



**HAL**  
open science

# Non-asymptotic deviation inequalities for smoothed additive functionals in non-linear state-space models with applications to parameter estimation

Cyrille Duguay, Sylvain Le Corff

► **To cite this version:**

Cyrille Duguay, Sylvain Le Corff. Non-asymptotic deviation inequalities for smoothed additive functionals in non-linear state-space models with applications to parameter estimation. 2010. hal-00548092v1

**HAL Id: hal-00548092**

**<https://hal.science/hal-00548092v1>**

Preprint submitted on 18 Dec 2010 (v1), last revised 25 Apr 2012 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Non-asymptotic deviation inequalities for smoothed additive functionals in non-linear state-space models with applications to parameter estimation

Cyrille Dufour\* and Sylvain Le Corff†

December 19, 2010

## Abstract

Approximating joint smoothing distributions using particle-based methods is a well-known issue in statistical inference when operating on general state space hidden Markov models (HMM). In this paper we focus on non-asymptotic bounds for the error generated by the computation of smoothed additive functionals. More precisely, this contribution provides new results on the forward filtering backward smoothing (FFBS)  $L_q$ -mean errors under appropriate mixing conditions on the Markov kernel's probability density function. The algorithm used has a computational complexity depending linearly on  $TN$  where  $T$  is the number of observations and  $N$  the number of particles. The main improvement concerns the rate of convergence of these norms which depends on  $T$  and  $N$  only through the ratio  $T/N$  for additive functionals (i.e. with norm proportional to  $T$ ). This paper relies mainly on recent exponential deviation inequalities on the smoothing error.

## 1 Introduction

State-space representations play a key role to model non-linear non-Gaussian time series in statistics, information engineering and financial econometrics; see [3, 16, 26]. Let  $\mathbb{X}$  be a general-state space endowed with a countably generated  $\sigma$ -field  $\mathcal{B}(\mathbb{X})$ . Let  $(M_\theta, \theta \in \Theta)$  be a family of transition kernels on  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$  depending on some (finite or infinite dimensional) parameter  $\theta$ . For  $\theta \in \Theta$ , consider a Markov chain  $\{X_t^\theta\}_{t \geq 0}$  on  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$  with initial distribution  $\chi$  and transition kernel  $M_\theta$ . The transition kernels are assumed to be dominated: there exists a probability measure  $\lambda$  on  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$  such that for all  $\theta \in \Theta$ ,  $x \in \mathbb{X}$ ,  $M_\theta(x, \cdot) \ll \lambda(\cdot)$ . We denote the transition density function  $m_\theta(x, \cdot)$ . The observed time series  $\{Y_t\}_{t \geq 0}$  takes values in a space  $\mathbb{Y}$  endowed with a countably generated  $\sigma$ -field  $\mathcal{B}(\mathbb{Y})$ . It is assumed that the observations  $\{Y_t\}_{t \geq 0}$  are independent conditionally to  $\{X_t\}_{t \geq 0}$  and that the conditional distribution given  $\{X_t\}_{t \geq 0}$  depends only on  $X_t$ . It is also assumed that there exist a probability measure  $\nu$  on  $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$  and a family of probability densities on  $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ ,  $(g_\theta, \theta \in \Theta)$ , such that for any bounded and measurable function  $h$  on  $\mathbb{Y}$  and any  $t \in \mathbb{N}$ :

$$\mathbb{E}_\theta [h(Y_t)|X_t] = \int h(y)g_\theta(X_t, y)\nu(dy) .$$

---

\*Département CITI, Institut Télécom/Télécom SudParis, France

†Département TSI, Institut Télécom/Télécom ParisTech, France

For  $T + 1$  observations,  $T \in \mathbb{N}^*$  and  $(s, t) \in \mathbb{N}^2, s \leq t \leq T$ , the joint smoothing distribution is defined, for all bounded and measurable function  $h$  on  $\mathbb{X}^{t-s+1}$ , by

$$\begin{aligned} \phi_{s:t|T}^\theta[h] &\stackrel{\text{def}}{=} \mathbb{E}_\theta [h(X_{s:t})|Y_{0:T}] \\ &= \frac{\int \chi(\mathrm{d}x_0)g_\theta(x_0, Y_0) \prod_{s=1}^T M_\theta(x_{s-1}, \mathrm{d}x_s)g_\theta(x_s, Y_s)h(x_{s:t})}{\int \chi(\mathrm{d}x_0)g_\theta(x_0, Y_0) \prod_{s=1}^T M_\theta(x_{s-1}, \mathrm{d}x_s)g_\theta(x_s, Y_s)}, \end{aligned}$$

where  $a_{u:v}$  is a short-hand notation for  $(a_u, \dots, a_v)$ . We denote by  $\phi_s^\theta \stackrel{\text{def}}{=} \phi_{s:s|s}^\theta$  the filtering distribution and by  $\phi_{s|T}^\theta \stackrel{\text{def}}{=} \phi_{s:s|T}^\theta$  the marginal smoothing distribution.

For any  $T \in \mathbb{N}^*$ , the log-likelihood of the observations  $Y_{0:T}$  may be written as

$$\theta \mapsto \ell(Y_{0:T}, \theta) = \log \left[ \int \chi(\mathrm{d}x_0)g_\theta(x_0, Y_0) \prod_{t=1}^T M_\theta(x_{t-1}, \mathrm{d}x_t)g_\theta(x_t, Y_t) \right].$$

Parameter estimates can be obtained by maximizing the log-likelihood function with respect to the parameter  $\theta \in \Theta$  using either a gradient ascent algorithm or the Expectation-Maximization (EM) algorithm introduced in [8]. Under appropriate regularity conditions, the score function can be computed using the so-called Fisher identity:

$$\nabla_\theta \ell(Y_{0:T}, \theta) = \sum_{t=0}^T \phi_{t|T}^\theta [\nabla_\theta \log g_\theta(\cdot, Y_t)] + \sum_{t=1}^T \phi_{t-1:t|T}^\theta [\nabla_\theta \log m_\theta(\cdot, \cdot)]. \quad (1)$$

The estimation of the score function therefore requires the computation of a path-space smoothed additive functional. Similarly, the E-step of the EM algorithm requires the computation of the conditional expectation of the complete-data log-likelihood defined, for any  $(\theta, \theta') \in \Theta^2$ , by:

$$Q(\theta, \theta') = \sum_{t=0}^T \phi_{t|T}^{\theta'} [\log g_\theta(\cdot, Y_t)] + \sum_{t=1}^T \phi_{t-1:t|T}^{\theta'} [\log m_\theta(\cdot, \cdot)]. \quad (2)$$

The computation of (1) and (2) may be seen as a specific example of estimation of a smoothed additive functional

$$\phi_{0:T|T}(S_T) = \sum_{t=1}^T \phi_{t-1:t|T}(h_t), \quad (3)$$

where for a family of bounded measurable functions  $\{h_t, t \geq 0\}$ ,  $h_t : \mathbb{X}^2 \rightarrow \mathbb{R}$ ,  $S_T$  is defined on the path-space  $\mathbb{X}^{T+1}$ , for any  $x_{0:T} \in \mathbb{X}^{T+1}$ , by

$$S_T(x_{0:T}) = \sum_{t=1}^T h_t(x_{t-1}, x_t). \quad (4)$$

Except for simple models such as linear Gaussian state-space models or when  $\mathbb{X}$  is a finite set, such smoothed additive functionals cannot be computed explicitly.

In this paper, we consider Sequential Monte Carlo algorithms, henceforth referred to as particle methods, to approximate these quantities. These methods combine sequential importance sampling and sampling importance resampling steps to produce a set of random particles with associated importance weights to approximate the fixed interval smoothing distributions and to estimate smoothed additive functionals defined in (3).

Implementations of such procedures are detailed in [1, 3, 10, 12]; the convergence of such approximations is addressed in [4, 7].

The most straightforward implementation of fixed-interval smoothing estimation is based on the so-called genealogical tree. The complexity of this algorithm per time-step grows only linearly with the number  $N$  of particles. However, as shown in [14], even under favorable mixing conditions on the kernel  $M_\theta$ , the variance of the estimate grows at least quadratically with the total number of observations  $T$ . Heuristically, the associated particle approximations become inaccurate when  $T$  goes to infinity since the genealogical tree estimates become progressively impoverished from the successive resampling steps: the number of distinct particles representing  $\phi_{t:t+\ell|T}$  for any  $t \ll T$  diminishes as  $T - t$  increases. Hence, whatever the number  $N$  of particles is,  $\phi_{t:t+\ell|T}$  will eventually be approximated by a single path when  $T$  becomes large. This problem is known in the literature as the path degeneracy issue; see [2, 17] for a discussion.

Several solutions have been proposed to solve this degeneracy problem. In this paper we consider the forward filtering backward smoothing algorithm (FFBS) and the forward filtering backward simulation algorithm (FFBSi) introduced in [11] and further developed in [17]. These can be seen as extrapolations to the non-linear hidden Markov model context of the Rauch-Tung-Striebel algorithm introduced for linear gaussian state-space. As shown in [5], when applied to additive functionals, the FFBS algorithm can be implemented forward in time, but its complexity grows quadratically with the number of particles. This algorithm has been applied in [14] to approximate the score function. The FFBSi algorithm uses a backward simulation step, and is thus mainly adapted for batch estimation problem: as shown in [9], it is possible to implement the FFBSi with a complexity growing only linearly with the number of particles.

It is conjectured in [14] that the variance of the FFBS score estimator grows only linearly with the number of observations  $T$  (instead of quadratically) under strong mixing conditions for the kernel  $M_\theta$  (which guarantee the uniform convergence of the filtering distribution). In an unpublished paper by [5], a more general result is shown: the FFBS estimator variance of any smoothed additive functional is upper bounded by terms depending on  $T$  and  $N$  only through the ratio  $T/N$ . Finally, in [5], for any  $q > 2$ , a  $L_q$ -mean error bound for smoothed additive functional is established. When applied to strongly mixing kernels, this bound amounts to be of order  $T/\sqrt{N}$  (see [5, Eq. (3.8), pp. 957]).

In this paper, we establish  $L_q$ -mean error and exponential deviation inequalities of both the FFBS and FFBSi smoothed additive functionals estimators. We show that, for any  $q \geq 2$ , the  $L_q$ -mean error for both algorithms is upper bounded by terms depending on  $T$  and  $N$  only through the ratio  $T/N$  under the strong mixing conditions. We also establish an exponential deviation inequality with the same functional dependence in  $T$  and  $N$ .

This paper is organized as follows. Section 2 introduces further definitions and notations and details the FFBS and FFBSi algorithms. In Section 3, upper bounds for the  $L_q$ -mean error and exponential deviation inequalities of these two algorithms are presented. In Section 4, some Monte Carlo experiments are presented to support our theoretical claims. The proofs are presented in Sections 5 and 6.

## 2 Framework

To keep the notation simple, the dependence on the parameter  $\theta$  is omitted and we define  $g_t(x) = g(x, Y_t)$ . In its original version, the FFBS algorithm work in two passes. In the forward pass the filtering distributions  $(\phi_t)_{t \in \{0, \dots, T\}}$  are approximated using weighted samples  $(\omega_t^{N, \ell}, \xi_t^{N, \ell})_{\ell \in \{1, \dots, N\}}$ , where  $T$  is the number of observations and  $N$  the num-

ber of particles: all sampled particles and weights are stored. In the backward pass of the FFBS, these importance weights are then modified (see [11, 21, 23]) while the particle positions are kept fixed. The importance weights are updated recursively backward in time to obtain an approximation of the fixed interval smoothing distributions  $(\phi_{s:T|T})_{0 \leq s \leq T}$ . The particle approximation is constructed as follows.

**Forward pass** Let  $\{\xi_0^{N,i}\}_{i=1}^N$  be i.i.d. random variables distributed according to the instrumental density  $\rho_0$  (which we denote  $\xi_0^{N,i} \sim \rho_0$ ) and set the unnormalized importance weights  $\omega_0^{N,i} \stackrel{\text{def}}{=} d\chi/d\rho_0(\xi_0^{N,i}) g_0(\xi_0^{N,i})$ . The weighted sample  $\{(\xi_0^{N,i}, \omega_0^{N,i})\}_{i=1}^N$  then targets the initial filter  $\phi_0$  in the sense that  $\phi_0^N(h) \stackrel{\text{def}}{=} \sum_{\ell=1}^N \omega_0^{N,\ell} h(\xi_0^{N,\ell}) / \sum_{\ell=1}^N \omega_0^{N,\ell}$  is a consistent estimator of  $\phi_0(h)$  for  $h \in \mathcal{F}_b(\mathbb{X})$ .

Let now  $\{(\xi_{s-1}^{N,i}, \omega_{s-1}^{N,i})\}_{i=1}^N$  be a weighted sample targeting  $\phi_{s-1}$ , we aim at computing new particles and importance weights targeting the probability distribution  $\phi_s^{N,t}$  defined, for any bounded measurable function  $h$  on  $\mathbb{X}$ , as

$$\phi_s^{N,t}(h) = \frac{\phi_{s-1}^N [Mg_s h]}{\phi_{s-1}^N [Mg_s]}, \quad (5)$$

in order to produce an updated particle sample approximating  $\phi_s$ . Following [24], this may be done by considering the auxiliary target distribution

$$\phi_s^{N,a}(i, h) \stackrel{\text{def}}{=} \frac{\omega_{s-1}^{N,i} M g_s h(\xi_{s-1}^{N,i})}{\sum_{\ell=1}^N \omega_{s-1}^{N,\ell} M g_s(\xi_{s-1}^{N,\ell})}, \quad (6)$$

on the product space  $(\{1, \dots, N\} \times \mathbb{X}, \mathcal{P}(\{1, \dots, N\}) \otimes \mathcal{B}(\mathbb{X}))$ . By construction,  $\phi_s^{N,t}$  is the marginal distribution of  $\phi_s^{N,a}$  with respect to the particle index. The target distribution  $\phi_s^{N,t}$  is approximated by simulating pairs  $\{(I_s^{N,i}, \xi_s^{N,i})\}_{i=1}^N$  of indices and particles from the instrumental distribution:

$$\pi_{s|s}(i, h) \propto \omega_{s-1}^{N,i} \vartheta_s(\xi_{s-1}^{N,i}) P_s(\xi_{s-1}^{N,i}, h), \quad (7)$$

on the product space  $\{1, \dots, N\} \times \mathbb{X}$ , where  $\{\vartheta_s(\xi_{s-1}^{N,i})\}_{i=1}^N$  are the adjustment multiplier weights and  $P_s$  is a Markovian proposal transition kernel. In the sequel, we assume for simplicity that  $P_s(x, \cdot)$  has, for any  $x \in \mathbb{X}$ , a density  $p_s(x, \cdot)$  with respect to the reference measure  $\lambda$ . For any  $i = 1, \dots, N$  we associate to the particle  $\xi_s^{N,i}$  its importance weight defined by:

$$\omega_s^{N,i} \stackrel{\text{def}}{=} \frac{m(\xi_{s-1}^{N,I_s^{N,i}}, \xi_s^{N,i}) g_s(\xi_s^{N,i})}{\vartheta_s(\xi_{s-1}^{N,I_s^{N,i}}) p_s(\xi_{s-1}^{N,I_s^{N,i}}, \xi_s^{N,i})}. \quad (8)$$

Finally, the indices  $\{I_s^{N,i}\}_{i=1}^N$  are discarded and  $\{(\xi_s^{N,i}, \omega_s^{N,i})\}_{i=1}^N$  is taken as an approximation of  $\phi_s$ . Setting for all  $x \in \mathbb{X}$ ,  $\vartheta_s(x) \equiv 1$  and  $p_s(x, \cdot) \equiv m(x, \cdot)$  leads to the bootstrap particle filter algorithm proposed by [18].

**Backward smoothing** For any probability measure  $\eta$  on  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ , denote by  $B_\eta$  the so-called backward smoothing kernel given, for all bounded measurable function  $h$  on  $\mathbb{X}$  and for all  $x \in \mathbb{X}$ , by:

$$B_\eta(x, h) \stackrel{\text{def}}{=} \frac{\int \eta(dx') m(x', x) h(x')}{\int \eta(dx') m(x', x)}, \quad (9)$$

For all  $0 \leq s < T$  and for all bounded measurable function  $h$  on  $\mathbb{X}^{T-s+1}$ ,  $\phi_{s:T|T}(h)$  may be computed recursively, backward in time, according to

$$\phi_{s:T|T}(h) = \int \cdots \int B_{\phi_s}(x_{s+1}, dx_s) \phi_{s+1:T|T}(dx_{s+1:T}) h(x_{s:T}), \quad (10)$$

## 2.1 The forward filtering backward smoothing algorithm

From the weighted samples  $\{(\xi_t^{N,i}, \omega_t^{N,i})\}_{i=1}^N$ ,  $1 \leq t \leq T$  drawn in the forward pass, an approximation of the joint smoothing distribution can be obtained using

$$\phi_{s:T|T}^N(h) = \int \cdots \int B_{\phi_s^N}(x_{s+1}, dx_s) \phi_{s+1:T|T}^N(dx_{s+1:T}) h(x_{s:T}), \quad (11)$$

and starting with  $\phi_{T:T|T}^N(h) = \phi_T^N(h)$ . Now, by definition, for all  $x \in \mathbb{X}$  and for all bounded measurable function  $h$  on  $\mathbb{X}$ ,

$$B_{\phi_s^N}(x, h) = \sum_{i=1}^N \frac{\omega_s^{N,i} m(\xi_s^{N,i}, x)}{\sum_{\ell=1}^N \omega_s^{N,\ell} m(\xi_s^{N,\ell}, x)} h(\xi_s^{N,i}),$$

and inserting this expression into (11) gives the following particle approximation of the fixed-interval smoothing distribution  $\phi_{0:T|T}(h)$

$$\begin{aligned} \phi_{0:T|T}^N(h) &= \sum_{i_0=1}^N \cdots \sum_{i_T=1}^N \left( \prod_{u=1}^T \frac{\omega_{u-1}^{N,i_{u-1}} m(\xi_{u-1}^{N,i_{u-1}}, \xi_u^{N,i_u})}{\sum_{\ell=1}^N \omega_{u-1}^{N,\ell} m(\xi_{u-1}^{N,\ell}, \xi_u^{N,i_u})} \right) \\ &\quad \times \frac{\omega_T^{N,i_T}}{\Omega_T^N} h(\xi_0^{N,i_0}, \dots, \xi_T^{N,i_T}), \end{aligned} \quad (12)$$

where  $h$  is a bounded measurable function on  $\mathbb{X}^{T+1}$  and

$$\Omega_t^N \stackrel{\text{def}}{=} \sum_{i=1}^N \omega_t^{N,i}. \quad (13)$$

The estimator of the fixed interval smoothing distribution  $\phi_{0:T|T}^N$  might seem impractical since the cardinality of its support is  $N^{T+1}$ . Nevertheless, the FFBS algorithm might be used to compute the fixed interval smoothing distributions of additive functionals of the form (4). In addition, perhaps surprisingly, as shown in [5], FFBS estimator of an additive functional can be computed forward in time.

Let  $\{R_t\}_{t \geq 1}$  be the family of Markov kernels on  $\mathbb{X} \times \mathcal{B}(\mathbb{X})^{\otimes(t+1)}$  defined, for any  $t \geq 1$ , any  $x_t \in \mathbb{X}$  and any measurable function  $h$  on  $\mathbb{X}^{t+1}$ , by

$$R_t(x_t, h) = \int B_{\phi_{t-1}}(x_t, dx_{t-1}) \cdots B_{\phi_0}(x_1, dx_0) h(x_{0:t}).$$

For any  $t \geq 1$ , by definition of the joint smoothing distribution  $\phi_{0:t|t}$  we have, for any measurable function  $h$  defined on  $\mathbb{X}^{t+1}$ ,  $\phi_t R_t(h) = \phi_{0:t|t}(h)$ . For an additive functional  $S_T$  of the form (4), the function  $R_T(\cdot, S_T)$  can be computed forward in time as follows

$$\begin{aligned} R_T(x_T, S_T) &= \int B_{\phi_{T-1}}(x_T, dx_{T-1}) \cdots B_{\phi_0}(x_1, dx_0) S_T(x_{0:T}) \\ &= \int B_{\phi_{T-1}}(x_T, dx_{T-1}) \cdots B_{\phi_0}(x_1, dx_0) S_{T-1}(x_{0:T-1}) \\ &\quad + \int B_{\phi_{T-1}}(x_T, dx_{T-1}) h_T(x_{T-1}, x_T) \\ &= \int B_{\phi_{T-1}}(x_T, dx_{T-1}) R_{T-1}(x_{T-1}, S_{T-1}) \\ &\quad + \int B_{\phi_{T-1}}(x_T, dx_{T-1}) h_T(x_{T-1}, x_T). \end{aligned}$$

The particle approximation of this forward recursion may be computed by replacing  $\phi_{T-1}$  in the last equation by its particle approximation  $\phi_{T-1}^N$ . For any  $1 \leq i \leq N$ ,

$$\mathbf{R}_T^N(\xi_T^{N,i}, S_T) = \sum_{j=1}^N \frac{\omega_{T-1}^{N,j} m(\xi_{T-1}^{N,j}, \xi_T^{N,i})}{\sum_{\ell=1}^N \omega_{T-1}^{N,\ell} m(\xi_{T-1}^{N,\ell}, \xi_T^{N,i})} \times \left[ \mathbf{R}_{T-1}^N(\xi_{T-1}^{N,j}, S_{T-1}) + h_T(\xi_{T-1}^{N,j}, \xi_T^{N,i}) \right],$$

and then, the forward-backward recursion defined by (12) can be computed in forward direction only.

**Example 1: The Fisher score** Going back to equation (1), the computation of the Fisher score relies on an additive function of the form (4) where for any  $1 \leq t \leq T$ ,  $h_t(x_{t-1}, x_t) = \nabla_{\theta} \log m_{\theta}(x_{t-1}, x_t) + \nabla_{\theta} \log g_{\theta}(x_t, Y_t)$ , leading to Algorithm 1. This algo-

---

**Algorithm 1** Computation of the Fisher score

---

**Require:** Parameter  $\theta$  and observations  $Y_{0:T}$

**Ensure:** Estimate of the Fisher score  $\nabla_{\theta} \ell^N(Y_{0:T}, \theta)$

**INITIALIZATION**

Simulate  $(\xi_0^{N,i}, \omega_0^{N,i})_{i=1}^N$  under  $\theta$

For all  $1 \leq i \leq N$ , compute

$$\mathbf{R}_0^N(\xi_0^{N,i}, S_0) = \nabla_{\theta} \log g_{\theta}(\xi_0^{N,i}, Y_0)$$

**FORWARD RECURSION**

**for**  $t = 1$  to  $T$  **do**

Simulate  $(\xi_t^{N,i}, \omega_t^{N,i})_{i=1}^N$  under  $\theta$

For all  $1 \leq i \leq N$ , compute

$$\begin{aligned} \mathbf{R}_t^N(\xi_t^{N,i}, S_t) &= \sum_{j=1}^N \frac{\omega_{t-1}^{N,j} m_{\theta}(\xi_{t-1}^{N,j}, \xi_t^{N,i})}{\sum_{\ell=1}^N \omega_{t-1}^{N,\ell} m_{\theta}(\xi_{t-1}^{N,\ell}, \xi_t^{N,i})} \\ &\times \left[ \mathbf{R}_{t-1}^N(\xi_{t-1}^{N,j}, S_{t-1}) + \nabla_{\theta} \log m_{\theta}(\xi_{t-1}^{N,j}, \xi_t^{N,i}) + \nabla_{\theta} \log g_{\theta}(\xi_t^{N,i}, Y_t) \right] \end{aligned}$$

**end for**

Set

$$\nabla_{\theta} \ell^N(Y_{0:T}, \theta) = \sum_{i=1}^N \frac{\omega_T^{N,i}}{\Omega_T^N} \mathbf{R}_T^N(\xi_T^{N,i}, S_T)$$


---

rithm is exactly the one presented in [14] as an alternative to the use of the genealogical tree method. Therefore, the results outlined in Section 3 hold for this method and confirm the conjecture mentioned in [14].

**Example 2: The EM algorithm** In the same way equation (2) can be understood as an instance of the same problem where for any  $1 \leq t \leq T$ ,  $h_t(x_{t-1}, x_t) = \log m_\theta(x_{t-1}, x_t) + \log g_\theta(x_t, Y_t)$ , leading to Algorithm 2.

---

**Algorithm 2** Computation of the intermediate quantity of the EM algorithm

---

**Require:** Parameters  $\theta$  and  $\theta'$  and observations  $Y_{0:T}$

**Ensure:** Estimate of the intermediate quantity  $Q^N(\theta, \theta')$

**INITIALIZATION**

Simulate  $(\xi_0^{N,i}, \omega_0^{N,i})_{i=1}^N$  under  $\theta'$

For all  $1 \leq i \leq N$ , compute

$$R_0^N(\xi_0^{N,i}, S_0) = \log g_\theta(\xi_0^{N,i}, Y_0)$$

**FORWARD RECURSION**

**for**  $t = 1$  to  $T$  **do**

Simulate  $(\xi_t^{N,i}, \omega_t^{N,i})_{i=1}^N$  under  $\theta'$

For all  $1 \leq i \leq N$ , compute

$$\begin{aligned} R_t^N(\xi_t^{N,i}, S_t) &= \sum_{j=1}^N \frac{\omega_{t-1}^{N,j} m_{\theta'}(\xi_{t-1}^{N,j}, \xi_t^{N,i})}{\sum_{\ell=1}^N \omega_{t-1}^{N,\ell} m_{\theta'}(\xi_{t-1}^{N,\ell}, \xi_t^{N,i})} \\ &\quad \times \left[ R_{t-1}^N(\xi_{t-1}^{N,j}, S_{t-1}) + \log m_\theta(\xi_{t-1}^{N,j}, \xi_t^{N,i}) + \log g_\theta(\xi_t^{N,i}, Y_t) \right] \end{aligned}$$

**end for**

Set

$$Q^N(\theta, \theta') = \sum_{i=1}^N \frac{\omega_T^{N,i}}{\Omega_T^N} R_T^N(\xi_T^{N,i}, S_T)$$


---

## 2.2 The forward filtering backward simulation algorithm

The FFBS algorithm both in the forward and backward interpretations outlined in Subsection 2.1 gives an approximation of smoothed additive functionals with a computational cost growing quadratically with the number  $N$  of particles. In many instances this complexity is prohibitive and the FFBS algorithm cannot be used. However, the estimator (12) may be understood alternatively by noting that the normalized smoothing weights define a probability distribution on the set  $\{1, \dots, N\}^{T+1}$  of trajectories associated with an inhomogeneous Markov chain. This interpretation leads to an algorithm whose complexity grows only linearly with the number of particles but the backward pass cannot be bypassed. Indeed, consider, for  $t \in \{0, \dots, T-1\}$ , the Markov transition matrix  $\{\Lambda_t^N(i, j)\}_{i,j=1}^N$  over the state-space  $\{1, \dots, N\}$  given by

$$\Lambda_t^N(i, j) = \frac{\omega_t^{N,j} m(\xi_t^{N,j}, \xi_{t+1}^{N,i})}{\sum_{\ell=1}^N \omega_t^{N,\ell} m(\xi_t^{N,\ell}, \xi_{t+1}^{N,i})}, \quad (i, j) \in \{1, \dots, N\}^2. \quad (14)$$

For  $1 \leq t \leq T$ , we define

$$\mathcal{F}_t^N \stackrel{\text{def}}{=} \sigma \{Y_{0:T}, (\xi_s^{N,i}, \omega_s^{N,i}); 0 \leq s \leq t, 1 \leq i \leq N\}. \quad (15)$$

The transition probabilities defined in (14) induce an inhomogeneous Markov chain  $\{J_u\}_{u=0}^T$  evolving backward in time as follows. At time  $T$ , the random index  $J_T$  is drawn



from the set  $\{1, \dots, N\}$  such that  $J_T$  takes the value  $i$  with a probability proportional to  $\omega_T^{N,i}$ . For any time  $0 \leq t \leq T-1$  the index  $J_t$  is sampled in the set  $\{1, \dots, N\}$  according to  $\Lambda_t^N(J_{t+1}, \cdot)$ . The joint distribution of  $J_{0:T}$  is therefore given by, for  $j_{0:T} \in \{1, \dots, N\}^{T+1}$ ,

$$\mathbb{P} [J_{0:T} = j_{0:T} | \mathcal{F}_T^N] = \frac{\omega_T^{N,j_T}}{\Omega_T^N} \Lambda_{T-1}^N(j_T, j_{T-1}) \dots \Lambda_0^N(j_1, j_0). \quad (16)$$

Thus, and this is a key observation, the FFBS estimator (12) of the joint smoothing distribution may be written as the conditional expectation

$$\phi_{0:T|T}^N(h) = \mathbb{E} \left[ h \left( \xi_0^{N,J_0}, \dots, \xi_T^{N,J_T} \right) \middle| \mathcal{F}_T^N \right], \quad (17)$$

where  $h$  is a bounded measurable function on  $\mathbb{X}^{T+1}$ . We may therefore construct an unbiased estimator of the FFBS estimator given by

$$\tilde{\phi}_{0:T|T}^N(h) = N^{-1} \sum_{\ell=1}^N h \left( \xi_0^{N,J_0^\ell}, \dots, \xi_T^{N,J_T^\ell} \right), \quad (18)$$

where  $\{J_{0:T}^\ell\}_{\ell=1}^N$  are  $N$  paths drawn independently given  $\mathcal{F}_T^N$  according to (16) and where  $h$  is a bounded measurable function on  $\mathbb{X}^{T+1}$ . This practical estimator was introduced in [17] (Algorithm 1, p. 158). The computational complexity for sampling a single path of  $J_{0:T}$  is linear in the number of particles; therefore, the overall computational effort spent when estimating  $\tilde{\phi}_{0:T|T}^N$  using the FFBSi sampler grows with  $N^2$ . Following [9] this complexity can be reduced so that it grows only linearly with  $N$  using a specific form of acceptance-rejection method.

### 3 Main results

In this Section, the  $L_q$ -mean error bounds and exponential deviation inequalities of the FFBS and FFBSi algorithms are established for additive functionals under the following assumptions.

- (A1)** (i) For all  $t \geq 0$  and all  $x \in \mathbb{X}$ ,  $g_t(x) > 0$ .  
(ii)  $\sup_{t \geq 0} |g_t|_\infty < \infty$ ,  $\sup_{t \geq 1} |\vartheta_t|_\infty < \infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty < \infty$  where

$$\omega_0(x) \stackrel{\text{def}}{=} \frac{d\chi}{d\rho_0}(x)g_0(x), \quad \omega_t(x, x') \stackrel{\text{def}}{=} \frac{m(x, x')g_t(x')}{\vartheta_t(x)p_t(x, x')}, \quad \forall t \geq 1.$$

- (A2)** (i)  $|m|_\infty < \infty$  and  $\sup_{t \geq 0} |p_t|_\infty < \infty$ .

- (A3)** (i) There exists  $(\sigma_-, \sigma_+) \in (0, \infty)^2$  such that  $\sigma_- < \sigma_+$  and for any  $(x, x') \in \mathbb{X}^2$ ,  $\sigma_- \leq m(x, x') \leq \sigma_+$ .  
(ii) There exists  $c_- \in \mathbb{R}_+^*$  such that  $\int \chi(dx)g_0(x) \geq c_-$  and for any  $t \in \mathbb{N}^*$ ,  $\inf_{x \in \mathbb{X}} \int M(x, dx')g_t(x') \geq c_-$ .

The last assumption, referred to as the strong mixing condition, is crucial to derive time-uniform exponential deviation inequalities and a time-uniform bound of the variance of the marginal smoothing distribution (see [6] and [9]).

For all function  $h$  from a space  $\mathbb{E}$  to  $\mathbb{R}$ ,  $\text{osc}(h)$  is defined by:

$$\text{osc}(h) \stackrel{\text{def}}{=} \sup_{(z, z') \in \mathbb{E}^2} |h(z) - h(z')|.$$

For sake of simplicity, our results are established for additive functionals of the form

$$S_T(x_{0:T}) = \sum_{t=0}^T h_t(x_t), \quad (19)$$

where  $\{h_t; 0 \leq t \leq T\}$  is a family of bounded measurable functions on  $\mathbb{X}$ . Note that  $S_T$  being defined in (19),  $\text{osc}(S_T) \leq \sum_{t=0}^T \text{osc}(h_t)$ . In the sequel  $C$  denotes a constant whose value may change upon each appearance.

**Theorem 1.** *Assume A1–3. Then, for all  $q \geq 2$ , there exists a constant  $C$  (depending only on  $q, \sigma_-, \sigma_+, c_-, \sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that for all  $T < \infty$  and for all additive functional  $S_T$  of the form (19), the  $L_q$ -mean error of the FFBS algorithms satisfies:*

$$\left\| \phi_{0:T|T}^N(S_T) - \phi_{0:T|T}(S_T) \right\|_q \leq C \max_{0 \leq t \leq T} \{\text{osc}(h_t)\} \left( \sqrt{\frac{T+1}{N}} + \frac{T+1}{N} \right).$$

The result provided in Theorem 1 can be extended to the FFBSi algorithm:

**Corollary 1.** *Under the assumptions of Theorem 1, the  $L_q$ -mean error of the FFBSi algorithm satisfies:*

$$\left\| \phi_{0:T|T}(S_T) - \tilde{\phi}_{0:T|T}^N(S_T) \right\|_q \leq C \max_{0 \leq t \leq T} \{\text{osc}(h_t)\} \left( \sqrt{\frac{T+1}{N}} + \frac{T+1}{N} \right),$$

where  $S_T$  is defined in (19).

The proofs of these results are postponed to Section 5. We provide below a brief outline of the main steps. Following [9], the proofs rely on a decomposition of the smoothing error. For all  $0 \leq t \leq T$  and all bounded and measurable function  $h$  on  $\mathbb{X}^{T+1}$  define the kernel  $L_{t,T} : \mathbb{X}^{t+1} \times \mathcal{B}(\mathbb{X})^{\otimes T+1} \rightarrow [0, 1]$  by

$$L_{t,T}h(x_{0:t}) \stackrel{\text{def}}{=} \int \cdots \int \prod_{u=t+1}^T M(x_{u-1}, dx_u) g(x_u, y_u) h(x_{0:T}). \quad (20)$$

The joint smoothing distribution may then be expressed, for all bounded and measurable function  $h$  on  $\mathbb{X}^{T+1}$ , by

$$\phi_{0:T|T}(h) = \frac{\phi_{0:t|t}[L_{t,T}h]}{\phi_{0:t|t}[L_{t,T}\mathbf{1}]},$$

and this suggests to decompose the smoothing error as follows

$$\begin{aligned} \Delta_T^N[h] &\stackrel{\text{def}}{=} \phi_{0:T|T}^N(h) - \phi_{0:T|T}(h) \\ &= \sum_{t=0}^T \frac{\phi_{0:t|t}^N[L_{t,T}h]}{\phi_{0:t|t}^N[L_{t,T}\mathbf{1}]} - \frac{\phi_{0:t-1|t-1}^N[L_{t-1,T}h]}{\phi_{0:t-1|t-1}^N[L_{t-1,T}\mathbf{1}]}, \end{aligned} \quad (21)$$

where we used the convention

$$\frac{\phi_{0:-1|t-1}^N[L_{t-1,T}h]}{\phi_{0:-1|t-1}^N[L_{t-1,T}\mathbf{1}]} = \frac{\phi_0[L_{0,T}h]}{\phi_0[L_{0,T}\mathbf{1}]} = \phi_{0:T|T}(h).$$

Furthermore, for all  $0 \leq t \leq T$ ,

$$\begin{aligned}\phi_{0:t|t}^N[\mathbf{L}_{t,T}h] &= \int \cdots \int \phi_{0:t|t}^N(dx_{0:t})\mathbf{L}_{t,T}h(x_{0:t}) \\ &= \int \cdots \int \phi_t^N(dx_t)\mathbf{B}_{\phi_{t-1}^N}(x_t, dx_{t-1}) \cdots \mathbf{B}_{\phi_0^N}(x_1, dx_0)\mathbf{L}_{t,T}h(x_{0:t}) \\ &= \int \phi_t^N(dx_t)\mathcal{L}_{t,T}^N h(x_t),\end{aligned}$$

where  $\mathcal{L}_{t,T}^N$  and  $\mathcal{L}_{t,T}$  are two kernels on  $\mathbb{X} \times \mathcal{B}(\mathbb{X})^{\otimes(T+1)}$  defined for all  $x_t \in \mathbb{X}$  by

$$\mathcal{L}_{t,T}h(x_t) \stackrel{\text{def}}{=} \int \cdots \int \mathbf{B}_{\phi_{t-1}}(x_t, dx_{t-1}) \cdots \mathbf{B}_{\phi_0}(x_1, dx_0)\mathbf{L}_{t,T}h(x_{0:t}) \quad (22)$$

$$\mathcal{L}_{t,T}^N h(x_t) \stackrel{\text{def}}{=} \int \cdots \int \mathbf{B}_{\phi_{t-1}^N}(x_t, dx_{t-1}) \cdots \mathbf{B}_{\phi_0^N}(x_1, dx_0)\mathbf{L}_{t,T}h(x_{0:t}). \quad (23)$$

For all  $1 \leq t \leq T$  we can write

$$\begin{aligned}\frac{\phi_{0:t|t}^N[\mathbf{L}_{t,T}h]}{\phi_{0:t|t}^N[\mathbf{L}_{t,T}\mathbf{1}]} - \frac{\phi_{0:t-1|t-1}^N[\mathbf{L}_{t-1,T}h]}{\phi_{0:t-1|t-1}^N[\mathbf{L}_{t-1,T}\mathbf{1}]} &= \frac{\phi_t^N[\mathcal{L}_{t,T}^N h]}{\phi_t^N[\mathcal{L}_{t,T}^N \mathbf{1}]} - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N h]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N \mathbf{1}]} \\ &= \frac{1}{\phi_t^N[\mathcal{L}_{t,T}^N \mathbf{1}]} \left( \phi_t^N[\mathcal{L}_{t,T}^N h] - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N h]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N \mathbf{1}]} \phi_t^N[\mathcal{L}_{t,T}^N \mathbf{1}] \right),\end{aligned}$$

and then,

$$\Delta_T^N[h] = \sum_{t=0}^T \frac{N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} G_{t,T}^N h(\xi_t^{N,\ell})}{N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})}, \quad (24)$$

with  $G_{t,T}^N$  is a kernel on  $\mathbb{X} \times \mathcal{B}(\mathbb{X})^{\otimes(T+1)}$  defined, for all  $x_t \in \mathbb{X}$  and all bounded and measurable function  $h$  on  $\mathbb{X}^{T+1}$ , by

$$G_{t,T}^N h(x_t) \stackrel{\text{def}}{=} \mathcal{L}_{t,T}^N h(x_t) - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N h]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N \mathbf{1}]} \mathcal{L}_{t,T}^N \mathbf{1}(x_t).$$

Two families of random variables  $(C_{i,T}^N(f))_{0 \leq i \leq T}$  and  $(D_{i,T}^N(f))_{0 \leq i \leq T}$  are now introduced to transform (24) into a suitable decomposition to compute an upper bound for the  $L_q$ -mean error. The idea is to replace  $N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})$  in (24) by its conditional expectation given  $\mathcal{F}_{t-1}^N$  to get a martingale difference (provided that  $\omega_t^{N,1} G_{t,T}^N f(\xi_t^{N,1})$  is centered given  $\mathcal{F}_{t-1}^N$ ). This conditional expectation is computed using the following intermediate result. For all measurable function  $h$  on  $\mathbb{X}$  and all  $0 \leq t \leq T$ ,

$$\mathbb{E} \left[ \omega_t^{N,1} h(\xi_t^{N,1}) \middle| \mathcal{F}_{t-1}^N \right] = \frac{\phi_{t-1}^N(Mg_t h)}{\phi_{t-1}^N(\vartheta_t)}. \quad (25)$$

Indeed,

$$\begin{aligned}
& \mathbb{E} \left[ \omega_t^{N,1} h(\xi_t^{N,1}) \middle| \mathcal{F}_{t-1}^N \right] \\
&= \mathbb{E} \left[ \frac{m(\xi_{t-1}^{N,1}, \xi_t^{N,1}) g_t(\xi_t^{N,1})}{\vartheta_t(\xi_{t-1}^{N,1}) p_t(\xi_{t-1}^{N,1}, \xi_t^{N,1})} h(\xi_t^{N,1}) \middle| \mathcal{F}_{t-1}^N \right] \\
&= \left( \sum_{i=1}^N \omega_{t-1}^{N,i} \vartheta_t(\xi_{t-1}^{N,i}) \right)^{-1} \sum_{i=1}^N \int \omega_{t-1}^{N,i} \vartheta_t(\xi_{t-1}^{N,i}) p_t(\xi_{t-1}^{N,i}, x) \frac{M(\xi_{t-1}^{N,i}, dx) g_t(x)}{\vartheta_t(\xi_{t-1}^{N,i}) p_t(\xi_{t-1}^{N,i}, x)} h(x) \\
&= \left( \sum_{i=1}^N \omega_{t-1}^{N,i} \vartheta_t(\xi_{t-1}^{N,i}) \right)^{-1} \sum_{i=1}^N \int \omega_{t-1}^{N,i} M(\xi_{t-1}^{N,i}, dx) g_t(x) h(x) \\
&= \frac{\phi_{t-1}^N(Mg_t h)}{\phi_{t-1}^N(\vartheta_t)}.
\end{aligned}$$

This result, applied with the function  $h = \mathcal{L}_{t,T} \mathbf{1}$ , yields

$$\mathbb{E} \left[ \omega_t^{N,1} \mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,1}) \middle| \mathcal{F}_{t-1}^N \right] = \frac{\phi_{t-1}^N(Mg_t \mathcal{L}_{t,T} \mathbf{1})}{\phi_{t-1}^N(\vartheta_t)} = \frac{\phi_{t-1}^N(\mathcal{L}_{t-1,T} \mathbf{1})}{\phi_{t-1}^N(\vartheta_t)}.$$

For any  $0 \leq t \leq T$ , define for all bounded and measurable function  $h$  on  $\mathbb{X}^{T+1}$ ,

$$\begin{aligned}
D_{t,T}^N(h) &\stackrel{\text{def}}{=} \mathbb{E} \left[ \omega_t^{N,1} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \middle| \mathcal{F}_{t-1}^N \right]^{-1} N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \\
&= \frac{\phi_{t-1}^N(\vartheta_t)}{\phi_{t-1}^N \left( \frac{\mathcal{L}_{t-1,T} \mathbf{1}}{|\mathcal{L}_{t-1,T} \mathbf{1}|_\infty} \right)} N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty},
\end{aligned} \tag{26}$$

$$\begin{aligned}
C_{t,T}^N(h) &\stackrel{\text{def}}{=} \left[ \frac{1}{N^{-1} \sum_{i=1}^N \omega_t^{N,i} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,i})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty}} - \frac{\phi_{t-1}^N(\vartheta_t)}{\phi_{t-1}^N \left( \frac{\mathcal{L}_{t-1,T} \mathbf{1}}{|\mathcal{L}_{t-1,T} \mathbf{1}|_\infty} \right)} \right] \\
&\quad \times N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^N h(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty}.
\end{aligned} \tag{27}$$

And then (24) is rewritten :

$$\Delta_T^N[h] = \sum_{t=0}^T D_{t,T}^N(h) + \sum_{t=0}^T C_{t,T}^N(h). \tag{28}$$

For any  $q \geq 2$ , the derivation of the upper bound relies on the triangle inequality:

$$\|\Delta_T^N[S_T]\|_q \leq \left\| \sum_{t=0}^T D_{t,T}^N(S_T) \right\|_q + \sum_{t=0}^T \|C_{t,T}^N(S_T)\|_q, \tag{29}$$

where  $S_T$  is defined in (19). The proof is completed by using Proposition 1 and Proposition 2. According to (28), the smoothing error can be decomposed into a sum of two terms which are considered separately. The first one is a martingale whose  $L_q$ -mean error is upper-bounded by  $\sqrt{(T+1)/N}$  as shown in Proposition 1. The second one is a sum of products,  $L_q$ -norm of which being bounded by  $1/N$  in Proposition 2.

The end of this section is devoted to the exponential deviation inequality for the error  $\Delta_T^N[S_T]$ . We use the decomposition of  $\Delta_T^N[S_T]$  obtained in (28) leading to a similar dependence on the ratio  $(T+1)/N$ . The martingale term  $D_{t,T}^N(S_T)$  is dealt with using the Azuma-Hoeffding inequality while the term  $C_{t,T}^N(S_T)$  needs a specific Hoeffding-type inequality for ratio of random variables.

**Theorem 2.** *Assume A1–3. Then, there exists a constant  $C$  (depending only on  $\sigma_-$ ,  $\sigma_+$ ,  $c_-$ ,  $\sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that for all  $T < \infty$ , all additive functional  $S_T$  of the form (19), all  $N \geq 1$  and for all  $\varepsilon > 0$ :*

$$\mathbb{P} \left\{ \left| \phi_{0:T|T}(S_T) - \phi_{0:T|T}^N(S_T) \right| > \varepsilon \right\} \leq 2 \exp \left( - \frac{CN\varepsilon^2}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\}^2 (T+1)} \right) + 8 \exp \left( - \frac{CN\varepsilon}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\} (T+1)} \right).$$

The extension of this theorem to the FFBSi algorithm is given in the following corollary.

**Corollary 2.** *Under the assumptions of Theorem 2, there exists a constant  $C$  (depending only on  $\sigma_-$ ,  $\sigma_+$ ,  $c_-$ ,  $\sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that for all  $T < \infty$ , all additive functional  $S_T$  of the form (19), all  $N \geq 1$  and for all  $\varepsilon > 0$ :*

$$\mathbb{P} \left\{ \left| \phi_{0:T|T}(S_T) - \tilde{\phi}_{0:T|T}^N(S_T) \right| > \varepsilon \right\} \leq 4 \exp \left( - \frac{CN\varepsilon^2}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\}^2 (T+1)} \right) + 8 \exp \left( - \frac{CN\varepsilon}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\} (T+1)} \right).$$

## 4 Monte-Carlo Experiments

In this section, the performance of the FFBSi algorithm is evaluated through simulations and compared to the so-called genealogical tree method. The main application of the estimation of smoothed additive functionals is the parameter estimation in non-linear and non-gaussian state-space model; see [13, 14] for an overview of such methods.

### 4.1 Linear gaussian model

Let us consider the following model:

$$\begin{cases} X_{t+1} &= \phi X_t + \sigma_u U_t, \\ Y_t &= X_t + \sigma_v V_t, \end{cases}$$

where  $X_0 \sim \mathcal{N}\left(0, \frac{\sigma_u^2}{1-\phi^2}\right)$ ,  $(U_t)_{t \geq 0}$  and  $(V_t)_{t \geq 0}$  are two sequences of independent and identically distributed standard gaussian random variables (which are independent from  $X_0$ ). The parameters  $(\phi, \sigma_u, \sigma_v)$  are assumed to be known. In Figure 1 the FFBSi algorithm is compared to the path-space method to compute the smoothed value of the empirical mean  $(T+1)^{-1} \sum_{t=0}^T \mathbb{E}[X_t | Y_{0:T}]$ . Data were generated using the linear

gaussian model with  $(\phi = 0.9, \sigma_u = 0.6, \sigma_v = 1)$ . The true value of this quantity was computed using the Kalman smoother. We display in Figure 1 the box and whisker plots of the estimations obtained with 100 independent Monte Carlo experiments. The FFBSi algorithm clearly outperforms the other method for comparable computational costs. In Table 1, the mean CPU times of each iteration corresponding to the two methods are given as a function of the number of particles (for  $T = 500$  and  $T = 1000$ ). Note that, as shown in Figure 7, the computational cost of the FFBSi algorithm grows linearly with the number of particles. In Figure 2 the score is computed under the same model with  $(\phi = 0.9, \sigma_u = 0.6, \sigma_v = 1)$ . Using the Fisher identity, the score (gradient of the log-likelihood) can be expressed as the conditional expectation of the gradient of the logarithm of the complete data likelihood given all the observations up to time  $T$ :

$$\begin{aligned} \frac{\partial}{\partial \phi} \log p(Y_{0:T}) &= -\frac{\phi}{1-\phi^2} + \frac{1}{\sigma_u^2} \mathbb{E} \left[ \phi X_0^2 + \sum_{t=1}^T X_{t-1}(X_t - \phi X_{t-1}) \middle| Y_{0:T} \right] \\ \frac{\partial}{\partial \sigma_u^2} \log p(Y_{0:T}) &= -\frac{T+1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \mathbb{E} \left[ (1-\phi^2)X_0^2 + \sum_{t=1}^T (X_t - \phi X_{t-1})^2 \middle| Y_{0:T} \right] \\ \frac{\partial}{\partial \sigma_v^2} \log p(Y_{0:T}) &= -\frac{T+1}{2\sigma_v^2} + \frac{1}{2\sigma_v^4} \mathbb{E} \left[ \sum_{t=0}^T (Y_t - X_t)^2 \middle| Y_{0:T} \right]. \end{aligned}$$

The box and whisker plots obtained with 100 independent Monte Carlo experiments give the same results as in the previous case.

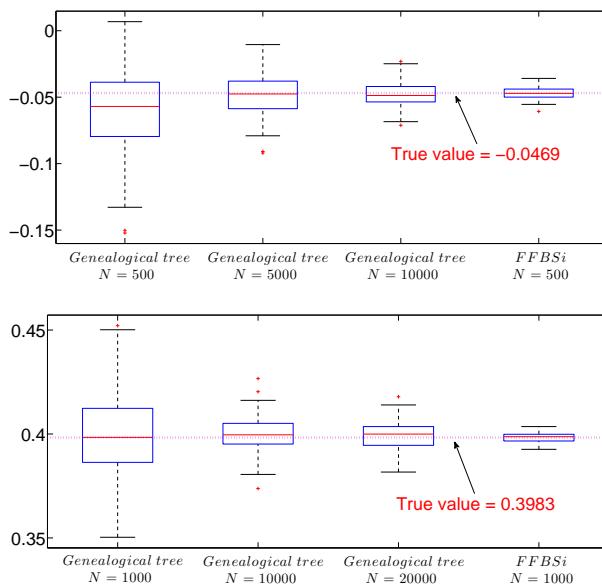


Figure 1: Computation of smoothed additive functionals in a linear gaussian model. The variance of the estimation given by the FFBSi algorithm is the smallest one in both cases. The first figure corresponds to the case  $T = 500$  and the second to  $T = 1000$ .

$T = 500$	FFBSi	Path-space method		
$N$	500	500	5000	10000
CPU time (s)	4.87	0.24	2.47	4.65

$T = 1000$	FFBSi	Path-space method		
$N$	1000	1000	10000	20000
CPU time (s)	16.5	0.9	8.5	17.2

Table 1: Average CPU time to compute the smoothed value of the empirical mean in the LGM

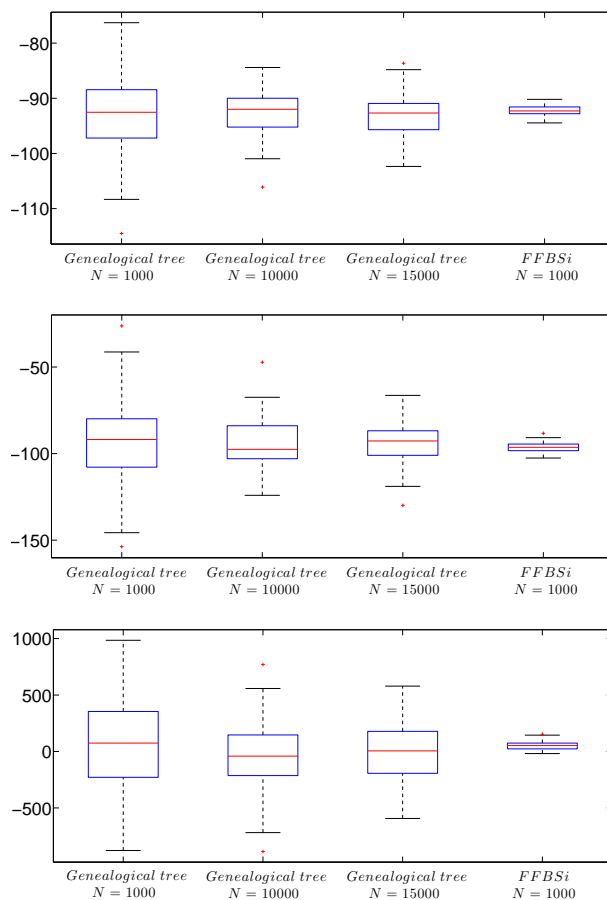


Figure 2: Computation of the score using the Fisher identity,  $T = 1000$ . These plots correspond to the partial derivatives with respect to  $\phi$ ,  $\sigma_u$  and  $\sigma_v$ .

## 4.2 Stochastic Volatility Model

Stochastic volatility models (SVM) have been introduced to provide better ways of modeling financial time series data than ARCH/GARCH models ([20]). Despite its apparent simplicity, parameter inference for this model proved to be challenging (see

[25] and references therein) and we focus on estimating the parameter  $\theta \stackrel{\text{def}}{=} (\phi, \sigma, \beta) \in \Theta$  characterizing the following SVM:

$$\begin{cases} X_{t+1} = \phi X_t + \sigma U_{t+1}, \\ Y_t = \beta e^{\frac{X_t}{2}} V_t, \end{cases}$$

where  $X_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{1-\phi^2}\right)$ ,  $U_t$  and  $V_t$  are independent standard gaussian random variables, and  $\Theta \stackrel{\text{def}}{=} (-1, 1) \times (0, \infty) \times (0, \infty)$ . The unknown parameter  $\theta^*$  can be recursively estimated according to the EM algorithm by a sequence  $(\hat{\theta}_n)_{n \geq 0}$  defined by:

$$\hat{\theta}_{n+1} = \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_{\hat{\theta}_n} [\log p_{\theta}(X_{0:T}, Y_{0:T}) | Y_{0:T}].$$

In this particular case, the sequence  $(\hat{\theta}_n)_{n \geq 0}$  can be computed by using the following updating formulas:

$$\begin{cases} \hat{\phi}_{n+1} = \frac{\sum_{t=1}^T \mathbb{E}_{\hat{\theta}_n} [X_{t-1} X_t | Y_{0:T}]}{\sum_{t=0}^{T-1} \mathbb{E}_{\hat{\theta}_n} [X_t^2 | Y_{0:T}]}, \\ \hat{\sigma}_{n+1}^2 = \frac{1}{T} \left( \sum_{t=1}^T \mathbb{E}_{\hat{\theta}_n} [X_t^2 | Y_{0:T}] + \hat{\phi}_{n+1}^2 \sum_{t=1}^T \mathbb{E}_{\hat{\theta}_n} [X_{t-1}^2 | Y_{0:T}] \right. \\ \quad \left. - 2 \hat{\phi}_{n+1} \sum_{t=1}^T \mathbb{E}_{\hat{\theta}_n} [X_{t-1} X_t | Y_{0:T}] \right), \\ \hat{\beta}_{n+1}^2 = \frac{1}{T+1} \sum_{t=0}^T Y_t^2 \mathbb{E}_{\hat{\theta}_n} [\exp(-X_t) | Y_{0:T}]. \end{cases} \quad (30)$$

This algorithm has been applied to foreign exchange data (EUR/USD) over the last five years. EUR/USD daily spot and returns (detrended according to [22]) are plotted in figure 3.

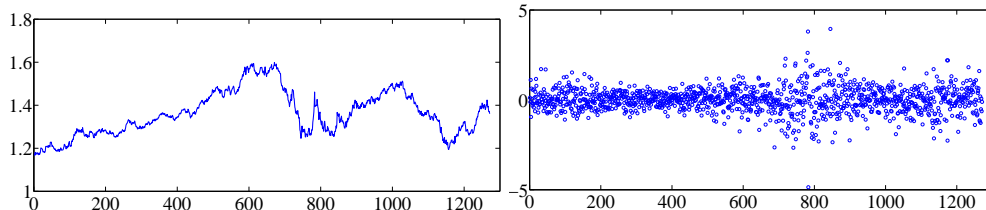


Figure 3: Daily spot (left) and daily detrended returns (right). EUR/USD data for the period 16/11/2005-16/11/2010

The path-space method and the FFBSi algorithm are used to approximate the conditional expectations of (30) with different numbers of particles  $N$ . The convergence of this EM algorithm is shown in Figures 4, 5 and 6. The number of particles has been chosen to be equal to the number of observations for the FFBSi algorithm, i.e.  $N = 1300$ . In order to make a fair comparison, the number of particles for the path-space method is such that the CPU time is almost the same (see Table 2), i.e.  $N = 20000$ . For the same CPU time, the FFBSi algorithm leads to better results in the sense that the variance of the estimator is lower than the one obtained with the path-space method. This can be seen through the smoother convergence of the EM algorithm when using the FFBSi rather than the path-space method.



	FFBSi	Path-space method	
$N$	1300	1300	20000
CPU time (s)	27	1.6	23

Table 2: Average CPU time for one EM iteration in the SVM

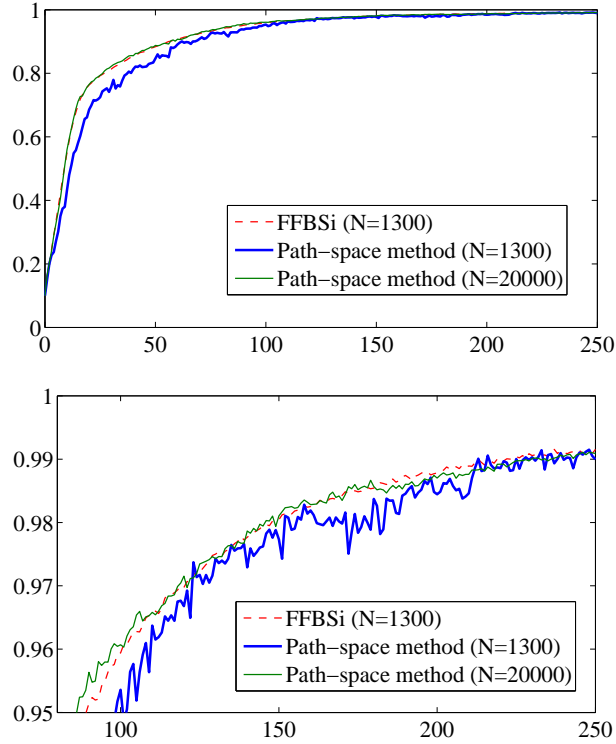


Figure 4: Estimation of  $\phi$  with  $T = 1270$  in the SVM. EM iterations from 0 to 250 and zoom on EM iterations from 80 to 250.

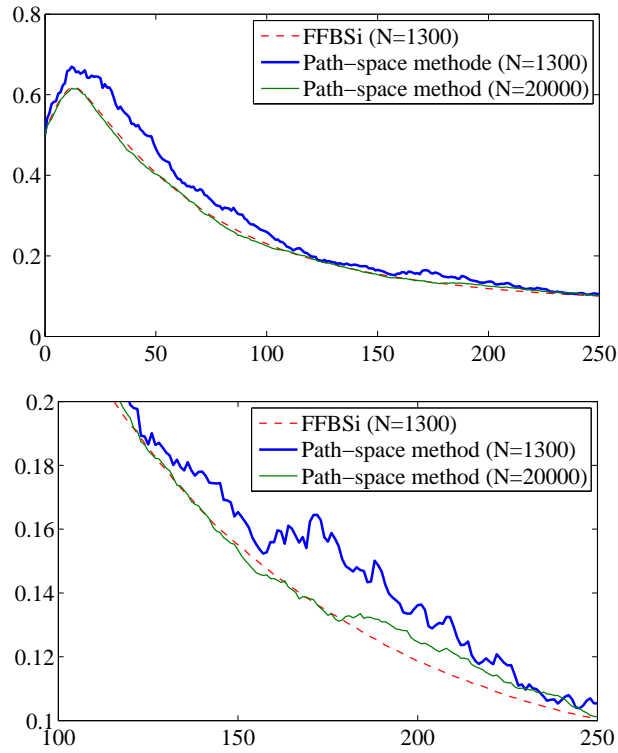


Figure 5: Estimation of  $\sigma$  with  $T = 1270$  in the SVM. EM iterations from 0 to 250 and zoom on EM iterations from 100 to 250.

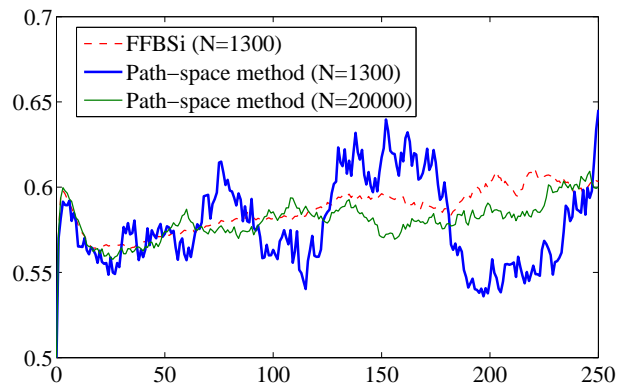


Figure 6: Estimation of  $\beta$  with  $T = 1270$  in the SVM.

Finally, Figure 7 outlines the linear complexity in the number of particles. It displays the empirical average CPU time needed to perform the FFBSi algorithm over 500 simulations for each number of particles in a stochastic volatility model with time horizon  $T = 1$ . Compared to the quadratic complexity of the classic FFBS algorithm, this will allow us to compute conditional expectations of additive functionals of the hidden states using a large number of particles.

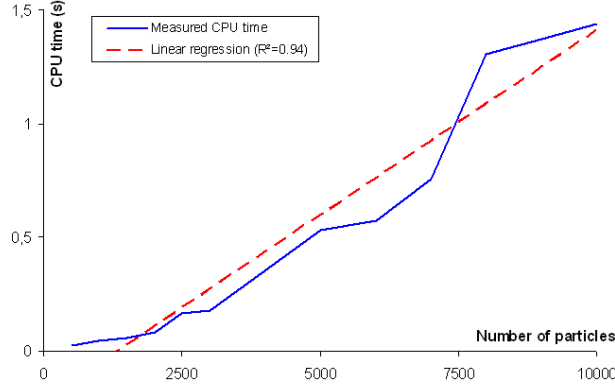


Figure 7: Computational cost

## 5 Proof of Theorem 1 and Corollary 1

We preface the proof of Proposition 1 by the following Lemma:

**Lemma 1.** *Under assumptions A1–3, we have, for any  $t \geq 0$ :*

(i) The random variables  $\left( \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right)_{1 \leq \ell \leq N}$  are, for all  $N \in \mathbb{N}$ :

(a) conditionally independent and identically distributed given  $\mathcal{F}_{t-1}^N$ ,

(b) centered conditionally to  $\mathcal{F}_{t-1}^N$ ,

(c) bounded:

$$\left| \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right| \leq \frac{2\sigma_+ - \sigma_-}{\sigma_-} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\}. \quad (31)$$

(ii) For all  $x \in \mathbb{X}$ ,  $\frac{\mathcal{L}_{t,T}\mathbf{1}(x)}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \geq \frac{\sigma_-}{\sigma_+}$  and  $\frac{\mathcal{L}_{t-1,T}\mathbf{1}(x)}{|\mathcal{L}_{t-1,T}\mathbf{1}|_\infty} \geq c_- \frac{\sigma_-}{\sigma_+}$ .

*Proof.* Proof of (i) (a) and (b) are directly given by [9, Lemma 3]. Define  $\Pi_{s,T}$  the operator from  $\mathcal{F}_b(\mathbb{X})$  to  $\mathcal{F}_b(\mathbb{X}^{T+1})$  by

$$\forall (x_0, \dots, x_T) \in \mathbb{X}^{T+1}, \quad \Pi_{s,T} h(x_{0:T}) \stackrel{\text{def}}{=} h(x_s).$$

Using this notation we may write  $S_T = \sum_{s=0}^T \Pi_{s,T} h_s$  and  $G_{t,T}^N S_T = \sum_{s=0}^T G_{t,T}^N \Pi_{s,T} h_s$ . Furthermore, [9, Lemma 10] shows that

$$|G_{t,T}^N \Pi_{s,T} h_s|_\infty \leq \rho^{|t-s|} \text{osc}(h_s) |\mathcal{L}_{t,T}\mathbf{1}|_\infty, \quad \forall (s, t) \in \{0, \dots, T\}^2,$$

where  $\rho = 1 - \sigma_-/\sigma_+$ . Consequently,

$$|G_{t,T}^N S_T|_\infty \leq \sum_{s=0}^T |G_{t,T} \Pi_{s,T} h_s|_\infty \leq \sum_{s=0}^T \rho^{|t-s|} \text{osc}(h_s) |\mathcal{L}_{t,T} \mathbf{1}|_\infty,$$

which implies that, for all  $T \in \mathbb{N}$  and all  $N \in \mathbb{N}$ ,

$$\left| \frac{G_{t,T}^N S_T}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \right| \leq \frac{1+\rho}{1-\rho} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\},$$

which shows (c).

Proof of (ii) From the definition (22), for all  $x \in \mathbb{X}$  and all  $t \in \{1, \dots, T\}$ ,

$$\mathcal{L}_{t,T} \mathbf{1}(x) = \int m(x, x_{t+1}) g_{t+1}(x_{t+1}) \prod_{u=t+2}^T M(x_{u-1}, dx_u) g_u(x_u) \lambda(dx_{t+1}),$$

hence, by assumption A3,

$$\begin{aligned} |\mathcal{L}_{t,T} \mathbf{1}|_\infty &\leq \sigma_+ \int g_{t+1}(x_{t+1}) \mathcal{L}_{t+1,T} \mathbf{1}(x_{t+1}) \lambda(dx_{t+1}) \\ \mathcal{L}_{t,T} \mathbf{1}(x) &\geq \sigma_- \int g_{t+1}(x_{t+1}) \mathcal{L}_{t+1,T} \mathbf{1}(x_{t+1}) \lambda(dx_{t+1}), \end{aligned}$$

which concludes the proof of the first statement. By construction, for any  $x \in \mathbb{X}$  and any  $t \in \{1, \dots, T\}$ ,

$$\mathcal{L}_{t-1,T} \mathbf{1}(x) = \int M(x, dx') g_t(x') \mathcal{L}_{t,T} \mathbf{1}(x'), \quad (32)$$

and then, by assumption A3,

$$\frac{\mathcal{L}_{t-1,T} \mathbf{1}(x)}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} = \int M(x, dx') g_t(x') \frac{\mathcal{L}_{t,T} \mathbf{1}(x')}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \geq c_- \frac{\sigma_-}{\sigma_+}.$$

□

**Proposition 1.** *Assume A1–3. Then, for all  $q \geq 2$ , there exists a constant  $C$  (depending only on  $q, \sigma_-, \sigma_+, c_-, \sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that for all  $T < \infty$ ,*

$$\left\| \sum_{t=0}^T D_{t,T}^N(S_T) \right\|_q \leq C \max_{0 \leq t \leq T} \{\text{osc}(h_t)\} \sqrt{\frac{T+1}{N}}, \quad (33)$$

where  $D_{t,T}^N(S_T)$  is defined in (26).

*Proof.* Since  $(D_{t,T}^N(S_T))_{t \leq T}$  is a sequence of  $\mathcal{F}_t^N$ -measurable random variables such that, for all  $0 \leq t \leq T$ ,  $\mathbb{E}[D_{t,T}^N(S_T) | \mathcal{F}_{t-1}^N] = 0$ ,  $(D_{t,T}^N(S_T))_{t \leq T}$  is a forward martingale difference. Hence, since  $q \geq 2$ , Burkholder's inequality (see [19, Theorem 2.10, page 23]) states the existence of a constant  $C$  such that:

$$\mathbb{E} \left[ \left| \sum_{t=0}^T D_{t,T}^N(S_T) \right|^q \right] \leq C \mathbb{E} \left[ \left| \sum_{t=0}^T D_{t,T}^N(S_T)^2 \right|^{\frac{q}{2}} \right].$$

Moreover, by application of the last statement of Lemma 1,

$$\frac{\phi_{t-1}^N(\vartheta_t)}{\phi_{t-1}^N\left(\frac{\mathcal{L}_{t,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}\right)} \leq \frac{\sigma_+ \sup_{t \geq 0} |\vartheta_t|_\infty}{\sigma_- c_-},$$

and thus,

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{t=0}^T D_{t,T}^N(S_T)^2 \right|^{\frac{q}{2}} \right] &\leq \left( \frac{\sigma_+ \sup_{t \geq 0} |\vartheta_t|_\infty}{\sigma_- c_-} \right)^q \\ &\quad \times \mathbb{E} \left[ \left| \sum_{t=0}^T \left( N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right)^2 \right|^{\frac{q}{2}} \right], \end{aligned}$$

which implies, using the Jensen inequality,

$$\mathbb{E} \left[ \left| \sum_{t=0}^T D_{t,T}^N(S_T) \right|^q \right] \leq C (T+1)^{\frac{q}{2}-1} \sum_{t=0}^T \mathbb{E} \left[ \left| N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right|^{2q} \right].$$

Since for any  $t \geq 0$  the random variables  $\left( \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right)_{1 \leq \ell \leq N}$  are conditionally independent and centered conditionally to  $\mathcal{F}_{t-1}^N$ , using again the Burkholder and the Jensen inequalities we obtain

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right|^q \middle| \mathcal{F}_{t-1}^N \right] &\leq C N^{q/2-1} \sum_{\ell=1}^N \mathbb{E} \left[ \left| \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right|^q \middle| \mathcal{F}_{t-1}^N \right] \\ &\leq C \max_{0 \leq t \leq T} \{ \text{osc}(h_t) \}^q N^{q/2}, \end{aligned}$$

where the last inequality comes from the first statement of Lemma 1.  $\square$

**Proposition 2.** *Assume A1-3. Then, for all  $q \geq 2$ , there exists a constant  $C$  (depending only on  $q, \sigma_-, \sigma_+, c_-, \sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that for all  $T < \infty$  and  $0 \leq t \leq T$ :*

$$\|C_{t,T}^N(S_T)\|_q \leq \frac{C}{N} \max_{0 \leq t \leq T} \{ \text{osc}(h_t) \}, \quad (34)$$

where  $C_{t,T}^N(S_T)$  is defined in (27).

*Proof.* According to (27),  $C_{t,T}^N(S_T)$  can be written

$$C_{t,T}^N(S_T) = U_{t,T}^N V_{t,T}^N W_{t,T}^N, \quad (35)$$

where

$$\begin{aligned} U_{t,T}^N &= \frac{N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}}{N^{-1} \Omega_t^N} \\ V_{t,T}^N &= N^{-1} \sum_{\ell=1}^N \left( \mathbb{E} \left[ \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \middle| \mathcal{F}_{t-1} \right] - \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right) \\ W_{t,T}^N &= \frac{N^{-1} \Omega_t^N}{\mathbb{E} \left[ \omega_t^{N,1} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \middle| \mathcal{F}_{t-1} \right] N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T}\mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty}}. \end{aligned}$$

Using the last statement of Lemma 1, we get the following bound:

$$\begin{aligned} \mathbb{E} \left[ \omega_t^{N,1} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \middle| \mathcal{F}_{t-1} \right] &= \frac{\phi_{t-1}^N (\mathcal{L}_{t-1,T} \mathbf{1} / |\mathcal{L}_{t,T} \mathbf{1}|_\infty)}{\phi_{t-1}^N (\vartheta_t)} \\ &\geq \frac{c_- \sigma_-}{|\vartheta_t|_\infty \sigma_+} \\ \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} &\geq \frac{\sigma_-}{\sigma_+}, \end{aligned}$$

which implies

$$|W_{t,T}^N| \leq \left( \frac{\sigma_+}{\sigma_-} \right)^2 \frac{|\vartheta_t|_\infty}{c_-}. \quad (36)$$

Then, the Cauchy-Schwarz inequality leads to

$$\|C_{t,T}^N(S_T)\|_q \leq C \|U_{t,T}^N\|_{2q} \|V_{t,T}^N\|_{2q}. \quad (37)$$

The random variables  $\left( \mathbb{E} \left[ \omega_t^{N,1} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \middle| \mathcal{F}_{t-1} \right] - \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \right)_{\ell \in \{1, \dots, N\}}$  being conditionally independent, centered and bounded given  $\mathcal{F}_{t-1}^N$ , following the same steps as in the proof of Proposition 1, there exists a constant  $C$  (depending only on  $q$ ,  $\sigma_-$ ,  $\sigma_+$ ,  $c_-$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that

$$\|V_{t,T}^N\|_{2q} \leq \frac{C}{\sqrt{N}}.$$

In order to handle the first norm of the RHS of (37), we use Lemma 3 in the appendix with  $\widehat{Z}_N = N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{G_{t,T}^{N,1} S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty}$ ,  $\widehat{W}_N = N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell}$ ,  $W_N = \mathbb{E} \left[ N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \middle| \mathcal{F}_{t-1}^N \right]$ ,  $X_{\ell,N} = \omega_t^{N,\ell}$  and  $Y_{\ell,N} = \frac{G_{t,T}^{N,1} S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty}$ . By Lemma 1,  $\mathbb{E} \left[ \widehat{Z}_N \middle| \mathcal{F}_{t-1}^N \right] = 0$ ,  $|\widehat{Z}_N| \leq |\omega_t|_\infty \frac{2\sigma_+ - \sigma_-}{\sigma_-} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\} \stackrel{\text{def}}{=} K'$  and  $\max_{1 \leq \ell \leq N} |Y_{\ell,N}| \leq \frac{2\sigma_+ - \sigma_-}{\sigma_-} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\} \stackrel{\text{def}}{=} K$ . Using the same arguments than in the proof of Proposition 1, we have,

$$\begin{aligned} \mathbb{E} \left[ \left| \widehat{Z}_N \right|^{2q} \middle| \mathcal{F}_{t-1}^N \right]^{1/2q} &\leq \frac{C}{\sqrt{N}} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\} \stackrel{\text{def}}{=} v_N \\ \mathbb{E} \left[ \left| \widehat{W}_N - W_N \right|^{4q} \middle| \mathcal{F}_{t-1}^N \right]^{1/4q} &\leq \frac{C}{\sqrt{N}} \stackrel{\text{def}}{=} v'_N. \end{aligned}$$

This Lemma gives

$$W_N \mathbb{E} \left[ |U_{t,T}|^{2q} \middle| \mathcal{F}_{t-1}^N \right]^{1/2q} \leq v_N + \left( \frac{1}{W_N} K' + K \right) v'_N.$$

According to assumptions A1 and A3 and to (25),

$$W_N = \mathbb{E} \left[ N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \middle| \mathcal{F}_{t-1}^N \right] = \mathbb{E} \left[ \omega_t^{N,1} \middle| \mathcal{F}_{t-1}^N \right] = \frac{\phi_{t-1}^N (Mg_t)}{\phi_{t-1}^N (\vartheta_t)} \geq \frac{c_-}{\sup_{t \geq 0} |\vartheta_t|_\infty},$$

and then

$$\|U_{t,T}^N\|_{2q} \leq \frac{C}{\sqrt{N}} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\}. \quad (38)$$

□

We preface the proof of Corollary 1 by the following Lemma. We first define the backward filtration  $(\mathcal{G}_{t,T}^N)_{t \in \{0, \dots, T+1\}}$  by

$$\begin{cases} \mathcal{G}_{T+1,T}^N = \mathcal{F}_T^N, \\ \mathcal{G}_{t,T}^N = \mathcal{F}_t^N \vee \sigma \{J_u^\ell, 1 \leq \ell \leq N, t \leq u \leq T\}, \quad \forall 0 \leq t \leq T. \end{cases} \quad (39)$$

**Lemma 2.** *Under assumptions A1–3, we have for all bounded measurable function  $h$  on  $\mathbb{X}$ , all  $\ell \in \{1, \dots, N\}$ , and all  $u, t, T$  such that  $0 \leq t \leq u \leq T$ ,*

$$\left| \mathbb{E} \left[ h \left( \xi_t^{N, J_t^\ell} \right) \middle| \mathcal{G}_{u,T}^N \right] - \mathbb{E} \left[ h \left( \xi_t^{N, J_t^\ell} \right) \middle| \mathcal{G}_{u+1,T}^N \right] \right| \leq \rho^{u-t} \text{osc}(h),$$

where  $\rho = 1 - \sigma_- / \sigma_+$ .

*Proof.* According to Section 2.2, for all  $\ell \in \{1, \dots, N\}$ ,  $\{J_u^{N, \ell}\}_{u=0}^T$  is an inhomogeneous Markov chain evolving backward in time with backward kernel  $\{\Lambda_u^N\}_{u=0}^{T-1}$ . For any  $0 \leq t \leq u \leq T$ , we have

$$\begin{aligned} & \mathbb{E} \left[ h \left( \xi_t^{N, J_t^{N, \ell}} \right) \middle| \mathcal{G}_{u,T}^N \right] - \mathbb{E} \left[ h \left( \xi_t^{N, J_t^{N, \ell}} \right) \middle| \mathcal{G}_{u+1,T}^N \right] \\ &= \sum_{j_t:u} \left[ \delta_{J_u^{N, \ell}}(j_u) - \left( \Lambda_u(J_{u+1}^{N, \ell}, j_u) \mathbf{1}_{u < T} + \frac{\omega_T^{N, j_u}}{\Omega_u} \mathbf{1}_{u=T} \right) \right] \\ & \quad \times \Lambda_{u-1}^N(j_u, j_{u-1}) \dots \Lambda_t^N(j_{t+1}, j_t) h \left( \xi_t^{N, j_t} \right). \end{aligned}$$

The RHS of this equation is the difference between two expectations started with two different initial distributions. Under A3, the backward kernel satisfies the uniform Doeblin condition,

$$\forall (i, j) \in \{1, \dots, N\}^2 \quad \Lambda_s^N(i, j) \geq \frac{\sigma_- \omega_s^i}{\sigma_+ \Omega_s^N},$$

and the proof is completed by the exponential forgetting of the backward kernel (see [3, 6]).  $\square$

To compute an upper-bound for the  $L_q$ -mean error of the FFBSi algorithm, we may define the difference between the FFBS and the FFBSi estimators:

$$\delta_T^N = \tilde{\phi}_{0:T|T}^N(S_T) - \phi_{0:T|T}^N(S_T). \quad (40)$$

*Proof of Corollary 1.* The difference between the FFBS and the FFBSi estimators,  $\delta_T^N$ , defined in (40), can be written

$$\begin{aligned} \delta_T^N &= \frac{1}{N} \sum_{\ell=1}^N \sum_{t=0}^T h_t \left( \xi_t^{N, J_t^{N, \ell}} \right) - \mathbb{E} \left[ h_t \left( \xi_t^{N, J_t^{N, 1}} \right) \middle| \mathcal{F}_T^N \right] \\ &= \frac{1}{N} \sum_{\ell=1}^N \sum_{t=0}^T \sum_{u=t}^T \mathbb{E} \left[ h_t \left( \xi_t^{N, J_t^{N, \ell}} \right) \middle| \mathcal{G}_{u,T}^N \right] - \mathbb{E} \left[ h_t \left( \xi_t^{N, J_t^{N, \ell}} \right) \middle| \mathcal{G}_{u+1,T}^N \right] \\ &= \frac{1}{N} \sum_{\ell=1}^N \sum_{u=0}^T \zeta_u^{N, \ell}, \end{aligned}$$

where  $\zeta_u^{N, \ell} \stackrel{\text{def}}{=} \sum_{t=0}^u \mathbb{E} \left[ h_t \left( \xi_t^{N, J_t^{N, \ell}} \right) \middle| \mathcal{G}_{u,T}^N \right] - \mathbb{E} \left[ h_t \left( \xi_t^{N, J_t^{N, \ell}} \right) \middle| \mathcal{G}_{u+1,T}^N \right]$ .

For all  $\ell \in \{1, \dots, N\}$  and all  $u \in \{0, \dots, T\}$ , the random variable  $\zeta_u^{N, \ell}$  is  $\mathcal{G}_{u,T}^N$ -measurable and  $\mathbb{E} [\zeta_u^{N, \ell} | \mathcal{G}_{u+1,T}^N] = 0$  so that  $\zeta_u^{N, \ell}$  can be seen as the increment of a

backward martingale. Hence, since  $q \geq 2$ , using the Burkholder inequality (see [19, Theorem 2.10, page 23]) and the Jensen inequality, there exists a constant  $C$  (depending only on  $q, \sigma_-, \sigma_+, c_-, \sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that:

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{u=0}^T \frac{1}{N} \sum_{\ell=1}^N \zeta_u^{N,\ell} \right|^q \right] &\leq CN^{-q} \mathbb{E} \left[ \left| \sum_{u=0}^T \left( \sum_{\ell=1}^N \zeta_u^{N,\ell} \right)^2 \right|^{\frac{q}{2}} \right] \\ &\leq CN^{-q} (T+1)^{q/2-1} \sum_{u=0}^T \mathbb{E} \left[ \left| \sum_{\ell=1}^N \zeta_u^{N,\ell} \right|^q \right]. \end{aligned} \quad (41)$$

Then, since the random variables  $(\zeta_u^{N,\ell})_{\ell \in \{1, \dots, N\}}$  are conditionally independent and centered conditionally to  $\mathcal{G}_{u+1,T}^N$ , using the Burkholder inequality once again implies:

$$\mathbb{E} \left[ \left| \sum_{\ell=1}^N \zeta_u^{N,\ell} \right|^q \middle| \mathcal{G}_{u+1,T}^N \right] \leq CN^{q/2-1} \sum_{\ell=1}^N \mathbb{E} \left[ |\zeta_u^{N,\ell}|^q \middle| \mathcal{G}_{u+1,T}^N \right]. \quad (42)$$

Furthermore, according to Lemma 2,

$$\begin{aligned} |\zeta_u^{N,\ell}| &\leq \sum_{t=0}^u \left| \mathbb{E} \left[ h_t \left( \xi_t^{N, J_t^{N,\ell}} \right) \middle| \mathcal{G}_{u,T}^N \right] - \mathbb{E} \left[ h_t \left( \xi_t^{N, J_t^{N,\ell}} \right) \middle| \mathcal{G}_{u+1,T}^N \right] \right| \\ &\leq \sum_{t=0}^u \rho^{u-t} \text{osc}(h_t) \leq \frac{\sigma_+}{\sigma_-} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\}. \end{aligned} \quad (43)$$

Putting (41), (42) and (43) together leads to

$$\|\delta_T^N\|_q \leq C \sqrt{\frac{T+1}{N}} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\},$$

and the announced result is derived from the triangle inequality:

$$\left\| \phi_{0:T|T}(S_T) - \tilde{\phi}_{0:T|T}^N(S_T) \right\|_q \leq \|\Delta_T^N[S_T]\|_q + \|\delta_T^N\|_q.$$

□

## 6 Proof of Theorem 2 and Corollary 2

We preface the proof of the Theorem by showing that the martingale term of the error  $\Delta_T^N[S_T]$  satisfies an exponential deviation inequality in the following Proposition.

**Proposition 3.** *Assume A1–3. Then there exists a constant  $C$  (depending only on  $\sigma_-, \sigma_+, c_-, \sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that for all  $T < \infty$ , all  $N \geq 1$  and for all  $\varepsilon > 0$ ,*

$$\mathbb{P} \left\{ \left| \sum_{t=0}^T D_{t,T}^N(S_T) \right| > \varepsilon \right\} \leq 2 \exp \left( - \frac{CN\varepsilon^2}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\}^2 (T+1)} \right), \quad (44)$$

where  $D_{t,T}^N(S_T)$  is defined in (26).



*Proof.* According to the definition of  $D_{t,T}^N(S_T)$  given in (26), we can write

$$\sum_{t=0}^T D_{t,T}^N(S_T) = \sum_{k=1}^{N(T+1)} \Upsilon_k^N,$$

where for all  $0 \leq t \leq T$  and  $1 \leq \ell \leq N$ ,  $\Upsilon_{Nt+\ell}^N$  is defined by

$$\Upsilon_{Nt+\ell}^N = \frac{\phi_{t-1}^N(\vartheta_t)}{\phi_{t-1}^N \left( \frac{\mathcal{L}_{t-1,T}\mathbf{1}}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty} \right)} N^{-1} \omega_t^{N,\ell} \frac{G_{t,T}^N S_T(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T}\mathbf{1}|_\infty},$$

and is bounded by (see Lemma 1)

$$|\Upsilon_{Nt+\ell}^N| \leq CN^{-1} \max_{0 \leq t \leq T} \{\text{osc}(h_t)\}.$$

Furthermore, we define the filtration  $(\mathcal{H}_k^N)_{1 \leq k \leq N(T+1)}$ , for all  $0 \leq t \leq T$  and  $1 \leq \ell \leq N$ , by:

$$\mathcal{H}_{Nt+\ell}^N = \mathcal{F}_{t-1}^N \vee \sigma \left\{ \left( \omega_t^{N,i}, \xi_t^{N,i} \right), 1 \leq i \leq \ell \right\},$$

with the convention  $\mathcal{F}_{-1}^N = \sigma(Y_{0:T})$ . Then, according to Lemma 1,  $(\Upsilon_k)_{1 \leq k \leq N(T+1)}$  is martingale increment for the filtration  $\mathcal{H}$  and the Azuma-Hoeffding inequality completes the proof.  $\square$

**Proposition 4.** *Assume A1–3. Then there exists a constant  $C$  (depending only on  $\sigma_-$ ,  $\sigma_+$ ,  $c_-$ ,  $\sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ ) such that for all  $T < \infty$ , all  $N \geq 1$  and for all  $\varepsilon > 0$ ,*

$$\mathbb{P} \left\{ \left| \sum_{t=0}^T C_{t,T}^N(S_T) \right| > \varepsilon \right\} \leq 8 \exp \left( - \frac{CN\varepsilon}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\} (T+1)} \right), \quad (45)$$

where  $C_{t,T}^N(S_T)$  is defined in (27).

*Proof.* In order to apply Lemma 5 in the appendix, we first need to find an exponential deviation inequality for  $C_{t,T}^N(S_T)$  which is done by using the decomposition  $C_{t,T}^N(S_T) = U_{t,T}^N V_{t,T}^N W_{t,T}^N$  given in (35). First, the ratio  $U_{t,T}^N$  is dealt with through Lemma 4 in the appendix by defining

$$\begin{cases} a_N \stackrel{\text{def}}{=} N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} G_{t,T}^N S_T(\xi_t^{N,\ell}) / |\mathcal{L}_{t,T}\mathbf{1}|_\infty, \\ b_N \stackrel{\text{def}}{=} N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell}, \\ b \stackrel{\text{def}}{=} \mathbb{E}[\omega_t^1 | \mathcal{F}_{t-1}^N] = \phi_{t-1}^N [Mg_t] / \phi_{t-1}^N(\vartheta_t), \\ \beta \stackrel{\text{def}}{=} c_- / |\vartheta_t|_\infty. \end{cases}$$

Assumption A3 shows that  $b \geq \beta$  and Lemma 1 (i) shows that  $|a_N/b_N| \leq \frac{2\sigma_+ - \sigma_-}{\sigma_-} \text{osc}(h)$ . Therefore, Condition (I) of Lemma 4 is satisfied. The bounds  $0 < \omega_t^1 \leq |\omega_t|_\infty$  and the Hoeffding inequality lead to

$$\begin{aligned} \mathbb{P}[|b_N - b| \geq \varepsilon] &= \mathbb{E} \left[ \mathbb{P} \left[ \left| N^{-1} \sum_{\ell=1}^N \left( \omega_t^{N,\ell} - \mathbb{E}[\omega_t^{N,1} | \mathcal{F}_{t-1}^N] \right) \right| \geq \varepsilon \middle| \mathcal{F}_{t-1}^N \right] \right] \\ &\leq 2 \exp \left( - \frac{2N\varepsilon^2}{|\omega_t|_\infty^2} \right), \end{aligned}$$

establishing Condition (II) in Lemma 4. Finally, Lemma 1 (i) and the Hoeffding inequality imply that

$$\begin{aligned} \mathbb{P}[|a_N| \geq \varepsilon] &= \mathbb{E} \left[ \mathbb{P} \left[ \left| N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} G_{t,T}^N S_T(\xi_t^{N,\ell}) / |\mathcal{L}_{t,T} \mathbf{1}|_\infty \right| \geq \varepsilon \mid \mathcal{F}_{t-1}^N \right] \right] \\ &\leq 2 \exp \left( - \frac{N\varepsilon^2}{2|\omega_t|_\infty^2 \left( \frac{2\sigma_+ - \sigma_-}{\sigma_-} \right)^2 \max_{0 \leq t \leq T} \{\text{osc}(h_t)\}^2} \right). \end{aligned}$$

Lemma 4 therefore yields

$$\mathbb{P} \{ |U_{t,T}^N| \geq \varepsilon \} \leq 2 \exp \left( -C \frac{N\varepsilon^2}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\}^2} \right).$$

Then  $V_{t,T}^N$  is dealt with by using again the Hoeffding inequality and the bounds  $0 < \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \leq |\omega_t|_\infty$ :

$$\begin{aligned} &\mathbb{P} \left[ \left| N^{-1} \sum_{\ell=1}^N \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} - \mathbb{E} \left[ \xi_t^{N,1} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \mid \mathcal{F}_{t-1} \right] \right| \geq \varepsilon \right] \\ &= \mathbb{E} \left[ \mathbb{P} \left[ \left| N^{-1} \sum_{\ell=1}^N \left( \omega_t^{N,\ell} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,\ell})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} - \mathbb{E} \left[ \omega_t^{N,1} \frac{\mathcal{L}_{t,T} \mathbf{1}(\xi_t^{N,1})}{|\mathcal{L}_{t,T} \mathbf{1}|_\infty} \mid \mathcal{F}_{t-1}^N \right] \right) \right| \geq \varepsilon \mid \mathcal{F}_{t-1}^N \right] \right] \\ &\leq 2 \exp(-CN\varepsilon^2). \end{aligned}$$

Finally,  $W_{t,T}^N$  has been shown in (36) to be bounded by a constant depending only on  $\sigma_-$ ,  $\sigma_+$ ,  $c_-$ ,  $\sup_{t \geq 1} |\vartheta_t|_\infty$  and  $\sup_{t \geq 0} |\omega_t|_\infty$ :  $|W_{t,T}^N| \leq C$  so that

$$\begin{aligned} \mathbb{P} \{ |C_{t,T}^N(S_T)| > \varepsilon \} &\leq \mathbb{P} \{ |U_{t,T}^N V_{t,T}^N| > \varepsilon / C \} \\ &\leq \mathbb{P} \left\{ |U_{t,T}^N| > \sqrt{\varepsilon \max_{0 \leq t \leq T} \{\text{osc}(h_t)\} / C} \right\} + \mathbb{P} \left\{ |V_{t,T}^N| > \sqrt{\varepsilon / C \max_{0 \leq t \leq T} \{\text{osc}(h_t)\}} \right\} \\ &\leq 4 \exp \left( - \frac{CN\varepsilon}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\}} \right). \end{aligned}$$

The proof is finally completed by applying Lemma 5 with

$$X_t = C_{t,T}^N(S_T), \quad A = 4, \quad B = \frac{CN}{\max_{0 \leq t \leq T} \{\text{osc}(h_t)\}}, \quad \gamma = 1/2.$$

□

*Proof of Theorem 2.* The result is obtained by writing

$$\begin{aligned} \mathbb{P} \{ |\Delta_T^N[S_T]| > \varepsilon \} &\leq \mathbb{P} \left\{ \left| \sum_{t=0}^T C_{t,T}^N(S_T) \right| > \varepsilon / 2 \right\} \\ &\quad + \mathbb{P} \left\{ \left| \sum_{t=0}^T D_{t,T}^N(S_T) \right| > \varepsilon / 2 \right\}, \quad (46) \end{aligned}$$

and using (44) and (45). □

*Proof of Corollary 2.* We recall the decomposition used in the proof of Corollary 1:

$$\delta_T^N = \frac{1}{N} \sum_{\ell=1}^N \sum_{u=0}^T \zeta_u^{N,\ell}.$$

Since  $(\zeta_u^{N,\ell})_{\ell \in \{1, \dots, N\}}$  are  $\mathcal{G}_{u,T}^N$  measurable and centered conditionally to  $\mathcal{G}_{u+1,T}^N$  using the same steps as in the proof of Proposition 4, we get

$$\mathbb{P} \{ |\delta_T^N| > \varepsilon \} \leq 2 \exp \left( - \frac{\varepsilon^2 \sigma_- N}{2\sigma_+(T+1) \max_{0 \leq t \leq T} \{\text{osc}(h_t)\}^2} \right). \quad (47)$$

The proof is finally completed by writing  $\phi_{0:T|T}(S_T) - \tilde{\phi}_{0:T|T}^N(S_T) = \Delta_T^N[S_T] + \delta_T^N$  and by using Theorem 2.  $\square$

## A Technical results

**Lemma 3.** *Let  $N \in \mathbb{N}^*$ ,  $q \geq 2$  and  $(X_\ell)_{\ell \in \{1, \dots, N\}}$  and  $(Y_\ell)_{\ell \in \{1, \dots, N\}}$  two sequences of random variables. Define  $\widehat{Z}_N = N^{-1} \sum_{\ell=1}^N X_{\ell,N} Y_{\ell,N}$ ,  $\widehat{W}_N = N^{-1} \sum_{\ell=1}^N X_{\ell,N}$ ,  $Z_N = \mathbb{E}[\widehat{Z}_N]$  and  $W_N = \mathbb{E}[\widehat{W}_N]$ . Assume that  $Z_N = 0$  and that there exist some constants  $K$ ,  $K'$  and  $v_N$  and  $v'_N$  (depending only on  $N$ ) such that  $\|\widehat{Z}_N\|_{2q} \leq v_N$ ,  $\|W_N - \widehat{W}_N\|_{4q} \leq v'_N$ ,  $\max_{1 \leq \ell \leq N} |Y_{\ell,N}| \leq K$  and  $|\widehat{Z}_N| \leq K'$ , then*

$$W_N \left\| \frac{\widehat{Z}_N}{\widehat{W}_N} \right\|_{2q} \leq v_N + \left( \frac{1}{W_N} K' + K \right) v'_N.$$

*Proof.* We use [15, Lemma 1] to establish that, for any  $0 \leq \alpha \leq 1$ ,

$$W_N \left| \frac{\widehat{Z}_N}{\widehat{W}_N} \right| \leq |\widehat{Z}_N| + \frac{|\widehat{Z}_N|}{W_N} |W_N - \widehat{W}_N| + \max_{1 \leq \ell \leq N} |Y_{\ell,N}| \frac{|W_N - \widehat{W}_N|^{1+\alpha}}{|W_N|^\alpha}.$$

By using this inequality with  $\alpha = 0$  and the Hölder inequality,

$$\begin{aligned} W_N \left\| \frac{\widehat{Z}_N}{\widehat{W}_N} \right\|_{2q} &\leq \|\widehat{Z}_N\|_{2q} + \frac{1}{W_N} \|\widehat{Z}_N\|_{4q} \|W_N - \widehat{W}_N\|_{4q} \\ &\quad + \left\| \max_{1 \leq \ell \leq N} |Y_{\ell,N}| |W_N - \widehat{W}_N| \right\|_{2q}. \end{aligned}$$

Hence

$$\begin{aligned} W_N \left\| \frac{\widehat{Z}_N}{\widehat{W}_N} \right\|_{2q} &\leq v_N + \frac{1}{W_N} K' v'_N + \left\| \max_{1 \leq \ell \leq N} |Y_{\ell,N}| |W_N - \widehat{W}_N| \right\|_{2q} \\ &\leq v_N + \frac{1}{W_N} K' v'_N + K v'_N, \end{aligned}$$

as  $\|W_N - \widehat{W}_N\|_{2q} \leq \|W_N - \widehat{W}_N\|_{4q} \leq v'_N$ .  $\square$

**Lemma 4.** *Assume that  $a_N$ ,  $b_N$ , and  $b$  are random variables defined on the same probability space such that there exist positive constants  $\beta$ ,  $B$ ,  $C$ , and  $M$  satisfying*

- (I)  $|a_N/b_N| \leq M$ ,  $\mathbb{P}$ -a.s. and  $b \geq \beta$ ,  $\mathbb{P}$ -a.s.,  
 (II) For all  $\epsilon > 0$  and all  $N \geq 1$ ,  $\mathbb{P}[|b_N - b| > \epsilon] \leq B e^{-CN\epsilon^2}$ ,  
 (III) For all  $\epsilon > 0$  and all  $N \geq 1$ ,  $\mathbb{P}[|a_N| > \epsilon] \leq B e^{-CN(\epsilon/M)^2}$ .

Then,

$$\mathbb{P}\left(\left|\frac{a_N}{b_N}\right| > \epsilon\right) \leq B \exp\left(-CN \left(\frac{\epsilon\beta}{2M}\right)^2\right).$$

*Proof.* See [9, Lemma 4]. □

**Lemma 5.** For  $T \geq 0$ , let  $(X_t)_{0 \leq t \leq T}$  be  $(T+1)$  random variables. Assume that there exist two constants  $A \geq 1$  and  $B > 0$  such that for all  $0 \leq t \leq T$  and all  $\epsilon > 0$

$$\mathbb{P}\{|X_t| > \epsilon\} \leq A e^{-B\epsilon}.$$

Then, for all  $0 < \gamma < 1$  and all  $\epsilon > 0$ , we have

$$\mathbb{P}\left\{\left|\sum_{t=0}^T X_t\right| > \epsilon\right\} \leq \frac{A}{1-\gamma} e^{-\gamma B\epsilon/(T+1)}.$$

*Proof.* By the Bienayme-Tchebychev inequality, we have

$$\begin{aligned} \mathbb{P}\left\{\left|\sum_{t=0}^T X_t\right| > \epsilon\right\} &= \mathbb{P}\left\{\exp\left[\frac{\gamma B}{T+1} \left|\sum_{t=0}^T X_t\right|\right] > e^{\gamma B\epsilon/(T+1)}\right\} \\ &\leq e^{-\gamma B\epsilon/(T+1)} \mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1} \left|\sum_{t=0}^T X_t\right|\right]\right]. \end{aligned} \quad (48)$$

It remains to bound the expectation in the RHS of (48) by  $\frac{A}{1-\gamma}$ . First, by a convexity inequality,

$$\begin{aligned} \mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1} \left|\sum_{t=0}^T X_t\right|\right]\right] &= \sum_{q=0}^{\infty} \frac{B^q \gamma^q}{q!(T+1)^q} \mathbb{E}\left[\left|\sum_{t=0}^T X_t\right|^q\right] \\ &\leq 1 + \sum_{q=1}^{\infty} \frac{B^q \gamma^q}{q!(T+1)^q} \sum_{t=0}^T \mathbb{E}[|X_t|^q]. \end{aligned}$$

Moreover, for  $q \geq 1$ ,  $\mathbb{E}[|X_t|^q]$  can be bounded by

$$\mathbb{E}[|X_t|^q] = \int_0^{\infty} \mathbb{P}\{|X_t| > \varepsilon^{1/q}\} d\varepsilon \leq A \int_0^{\infty} e^{-B\varepsilon^{1/q}} d\varepsilon = \frac{Aq!}{B^q},$$

Finally,

$$\mathbb{E}\left[\exp\left[\frac{\gamma B}{T+1} \left|\sum_{t=0}^T X_t\right|\right]\right] \leq A \sum_{q=0}^{\infty} \gamma^q = \frac{A}{1-\gamma}.$$

□

## References

- [1] O. Cappé. Recursive computation of smoothed functionals of hidden Markovian processes using a particle approximation. *Monte Carlo Methods Appl.*, 7(1–2):81–92, 2001.
- [2] O. Cappé, R. Douc, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing with application to parameter estimation in non-linear state space models. *Bernoulli*, 14(1):155–179, 2008. arXiv:math.ST/0609514.
- [3] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.
- [4] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.
- [5] P. Del Moral, A. Doucet, and S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5):947–975, 2010.
- [6] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l’Institut Henri Poincaré*, 37:155–194, 2001.
- [7] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV*, volume 1729 of *Lecture Notes in Math.*, pages 1–145. Springer, Berlin, 2000.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39(1):1–38 (with discussion), 1977.
- [9] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *accepted for publications in Ann. Appl. Probab.*, 4 2010.
- [10] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.
- [11] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10:197–208, 2000.
- [12] A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *Oxford handbook of nonlinear filtering*, 2009.
- [13] A. Doucet, N. Kantas, J.M. Maciejowski, and S.S. Singh. An overview of sequential monte carlo methods for parameter estimation in general state-space models. *Proceedings IFAC System Identification (SysId) Meeting*, 2009.
- [14] A. Doucet, G. Poyiadjis, and S.S. Singh. Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, 2010.
- [15] P. Doukhan and G. Lang. Evaluation for moments of a ratio with application to regression estimation. *Bernoulli*, 15(4):1259–1286, 2009.
- [16] J. Durbin and S. J. Koopman. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Roy. Statist. Soc. B*, 62:3–29, 2000.

- [17] S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.*, 50:438–449, 2004.
- [18] N. Gordon, D. Salmond, and A. F. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar Signal Process.*, 140:107–113, 1993.
- [19] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, New York, London, 1980.
- [20] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *J. Finance*, 42:281–300, 1987.
- [21] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.*, 7:175–193, 1998.
- [22] S. Kim, N. Shephard, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.*, 65:361–394, 1998.
- [23] G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, 1:1–25, 1996.
- [24] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, 94(446):590–599, 1999.
- [25] Michael K Pitt and Neil Shephard. Time varying covariances: a factor stochastic volatility approach (with discussion). In J M Bernardo, J O Berger, A P Dawid, and A F M Smith, editors, *Bayesian Statistics 6*, pages 547–570. Oxford University Press, Oxford, 1999.
- [26] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 1989.