$.
Set $\bar{\mu} = \mu \otimes \mu$, and $\bar{x} = (x, x')$. For $\bar{f}$ a measurable non negative function, we have

$$\epsilon_-^2 \bar{\mu}(\bar{f}) \leq \bar{Q}(\bar{x}, \bar{f}) \leq \epsilon_+^2 \bar{\mu}(\bar{f}).$$

Setting

$$\bar{Q}_0(\bar{x}, \bar{f}) = \epsilon_-^2 \bar{\mu}(\bar{f}) \quad \text{and} \quad \bar{Q}_1(\bar{x}, \bar{f}) = \bar{Q}(\bar{x}, \bar{f}) - \bar{Q}_0(\bar{x}, \bar{f}),$$

we deduce that

$$0 \leq \bar{Q}_1(\bar{x}, \bar{f}) \leq \rho \bar{Q}(\bar{x}, \bar{f}).$$

Now let us compute the numerator:

$$R_k(\nu, \nu', A) = \mathbb{E}_{\nu\otimes\nu'}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i)\mathbf{1}_A(X_k)) - \mathbb{E}_{\nu'\otimes\nu}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X'_i)\mathbf{1}_A(X_k))$$

$$= \nu \otimes \nu'(\bar{g}_0 \bar{Q} \bar{g}_1 \ldots \bar{g}_{k-1} \bar{Q} \mathbf{1}_{A\times\mathcal{X}}) - \nu' \otimes \nu(\bar{g}_0 \bar{Q} \bar{g}_1 \ldots \bar{g}_{k-1} \bar{Q} \mathbf{1}_{A\times\mathcal{X}}).$$

It may be decomposed as

$$R_k(\nu, \nu', A) = \sum_{t_{0:k-1} \in \{0,1\}^k} R_k(A, t_{0:k-1})$$

where

$$R_k(A, t_{0:k-1}) := \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \ldots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A\times\mathcal{X}}) - \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \ldots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A\times\mathcal{X}}).$$

Assume that for an index i, $t_i = 0$. Then

$$\begin{aligned}
&\nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\
&= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \dots \bar{g}_{i-1} \bar{Q}_{t_{i-1}} \bar{g}_i) \times \epsilon_-^2 \bar{\mu} \left(\bar{g}_{i+1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}\right) \\
&= \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \dots \bar{g}_{i-1} \bar{Q}_{t_{i-1}} \bar{g}_i) \times \epsilon_-^2 \bar{\mu} \left(\bar{g}_{i+1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}\right) \\
&= \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\
&= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{\mathcal{X} \times A})
\end{aligned}$$

and $R_k(A, t_{0:k-1})$ vanishes except if all $t_i = 1$. Hence

$$R_k(\nu, \nu', A) = \nu \otimes \nu'(\bar{g}_0 \bar{Q}_1 \bar{g}_1 \dots \bar{g}_{k-1} \bar{Q}_1 (\mathbf{1}_{A \times \mathcal{X}} - \mathbf{1}_{\mathcal{X} \times A})).$$

Therefore

$$\sup_A \left| R_k(\nu, \nu', A) \right| \leq \rho^k \mathbb{E}_{\nu \otimes \nu'}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X_i')).$$

The result follows.□

Remark. Applying the result of Proposition 2 with $g_i \equiv 1$ and $\nu' = \pi$, we get $\|\nu Q^k - \pi\|_{TV} \leq \rho^k$. Thus, **(A1)-(A2)** imply the geometric ergodicity of (X_k).

The following propositions give upper bounds for $\Delta_{k|k-1:0}(f)$.

Proposition 3. *Assume* **(A0)-(A2) (B1)-(B2)** *For f a bounded measurable function and $(Y_{0:k})$ a sequence of observations, the following inequalities hold*

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \sum_{i=0}^{k} \eta_{i|i-1:0} \left(\left(\frac{L_{i,k-1}\mathbf{1}}{\eta_{i|i-1:0} L_{i,k-1}\mathbf{1}} \right)^2 \right) \rho^{2(k-i)} \tag{14}$$

Proof. Remark that, for all ν:

$$\eta_{\nu,k}[Y_{0:k-1}] = \Phi_{Y_{k-1}} \circ \dots \circ \Phi_{Y_i}(\eta_{\nu,i}[Y_{0:i-1}]).$$

By Proposition 2, we deduce, for $\nu = \eta_0$:

$$\begin{aligned}
|\eta_{k|k-1:i}(f)(x) - \eta_{k|k-1:0}^{\eta_0}(f)| &\leq \|f\|_\infty \|\eta_{\delta_x, k-i}[Y_{i:k-1}] - \eta_{\eta_0,k}(Y_{0:k-1})\|_{TV} \\
&\leq \|f\|_\infty \|\Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_i}(\delta_x) - \Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_i}(\eta_{\eta_0,i}[Y_{0:i-1}])\|_{TV} \\
&\leq \|f\|_\infty \rho^{k-i}.
\end{aligned}$$

Using (12), we get the result. □

We stress the fact that Proposition 3 only relies on the exponential stability which may hold even if **(A2)** is not satisfied (see Douc *et al.* (2009)).

Proposition 4. *Under Assumption* **(A2)** *and for f measurable and bounded, it holds that*

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \frac{\epsilon_+^2}{\epsilon_-^2} \sum_{i=0}^{k} \eta_{i|i-1:0} \left(\left(\frac{g_i}{\eta_{i|i-1:0} g_i} \right)^2 \right) \rho^{2(k-i)}. \tag{15}$$

Proof. We remark that for any probability measure ν

$$\epsilon_- \nu(g_i)\mu(g_{i+1}Q\dots g_{k-1}) \le \nu g_i Q g_{i+1} Q \dots g_{k-1} \le \epsilon_+ \nu(g_i)\mu(g_{i+1}Q\dots g_{k-1}).$$

By **(A2)**, since Q is Feller and the g_l's are positive continuous, $\mu(g_{i+1}Q\dots g_{k-1})$ is positive. Applying the left inequality with $\nu = \eta_{i|i-1:0}$ and the right inequality with $\nu = \delta_x$, it comes

$$\frac{L_{i,k-1}\mathbf{1}(x)}{\eta_{i|i-1:0}L_{i,k-1}\mathbf{1}} \le \frac{\epsilon_+}{\epsilon_-}\frac{g_i(x)\mu(g_{i+1}Q\dots g_{k-1})}{\eta_{i|i-1:0}(g_i)\mu(g_{i+1}Q\dots g_{k-1})}.$$

Thus,

$$\Delta_{k|k-1:0}(f) \le \|f\|_\infty^2 \frac{\epsilon_+^2}{\epsilon_-^2}\sum_{i=0}^{k}\eta_{i|i-1:0}\left(\left(\frac{g_i}{\eta_{i|i-1:0}g_i}\right)^2\right)\rho^{2(k-i)}.$$

□

We state the main result under an additional assumption that will be discussed in Section 4 where we show that, under **(A1)**, the additional assumption **(B3)** is especially easy to check.

Theorem 1. *Assume that*

(B3) *for some $\delta > 0$*

$$\sup_{k\ge 0}\mathbf{E}\left|\log\left(\eta_{k|k-1:0}\left(g_k\right)\right)\right|^{1+\delta} < \infty, \tag{16}$$

where $\mathbf{E}$ denotes the expectation with respect to the distribution of $(Y_k)_{k\ge 0}$.

Then, for all bounded function f, the sequences of variances $(\Delta_{k|k-1:0}(f))$ and $(\Gamma_{k|k:0}(f))$ are tight.

Proof. Using that $g_i \le 1$

$$\eta_{i|i-1:0}\left(\left(\frac{g_i}{\eta_{i|i-1:0}g_i}\right)^2\right) \le \frac{1}{\eta_{i|i-1:0}g_i}.$$

Setting $B_i = -\log(\eta_{i|i-1:0}g_i)$, Lemma 1 (see the Appendix) implies that the sequence $\sum_{i=0}^k e^{B_i}\rho^{2(k-i)}$ is tight with respect to k. With Proposition 4, we deduce that $(\Delta_{k|k-1:0}(f))_{k\ge 0}$ is tight.

Using (10) and $g_k \le 1$, we obtain:

$$\Gamma_{k|k:0}(f) \le \frac{1}{(\eta_{k|k-1:0}g_k)^2}\Delta_{k|k-1:0}\left(f - \pi_{k|k:0}f\right).$$

Since $\|f - \pi_{k|k:0}f\|_\infty \le 2\|f\|_\infty$, the first part implies that $(\Delta_{k|k-1:0}\left(f - \pi_{k|k:0}f\right))$ is tight. By **(B3)**, $(\eta_{k|k-1:0}(g_k))$ is also tight. The result follows. □

4 Discussion and examples

4.1 Checking of (B3)

Let us consider a hidden chain with state space $\mathcal{X} = [a, b]$ a compact interval of $\mathbb{R}$ satisfying **(A0)-(A2)** (for instance a discrete sampling of a diffusion on $[a, b]$ with reflecting boundaries). Under **(B2)**, $r(y) = \inf_{x\in\mathcal{X}} f(y|x)$ is well defined and positive. Thus, we have

$$r(Y_k) \le \eta_{k|k-1:0}(g_k) \le 1.$$

Therefore, the condition $\sup_{k\ge 0} \mathbf{E} \left|\log\left(r(Y_k)\right)\right|^{1+\delta} < \infty$ implies **(B3)**. In particular, when (Y_k) is stationary *i.e.* when the initial distribution of the chain is $\eta_0 = \pi$ the stationary distribution, the condition is simply $\mathbf{E}\left|\log\left(r(Y_0)\right)\right|^{1+\delta} < \infty$. Let us compute $r(y)$ in some typical examples.

Example 1. Assume that $Y_k = X_k + \varepsilon_k$ with $\varepsilon_k \sim_{i.i.d.} \mathcal{N}(0, 1)$ and (X_k) independent of (ε_k). The observation kernel is $F(x, dy) = \mathcal{N}(x, 1)$. Choosing the dominating measure $\kappa(dy) = \frac{1}{\sqrt{2\pi}}dy$,

$$f(y|x) = \exp\left(-\frac{(y-x)^2}{2}\right) \ge r(y)$$

where

$$|\log(r(y))| \le \frac{1}{2}\left((y-a)^2 + (y-b)^2\right).$$

Example 2. Assume that $Y_k = \sqrt{X_k}\varepsilon_k$ with $\varepsilon_k \sim_{i.i.d.} \mathcal{N}(0, 1)$, (X_k) independent of (ε_k) and $0 < a < b$. The observation kernel is $F(x, dy) = \mathcal{N}(0, x)$. Note that

$$\frac{1}{\sqrt{2\pi b}}\exp\left(-\frac{y^2}{2a}\right) \le \frac{1}{\sqrt{2\pi x}}\exp\left(-\frac{y^2}{2x}\right) \le \frac{1}{\sqrt{2\pi a}}.$$

Taking $\kappa(dy) = \frac{1}{\sqrt{2\pi a}}dy$, we get that

$$|\log(r(y))| \le C + \frac{y^2}{2a}.$$

Thus, assumption **(B3)** is a simple moment condition on the observations which is evidently satisfied on these examples.

4.2 The case of a diffusion on a compact manifold

Consider the stochastic differential equation

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dW_t$$

with a one-dimensional observation process

$$Y_{t_i} = g(Z_{t_i}) + \varepsilon_i$$

where W is a standard Brownian motion, (ε_i) is an i.i.d. sequence of $\mathcal{N}(0,1)$ random variables, and b, σ are Lipschitz and g is smooth enough.

Assume that the diffusion process Z valued in a compact manifold M of dimension m embedded in $\mathbb{R}^d$. Assume that b and σ lead to a strictly elliptic generator on M, with heat kernel $G_t(x,y)$. We refer to Atar and Zeitouni (1997) and Davies (1989) for the following inequality

$$c_0 e^{-c_1/t} \leq G_t(x,y) \leq c_2 t^{-m/2}.$$

where c_0, c_1 and c_2 are numerical constants. Assume that $t_i = i\delta, \quad \delta > 0, \quad i \in \mathbb{N}$, hence the observations are equally spaced in time. Then we obtain the inequality for **(A2)** with μ a probability distribution with positive density with respect to Lebesgue measure on M, because the transition density of the hidden Markov chain is bounded from below by a positive value. Due to the underlying diffusion process, other assumptions on the chain are verified.

4.3 A toy-example

Consider two continuous densities u and v on $(0,1)$, a distribution π on $(0,1)$ with a continuous density with respect to Lebesgue measure and a real α of $(0,1)$. Define the Markov chain (X_k) by

$$X_0 \sim \pi, \qquad X_{k+1} = \mathbf{1}_{X_k<\alpha} U_{k+1} + \mathbf{1}_{X_k \geq \alpha} V_{k+1} \tag{17}$$

where (U_k) and (V_k) are two independent sequences of i.i.d. random variables, independent of X_0, with respective distributions $u(x)dx$ and $v(x)dx$. Set $p(x,x') = \mathbf{1}_{x<\alpha} u(x') + \mathbf{1}_{x\geq\alpha} v(x')$ the transition kernel density. The transition kernel admits an invariant distribution $\pi(x)dx$ with $\pi(x) = Au(x) + (1-A)v(x)$ and

$$A = \frac{\int_0^\alpha v(x)dx}{\int_\alpha^1 u(x)dx + \int_0^\alpha v(x)dx}.$$

For example, with

$$u(x) = \left\{ \begin{array}{c} 6x \text{ if } x \in [0, \frac{1}{3}] \\ -3x+3 \text{ if } x \in]\frac{1}{3}, 1] \end{array} \right. \quad \text{and } v(x) = \left\{ \begin{array}{c} 3x \text{ if } x \in [0, \frac{2}{3}] \\ -6x+6 \text{ if } x \in]\frac{2}{3}, 1] \end{array} \right.$$

the transition kernel Q of the chain (X_k) satisfies **(A2)** (see Figure 1) with $\mu(dx) = 4(x \wedge 1-x)dx$ and $\epsilon_- = \frac{1}{4}, \quad \epsilon_+ = \frac{3}{2}$:

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}(\mathcal{X}) \quad \epsilon_- \mu(B) \leq Q(x,B) \leq \epsilon_+ \mu(B).$$

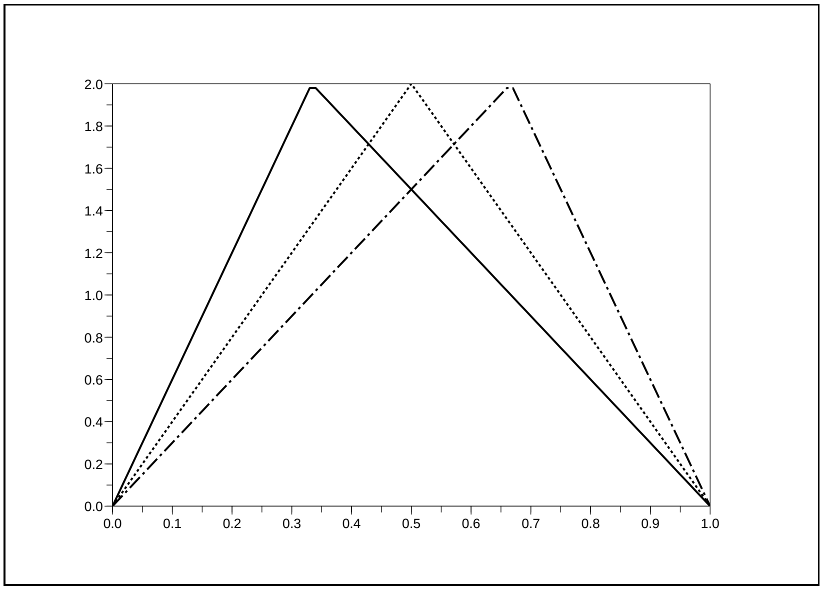

Figure 1: Densities involved in the toy-example. Functions u (solid), v (bigdash dot) and $\frac{d\mu}{dx}$ (dash dot).

In Figure 1, the graph of u is plotted in solid line, the graph of v is plotted in bigdash dotted line, and the density of μ is plotted in dash dotted line. For this example, the transition density, $p(x, x')$ is not bounded from below by a positive constant. This shows that **(A2)** is strictly weaker than the assumption of Theorem 5 in Chopin (2004). Although Assumption **(A0)** is not verified on this example, the proof of the tightness still holds. Indeed, all the denominators involved in the computations of the upper bounds are well-defined and positive. Assumption **(A1)** is clearly verified and the stationary distribution can be explicitely computed, and the bounds in **(A2)** too.

5 The Gaussian case

In the context of the one-dimensional Kalman filter model, Assumptions **(A0)-(A1) (B1)-(B2)** hold but not **(A2)**. On the other hand, we are able to study the expression (12) of $\Delta_{k|k-1:0}(f)$ by explicit computations. In Del Moral and Jacod (2001b), (13) is proved to be tight by direct computations too. We do the analogous calculus for (12) which is simpler. Recall the model: the hidden Markov chain is a Gaussian AR(1) process

$$\begin{aligned} X_0 &\sim \mathcal{N}(0, \sigma_s^2) \\ X_{k+1} &= aX_k + \beta U_{k+1} \end{aligned}$$

where (U_k) is a sequence of $\mathcal{N}(0,1)$ independent random variables, independent of X_0. The observations are given by

$$Y_k = bX_k + \beta' V_k$$

where (V_k) is a sequence of $\mathcal{N}(0,1)$ independent random variables, independent of (X_k). We assume that $|a| < 1$ and $\sigma_s^2 = \frac{\beta^2}{1-a^2}$. Hence, the process (X_k, Y_k) is stationary.

5.1 Preliminary computations

Denote by η_{m,σ^2} the Gaussian distribution $\mathcal{N}(m,\sigma^2)$ and by ϕ_{m,σ^2} its density. Due to the identity

$$\begin{aligned}
\frac{1}{\sigma^2}(x-m)^2 + \frac{1}{v^2}(x-u)^2 &= \left(\frac{1}{\sigma^2}+\frac{1}{v^2}\right)\left(x - \frac{\frac{m}{\sigma^2}+\frac{u}{v^2}}{\frac{1}{\sigma^2}+\frac{1}{v^2}}\right)^2 - \frac{\left(\frac{m}{\sigma^2}+\frac{u}{v^2}\right)^2}{\frac{1}{\sigma^2}+\frac{1}{v^2}} \\
&\quad + \left(\frac{m^2}{\sigma^2}+\frac{u^2}{v^2}\right),
\end{aligned}$$

we get

$$\eta_{m,\sigma^2}(\phi_{u,v^2}) = \frac{1}{\sqrt{2\pi\left(\sigma^2+v^2\right)}}\exp\left(-\frac{(m-u)^2}{2(v^2+\sigma^2)}\right). \tag{18}$$

The prediction operator L_k is given by $L_k(x,.) = \frac{1}{b}\phi_{\frac{Y_k}{b},\frac{\beta'^2}{b^2}}(x)\mathcal{N}(ax,\beta^2)$. To compute (12), following Del Moral and Jacod (2001b) we search the compound operator $L_{i,j} = L_i \ldots L_j$ in the following form:

$$L_{i,j}(x,.) = u_{i,j}\phi_{v_{i,j},w_{i,j}}(x)\mathcal{N}(\theta_{i,j}x+\gamma_{i,j},\delta_{i,j}).$$

After some technical computations, the parameters are recursively given by:

$$u_{i,i} = \frac{1}{b},\quad v_{i,i} = \frac{Y_i}{b},\quad w_{i,i} = \frac{\beta'^2}{b^2},\quad \theta_{i,i} = a,\quad \gamma_{i,i} = 0,\quad \delta_{i,i} = \beta^2.$$

For $i < j$

$$\begin{aligned}
\theta_{i,j} &= \tfrac{a w_{i+1,j}\theta_{i+1,j}}{\beta^2 + w_{i+1,j}}, & \gamma_{i,j} &= \gamma_{i+1,j} + \theta_{i+1,j}\left(\tfrac{\beta^2 v_{i+1,j}}{\beta^2+w_{i+1,j}}\right), \\
\delta_{i,j} &= \delta_{i+1,j} + \theta_{i+1,j}^2\left(\tfrac{\beta^2 w_{i+1,j}}{\beta^2+w_{i+1,j}}\right), & v_{i,j} &= \tfrac{\frac{Y_i}{b}(\beta^2+w_{i+1,j}) + a\frac{\beta'^2}{b^2}v_{i+1,j}}{(\beta^2+w_{i+1,j})+\frac{\beta'^2}{b^2}a^2}, \\
w_{i,j} &= \tfrac{\beta'^2}{b^2}\tfrac{\beta^2+w_{i+1,j}}{(\beta^2+w_{i+1,j})+\frac{\beta'^2}{b^2}a^2}, & &
\end{aligned}$$

$$u_{i,j} = \frac{u_{i+1,j}}{\sqrt{2\pi\left((\beta^2+w_{i+1,j})+\frac{\beta'^2}{b^2}a^2\right)}}\exp\left(-\frac{1}{2}\frac{(v_{i+1,j}-a\frac{Y_i}{b})^2}{(\beta^2+w_{i+1,j})+\frac{\beta'^2}{b^2}a^2}\right).$$

It appears that the recursion for $w_{i,j}$ is driven by an autonomous algorithm, which has a unique fixed point $\underline{w}$ defined as the solution of

$$\underline{w} = \frac{\beta'^2}{b^2} \frac{\beta^2 + \underline{w}}{(\beta^2 + \underline{w}) + \frac{\beta'^2}{b^2} a^2}.$$

The downward recursions for all parameters can be solved by elementary but extremely cumbersome computations. Since we just want to illustrate the way tightness is obtained for (12), we do not proceed to the exact computations. Instead, we introduce a *stabilized* operator

$$\underline{L}_k(x, .) = \frac{1}{b} \phi_{\frac{Y_k}{b}, \underline{w}}(x) \mathcal{N}(ax, \beta^2)$$

and compute the compound operator $\underline{L}_{i,j} = \underline{L}_i \dots \underline{L}_j$. We denote by $\underline{\Delta}_{k|k-1:0}(f)$ the *stabilized* asymptotic variance, *i.e.* (12) computed with $\underline{L}_k(., .)$ instead of the exact L_k.

5.2 Solving recursions for the stabilized operators

Define

$$\underline{L}_{i,j}(x, .) = u_{i,j} \phi_{v_{i,j}, \underline{w}}(x) \mathcal{N}(\theta_{i,j} x + \gamma_{i,j}, \delta_{i,j}) \tag{19}$$

and set

$$\alpha = \frac{\underline{w}}{\beta^2 + \underline{w}}, \quad \tau = \frac{\beta^2 + \underline{w}}{\beta^2 + \underline{w} + \frac{\beta'^2}{b^2} a^2}. \tag{20}$$

The recursions are as follows:

$$\theta_{i,j} = a\alpha\theta_{i+1,j}, \quad \gamma_{i,j} = \gamma_{i+1,j} + (1 - \alpha)\theta_{i+1,j} v_{i+1,j},$$
$$\delta_{i,j} = \delta_{i+1,j} + \beta^2 \alpha \theta^2_{i+1,j}, \quad v_{i,j} = \tau \frac{Y_i}{b} + a\alpha v_{i+1,j})$$

where the initial values are as previously, except that $w_{i,i} = \underline{w}$. Now the recursions are easily solved. We get:

$$\theta_{i,j} = a(a\alpha)^{j-i}, \qquad \delta_{i,j} = \beta^2 (1 + a \sum_{l=0}^{j-i-1} (a\alpha)^{2l+1}), \tag{21}$$

$$\gamma_{i,j} = (1-\alpha) a \sum_{l=0}^{j-i-1} (a\alpha)^l v_{j-l,j}, \qquad v_{i,j} = \frac{\tau}{b} \left(\sum_{l=1}^{j-i} (a\alpha)^{j-i-l} Y_{j-l} \right) + (a\alpha)^{j-i} \frac{Y_j}{b}.$$

We finally obtain closed formulae for $\gamma_{i,j}$ and $u_{i,j}$. By (12) we must compute

$$f_{i,k-1}(x) = \frac{\underline{L}_{i,k-1} \mathbf{1}(x)}{\eta_{i|i-1:0} \underline{L}_{i,k-1} \mathbf{1}} = \frac{u_{0,i-1} \phi_{v_{0,i-1}, \underline{w}}(x)}{u_{0,i-1} \eta_{i|i-1:0}(\phi_{v_{i,k}, \underline{w}})}$$

Hence the term $u_{0,i-1}$ is compensated. Then, we obtain $\eta_{i|i-1:0} = \mathcal{N}(m_{0,i-1}, s^2_{0,i-1})$ with

$$m_{0,i-1} = \gamma_{0,i-1} + \theta_{0,i-1}\frac{\sigma_s^2 v_{0,i-1}}{\sigma_s^2 + v_{0,i-1}} \text{ and } s^2_{0,i-1} = \delta_{0,i-1} + \theta^2_{0,i-1}\frac{\sigma_s^2 \underline{w}}{\sigma_s^2 + \underline{w}}.$$

Finally

$$f^2_{i,k-1}(x) = \left(1 + \frac{s^2_{0,i-1}}{\underline{w}}\right)\exp\left(-\frac{(x - v_{i,k-1})^2}{\underline{w}}\right)\exp\left(\frac{(m_{0,i-1} - v_{i,k-1})^2}{s^2_{0,i-1} + \underline{w}}\right). \tag{22}$$

Now we study

$$\left(\eta_{k|k-1:i} f(x) - \eta_{k|k-1:0} f\right)^2.$$

The distributions $\eta^{\delta_x}_{k|k-1:i}$ and $\eta_{k|k-1:0}$ are Gaussian with variances respectively given by $\delta_{i,k-1}$ and $s^2_{0,k-1}$. By (21), these variances belong to a compact interval $[k_1, k_2] \subset (0, +\infty)$. By Lemma 2 of the Appendix, we have

$$|\eta_{k|k-1:i} f(x) - \eta_{k|k-1:0} f| \leq K Z_{i,k} \tag{23}$$

with

$$Z_{i,k} = |\theta_{i,k-1}x - \theta_{0,k-1}\frac{\sigma_s^2 v_{0,k-1}}{\sigma_s^2 + \underline{w}} + \gamma_{i,k-1} - \gamma_{0,k-1}| + |\delta_{i,k-1} - \delta_{0,k-1} - \frac{\sigma_s^2 \underline{w}\theta^2_{0,k-1}}{\sigma_s^2 + \underline{w}}|, \tag{24}$$

and K is a constant depending on $\|f\|_\infty$ and k_1, k_2.

Notice that:

$$\begin{aligned}
\gamma_{i,k-1} - \gamma_{0,k-1} &= (1-\alpha)a\left(\sum_{l=0}^{k-i-1}(a\alpha)^l v_{k-1-l,k-1} - \sum_{l=0}^{k-1}(a\alpha)^l v_{k-1-l,k-1}\right) \\
&= (1-\alpha)(a\alpha)^{k-1-i}\sum_{l=k-1-i}^{k-1}(a\alpha)^{l-(k-1-i)} v_{k-1-l,k-1}
\end{aligned}$$

and

$$\begin{aligned}
\delta_{i,k-1} - \delta_{0,k-1} &= \beta^2 a\left(\sum_{l=0}^{k-i-1}(a\alpha)^{2l+1} - \sum_{l=0}^{k-1}(a\alpha)^{2l+1}\right) \\
&= \beta^2 a\sum_{l=k-i}^{k-1}(a\alpha)^{2l+1} \asymp (a\alpha)^{2(k-i)}
\end{aligned}$$

In addition, the quantity $\left(1 + \frac{s^2_{0,i-1}}{\underline{w}}\right)$ is uniformly bounded. We have

$$\underline{\Delta}_{k|k-1:0}(f) = \sum_{i=0}^{k}\underline{\Delta}_{i,k}$$

with

$$\underline{\Delta}_{i,k} = \eta_{i|i-1,0}(f_{i,k-1}^2(.)(\eta_{k|k-1,i}f(.) - \eta_{k|k-1,0}f)^2).$$

Now we use (18) and (22)-(24) and the above computations to get

$$\underline{\Delta}_{k|k-1:0}(f) \leq M \sum_{i=0}^{k} \alpha^i e^{B_{i,k}} A_{i,k}$$

where M is a constant depending on $\|f\|_\infty$ and $A_{i,k}, B_{i,k}$ are random variables depending on the observations $Y_{0:k}$ through $v_{i,j}$ and $\gamma_{i,j}$. Due to the stationarity, one can see that

$$\sup_{i \leq j} \mathbf{E}|v_{i,j}|^2 < \infty \text{ and } \sup_{i \leq j} \mathbf{E}|\gamma_{i,j}|^2 < \infty$$

where the expectation is taken with respect to the Y_k random variables. This allows to prove $\sup_{i \leq k} \mathbf{E}(A_{i,k}^2 + B_{i,k}^2) < \infty$. We conclude the proof with Lemma 1 (see appendix). Analogously, we prove the tightness of $\underline{\Gamma}_{k|k:0}(f)$.

6 Numerical simulations

6.1 Simulations based on the toy-example

We consider the example of 4.3. Consider the observations $Y_k = X_k + \varepsilon_k$ where $(\varepsilon_k)_k$ is a sequence of i.i.d. $\mathcal{N}(0, 0.2^2)$ random variables. In figure 2, we have plotted in plain line, with square marks, a trajectory of the hidden Markov chain, in longdashed line, with diamond marks, the noisy observations. We have plotted in dashed line, with plus marks, the result of the bootstrap particle filter associated to these observations, with $N = 500$ particles and $f(x) = x$. We observe that the result of the particle filter is close to the hidden chain, uniformly in time.

6.2 Simulations based on a Gaussian AR(1) model

Consider the observations given by

$$Y_i = X_i + \varepsilon_i.$$

where (ε_i) is a sequence of i.i.d. $\mathcal{N}(0, 0.5^2)$ random variables and $X_i = aX_{i-1} + \beta U_i, \quad a = 0.5, \beta = 0.5$. Figure 3 presents in plain line with square marks a trajectory of the hidden chain, in longdashed line, with diamond marks, the observations. We have plotted in dashed line, with plus marks, the result of the particle filter with $N = 500$ particles and $f(x) = x$. We also observe that the result of the particle filter is close to the hidden chain, uniformly in time.

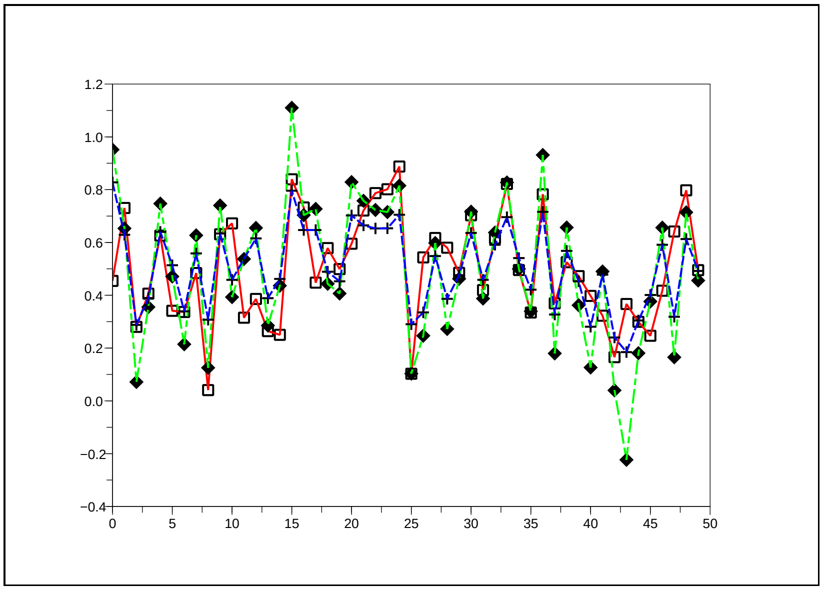

Figure 2: Toy-example ($\alpha = 0.4$). Hidden Markov chain (plain, square marks). Observations (longdashed line, diamond marks). Particle filter (dashed line, plus marks)

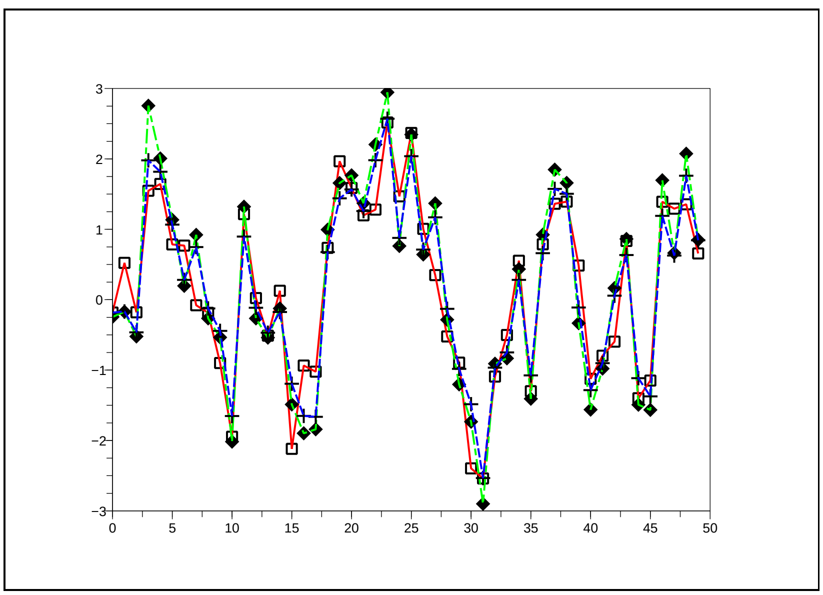

Figure 3: Kalman model. Hidden Markov chain (plain, square marks). Observations (longdashed line, diamond marks). Particle filter (dashed line, plus marks)

7 Acknowledgements

The author wishes to thank Pr. Genon-Catalot, my PhD advisor, for her help and the numerous discussions. The author wishes to thank Prs. Del Moral and Jacod for their advices and bibliographical references. The author wishes to thank Mahendra Mariadassou for discussions about the numerical examples.

References

Atar, R. and Zeitouni, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.* **33**, 697–725.

Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in Hidden Markov Models (Springer Series in Statistics).* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Chaleyat-Maurel, M. and Genon-Catalot, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Process. Appl.* **116**, 1447–1467.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32**, 2385–2411.

Davies, E. B. (1989). *Heat kernels and spectral theory*, vol. 92 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.

Del Moral, P. (2004). *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York. Genealogical and interacting particle systems with applications.

Del Moral, P. and Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.* **37**, 155–194.

Del Moral, P. and Jacod, J. (2001a). Interacting particle filtering with discrete observations. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pp. 43–75. Springer, New York.

Del Moral, P. and Jacod, J. (2001b). Interacting particle filtering with discrete-time observations: asymptotic behaviour in the Gaussian case. In *Stochastics in finite and infinite dimensions*, Trends Math., pp. 101–122. Birkhäuser Boston, Boston, MA.

Douc, R., Fort, G., Moulines, E. and Priouret, P. (2009). Forgetting of the initial distribution for hidden markov models. *Stochastic Process. Appl.* **119**, 1235–1256.

Douc, R., Guillin, A. and Najim, J. (2005). Moderate deviations for particle filtering. *Ann. Appl. Probab.* **15**, 587–614.

Doucet, A., de Freitas, N. and Gordon, N., eds. (2001). *Sequential Monte Carlo methods in practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York.

Genon-Catalot, V. and Kessler, M. (2004). Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter. *Bernoulli* **10**, 701–720.

Künsch, H. R. (2001). State space and hidden Markov models. In *Complex stochastic systems (Eindhoven, 1999)*, vol. 87 of *Monogr. Statist. Appl. Probab.*, pp. 109–173. Chapman & Hall/CRC, Boca Raton, FL.

Oudjane, N. and Rubenthaler, S. (2005). Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.* **23**, 421–448.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edn.

Van Handel, R. (2008). Uniform time average consistency of Monte Carlo particle filters. *http://arxiv.org/abs/0812.0350* .

8 Appendix

8.1 Bootstrap particle filter

The aim is to build a sequence of measures $(\eta^N_{k|k-1:0})_k$, where N is the number of interacting particles, so that $\eta^N_{k|k-1:0}f$ is a good approximation of $\eta_{k|k-1:0}f$ for f bounded. We assume that the distribution of X_0 is known and we set it as $\eta_0 = \eta_{0|-1:0}$. We assume that we are able to simulate random variables under η_0 and under $Q(x, dx')$.

Step 0: Simulate $(X_0^j)_{1\leq j\leq N}$ i.i.d. with distribution η_0 and compute $\eta^N_{0|-1:0} = \frac{1}{N}\sum_{j=1}^N \delta_{X_0^j}$.

Step 1-a: Simulate ${X'}_0^j$ i.i.d. with distribution $\pi^N_{0|0:0} = \sum_{j=1}^N \frac{g_0(X_0^j)}{\sum_{j=1}^N g_0(X_0^j)}\delta_{X_0^j}$.

Step 1-b: Simulate N random variables $(X_1^j)_j$ independantly with $X_1^j \sim Q({X'}_0^j, dx)$. Set $\eta^N_{1|0:0} = \frac{1}{N}\sum_{i=1}^N \delta_{X_1^j}$.

Step k-a: (updating) Suppose that $\eta^N_{k|k-1:0}$ is known. Simulate $(X^j_k)_{1\le j\le N}$ i.i.d. with distribution $\eta^N_{k|k-1:0}$ and simulate ${X'}^j_k$ i.i.d.with distribution $\pi^N_{k|k:0} = \sum_{j=1}^N \frac{g_k(X^j_k)}{\sum_{j=1}^N g_k(X^j_k)} \delta_{X^j_k}$.

Step k-b: (prediction) Simulate X^N_{k+1} independantly with $X^N_{k+1} \sim Q({X'}^j_k, dx)$. Set $\eta^N_{k+1|k:0} = \frac{1}{N}\sum_{j=1}^N \delta_{X^j_{k+1}}$.

8.2 Tightness lemma

The following lemma is proved (with $\delta = 1$) in Del Moral and Jacod (2001b).

Lemma 1. *(Tightness lemma) Let $\alpha \in (0,1)$ and consider two sequences $(A_{i,k})_{1\le i\le k}$ and $(B_{i,k})_{1\le i\le k}$ of non negative random variables such that*

$$\sup_{i,k} \mathbb{E}(A_{i,k}) + \sup_{i,k} \mathbb{E}(B^{1+\delta}_{i,k}) = K < \infty \tag{25}$$

then the sequence

$$\Upsilon_k = \sum_{i=1}^k \alpha^{k-i} A_{i,k} e^{B_{i,k}} \tag{26}$$

is tight.

Proof. Choose $\gamma > 1$ such that $\alpha\gamma < 1$. Set $\Omega_{j,k} = \cap_{i=1}^{k-j}\{|B_{i,k}| \le (k-i)\log\gamma\}$ for $1 \le j \le k$. Set also $\epsilon_j = \sum_{i\ge j} \frac{1}{i^{1+\delta}}$. Then

$$\mathbb{P}(\Omega^c_{j,k}) \le \frac{K}{(\log\gamma)^{1+\delta}} \sum_{i=1}^{k-j} \frac{1}{(k-i)^{1+\delta}} \le K\epsilon_j.$$

On $\Omega_{j,k}$ we have

$$\begin{aligned}
\sum_{i=1}^k \alpha^i A_{i,k} e^{B_{i,k}} &= \sum_{i=1}^{k-j} \alpha^{k-i} A_{i,k} e^{B_{i,k}} + \sum_{i=k-j+1}^{k} \alpha^{k-i} A_{i,k} e^{B_{i,k}} \\
&\le \sum_{i=1}^{k-j} (\gamma\alpha)^{k-i} A_{i,k} + \sum_{i=k-j+1}^{k} \alpha^{k-i} A_{i,k} e^{B_{i,k}}
\end{aligned}$$

Finally, we get for $1 \le j \le k$

$$\mathbb{P}(\Upsilon_k > M) \le K\epsilon_j + \frac{K}{M} + \sum_{i=k-j+1}^{k} \mathbb{P}(A_{i,k} e^{B_{i,k}} > \frac{M}{2j})$$

With our assumption, the sequence $(A_{i,k}e^{B_{i,k}})_{1\le i\le k}$ is tight. For $\epsilon > 0$ we first choose j then M, hence the result.

8.3 Inequality between Gaussian distributions

Lemma 2. *Let f a measurable function and $C > 0$ such that $\forall x \in \mathbb{R}, \quad |f(x)| \leq C$. Then for $m, m' \in \mathbb{R}$ and $\sigma, \sigma' \in [\epsilon, \frac{1}{\epsilon}]$ there is a constant $K > 0$ where K only depends on C and ε, such that*

$$\left|\mathcal{N}(m,\sigma^2)(f) - \mathcal{N}(m',\sigma'^2)(f)\right| \leq K\left(|m-m'| + |\sigma^2 - \sigma'^2|\right).$$

We start with the case $\sigma = \sigma'$:

$$\begin{aligned} & |\int_{\mathbb{R}} f(x)(\phi_{m,\sigma^2}(x) - \phi_{m',\sigma^2})dx| \\ \leq \quad & C\int_{\mathbb{R}}\left|\frac{1}{\sqrt{2\pi}\sigma}(\exp(-\frac{1}{2}(\frac{x-m}{\sigma})^2) - \exp(-\frac{1}{2}(\frac{x-m'}{\sigma})^2))\right|dx. \end{aligned}$$

With the changing of variables $x = \sigma z$

$$\begin{aligned} & \int_{\mathbb{R}}\left|\frac{1}{\sqrt{2\pi}\sigma}(\exp(-\frac{1}{2}(\frac{x-m}{\sigma})^2) - \exp(-\frac{1}{2}(\frac{x-m'}{\sigma})^2))\right|dx \\ = \quad & \int_{-\infty}^{\frac{m}{\sigma}+\frac{t}{2}}\frac{1}{\sqrt{2\pi}}\left(\exp(-\frac{1}{2}(x-\frac{m}{\sigma})^2) - \exp(-\frac{1}{2}(x-\frac{m}{\sigma}-t)^2)\right)dx \\ & + \int_{\frac{m}{\sigma}+\frac{t}{2}}^{+\infty}\frac{1}{\sqrt{2\pi}}\left(\exp(-\frac{1}{2}(x-\frac{m}{\sigma}-t)^2) - \exp(-\frac{1}{2}(x-\frac{m}{\sigma})^2)\right)dx \end{aligned}$$

where $t = \frac{m'}{\sigma} - \frac{m}{\sigma}$ and we assumed $m' > m$. With m and σ fixed, we consider the former quantity as a function of t: let $g(t) = g_1(t) + g_2(t)$, where $g_1(t) = G_1(t, \frac{m}{\sigma} + \frac{t}{2})$, with G_1 is defined by

$$G_1(t,y) = \int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}\left(\exp(-\frac{1}{2}(x-\frac{m}{\sigma})^2) - \exp(-\frac{1}{2}(x-\frac{m}{\sigma}-t)^2)\right)dx.$$

and with similar notations $g_2(t) = G_2(t, \frac{m}{\sigma} + \frac{t}{2})$. Then:

$$\frac{\partial G_1}{\partial y}(t, \frac{m}{\sigma} + \frac{t}{2}) = \frac{\partial G_2}{\partial y}(t, \frac{m}{\sigma} + \frac{t}{2}) = 0$$

and

$$g_1'(t) = \frac{\partial G_1}{\partial t}(t, \frac{m}{\sigma} + \frac{t}{2}) + \frac{1}{2}\frac{\partial G_1}{\partial y}(t, \frac{m}{\sigma} + \frac{t}{2}).$$

Differentiating the two members, it comes

$$\begin{aligned} |g'(t)| \quad & \leq \quad \int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}}|x-\frac{m}{\sigma}-t|\exp(-\frac{1}{2}(x-\frac{m}{\sigma}-t)^2)dx \\ & \leq \quad \int_{\mathbb{R}}|x|\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)dx \end{aligned}$$

where the last inequality holds for $t \in [0, \frac{m'-m}{\sigma}]$. By applying the mean value theorem to g on $[0, \frac{m'-m}{\sigma}]$ we have

$$|g(\frac{m'-m}{\sigma}) - g(0)| \leq \left(\int_{\mathbb{R}} |x| \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) dx \right) |\frac{m'-m}{\sigma}|$$

and then

$$|\mathcal{N}(m, \sigma^2)(f) - \mathcal{N}(m', \sigma^2)(f)| \leq K_\epsilon |m - m'|.$$

Analogous computations work when $m = m'$. With $s = \frac{1}{\sigma}$ et $s' = \frac{1}{\sigma'}$, we have:

$$\begin{aligned} & \left| \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} \left(s \exp(-\frac{1}{2}s^2(x-m)^2) - s' \exp(-\frac{1}{2}s'^2(x-m)^2) \right) dx \right| \\ \leq \; & C \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \left| \left(s \exp(-\frac{1}{2}s^2(x-m)^2) - s' \exp(-\frac{1}{2}s'^2(x-m)^2) \right) \right| dx \\ \leq \; & C \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \left| \left(s \exp(-\frac{1}{2}s^2x^2) - st \exp(-\frac{1}{2}s^2t^2x^2) \right) \right| dx \end{aligned}$$

with $t = \frac{s'}{s}$, assuming $s' > s$. The sign changes occur at:

$$x = \pm \sqrt{\frac{2 \log t}{s^2(t^2-1)}}.$$

Thus, we have

$$\begin{aligned} & \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \left| \left(s \exp(-\frac{1}{2}s^2x^2) - st \exp(-\frac{1}{2}s^2t^2x^2) \right) \right| dx \\ \leq \; & \int_{-\infty}^{-\sqrt{\frac{2\log t}{s^2(t^2-1)}}} \frac{1}{\sqrt{2\pi}} \left(s \exp(-\frac{1}{2}s^2x^2) - st \exp(-\frac{1}{2}s^2t^2x^2) \right) dx \\ & + \int_{-\sqrt{\frac{2\log t}{s^2(t^2-1)}}}^{+\sqrt{\frac{2\log t}{s^2(t^2-1)}}} \frac{1}{\sqrt{2\pi}} \left(st \exp(-\frac{1}{2}s^2t^2x^2) - s \exp(-\frac{1}{2}s^2x^2) - \right) dx \\ & + \int_{+\sqrt{\frac{2\log t}{s^2(t^2-1)}}}^{+\infty} \frac{1}{\sqrt{2\pi}} \left(s \exp(-\frac{1}{2}s^2x^2) - st \exp(-\frac{1}{2}s^2t^2x^2) \right) dx. \end{aligned}$$

Proceeding as above, we derive:

$$\left| \mathcal{N}(m, \sigma^2)(f) - \mathcal{N}(m, \sigma'^2)(f) \right| \leq K_\epsilon |\sigma^2 - \sigma'^2|$$

remarking that $\sigma, \sigma^2 \in [\epsilon, \frac{1}{\epsilon}]$ and $|\frac{1}{\sigma} - \frac{1}{\sigma'}| \leq C_\epsilon |\sigma^2 - \sigma'^2|$.