



**HAL**  
open science

# On the Forward Filtering Backward Smoothing particle approximations of the smoothing distribution in general state spaces models

Randal Douc, Aurélien Garivier, Éric Moulines, Jimmy Olsson

► **To cite this version:**

Randal Douc, Aurélien Garivier, Éric Moulines, Jimmy Olsson. On the Forward Filtering Backward Smoothing particle approximations of the smoothing distribution in general state spaces models. 2009. hal-00370685

**HAL Id: hal-00370685**

**<https://hal.science/hal-00370685v1>**

Preprint submitted on 1 Apr 2009

**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.

ON THE FORWARD FILTERING BACKWARD SMOOTHING  
PARTICLE APPROXIMATIONS OF THE SMOOTHING  
DISTRIBUTION IN GENERAL STATE SPACES MODELS

BY RANDAL DOUC

*Institut Télécom/Télécom SudParis*

BY AURÉLIEN GARIVIER,

*CNRS/Télécom ParisTech, CNRS UMR 5141*

BY ERIC MOULINES,

*Institut Télécom/Télécom ParisTech, CNRS UMR 5141*

BY JIMMY OLSSON

*Center of Mathematical Sciences*

A prevalent problem in general state-space models is the approximation of the smoothing distribution of a state, or a sequence of states, conditional on the observations from the past, the present, and the future. The aim of this paper is to provide a rigorous foundation for the calculation, or approximation, of such smoothed distributions, and to analyse in a common unifying framework different schemes to reach this goal. Through a cohesive and generic exposition of the scientific literature we offer several novel extensions allowing to approximate joint smoothing distribution in the most general case with a cost growing linearly with the number of particles.

---

*Keywords and phrases:* Sequential Monte-carlo methods, particle filter, smoothing, Hidden Markov Models.

*AMS 2000 subject classifications:* Primary 60G10, 60K35; secondary 60G18

**1. Introduction.** Consider the nonlinear *state space model*, where the *state process*  $\{X_t\}_{t \geq 0}$  is a Markov chain on some general state space  $(X, \mathcal{B}(X))$  having initial distribution  $\chi$  and transition kernel  $Q$ . The state process is hidden but partially observed through the *observations*  $\{Y_t\}_{t \geq 0}$ , which are  $Y$ -valued random variables being independent conditionally on the latent state sequence  $\{X_t\}_{t \geq 0}$ ; in addition, there exists a  $\sigma$ -finite measure  $\lambda$  on  $(Y, \mathcal{B}(Y))$ , and a transition density function  $x \mapsto g(x, y)$ , referred to as the *likelihood*, such that  $\mathbb{P}(Y_t \in A | X_t) = \int_A g(X_t, y) \lambda(dy)$  for all  $A \in \mathcal{B}(Y)$ . The kernel  $Q$  and the likelihood function  $x \mapsto g(x, y)$  are assumed to be known. We shall consider the case in which the observations have arbitrary but fixed values  $y_{0:T} \stackrel{\text{def}}{=} [y_0, \dots, y_T]$ .

Statistical inference in general state space models involves computing the *posterior distribution* of a batch of state variables  $X_{s:s'}$  conditioned on a batch of observations  $Y_{t:T}$ , which we denote by  $\phi_{s:s'|t:T}$  (the dependence on the observations  $Y_{t:T}$  is implicit). The posterior distribution can be computed in closed form only in very specific cases, principally, when the state space model is linear and Gaussian or when the state space  $X$  is a finite set. In the vast majority of cases, nonlinearity or non-Gaussianity render analytic solutions intractable [1, 2, 17, 24].

These limitations have stimulated the interest in alternative strategies being able to handle more general state and measurement equations without putting strong a priori constraints on the behaviour of the posterior distributions. Among these, *Sequential Monte Carlo* (SMC) *methods* play a central role. SMC methods refer to a class of algorithms for approximating a *sequence of probability distributions* over a *sequence of probability spaces* by updating recursively a set of random *particles* with associated nonnegative *weights*. These algorithms can be seen as a combination of the *sequential importance sampling* and *sampling importance resampling* methods introduced in [15] and [25], respectively. SMC methods have emerged as a key tool for approximating state posterior distributions in general state space models; see, for instance, [11, 21, 22, 24] and the references therein.

The recursive formulas generating the filtering distribution  $\phi_{T|0:T}$  and the joint smoothing distributions  $\phi_{0:T|0:T}$  are closely related. Using the basic *filtering* version of the particle filter actually provides as a by-product an approximation of the joint smoothing distribution in the sense that the particle *paths* and their associated weights can be considered as a weighted sample approximating  $\phi_{0:T|0:T}$ . From these joint draws one may readily obtain fixed lag or fixed interval smoothed samples by simply extracting the required components from the sampled particle paths and retaining the same weights. This appealingly simple scheme can be used successfully for estimating the smoothing joint smoothing distribution for small values of  $T$  or any marginal smoothing distribution  $\phi_{s|0:T}$ , with  $s \leq T$ , when  $s$  and  $T$  are close; however, when  $T$  is large or when  $s$  and  $T$  are remote, the associated particle approximations are inaccurate [13].

In this article, we consider the forward filtering backward smoothing (FFBSm) algorithm and the forward filtering backward simulation (FFBSi) sampler. These algorithms share some similarities with the forward-backward algorithm for discrete state-space HMM. The FFBSm algorithm consists in reweighting,

in a backward pass, the weighted sample approximating the filtering distribution (see [18], [16], [12]). The FFBSi sample, conditionally independently to the particles and the weights obtained in the forward path, realizations of the joint smoothing fixed interval smoothing distribution; see [13].

The complexity of the FFBSm algorithm to estimate the marginal fixed interval smoothing distribution or of the original formulation of the FFBSi sampler grows generally as the square of the number of particles  $N$  multiplied by the time horizon  $T$ . This complexity can be linear in  $N$  for some specific examples. Otherwise, some tricky algorithms should be developed to overcome this problem, see for example [19]. Note that these computational techniques lead to algorithms with complexity of order  $N \log(N)$ , but this reduction in complexity comes at the price of introducing some level of approximations (truncation) which in practice introduce some bias which might be difficult to control. In this paper, a modification of the original FFBSi algorithm is presented, having a complexity which grows linearly in  $N$ , without having to truncate the density or to use intricate data structures.

The FFBSm and FFBSi algorithms are very challenging to analyze and, up to now, only a consistency result is available in [13] (the proof of this result being plagued by an error). The FFBSm estimate and the FFBSi trajectories explicitly depend upon all the particles and weights drawn before and after this time instant. It is therefore impossible to analyze directly the convergence of this approximation using the standard techniques developed to study the interacting particle approximations of the Feynman-Kac flows (see [5] or [9]).

The paper is organized as follows. In Section 2, the FFBSm algorithm and the FFBSi sampler are introduced. An exponential deviation inequality is first provided in Section 3 for the fixed-interval joint smoothing distribution. A Central Limit Theorem (CLT) for this quantity is then obtained in Section 4. Time-uniform exponential bounds are then computed for the FFBSm marginal smoothing distribution estimator, under mixing conditions on the kernel  $Q$ , in Section 5. Finally, under the same mixing condition, an explicit bound for the variance of the marginal smoothing distribution estimator is derived in Section 6.

*Notations and Definitions.* We denote  $a_{m:n} \stackrel{\text{def}}{=} (a_m, \dots, a_n)$  and  $[a_{m:n}, b_{p:q}] \stackrel{\text{def}}{=} (a_m, \dots, a_n, b_p, \dots, b_q)$ . We assume that all random variables are defined on a common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . A state space  $\mathsf{X}$  is said to be *general* if it is a Polish space and its topology is metrizable by some metric  $d$  such that  $(\mathsf{X}, d)$  is a complete separable metric space. We denote by  $\mathcal{B}(\mathsf{X})$  the associated Borel  $\sigma$ -algebra and by  $\mathbb{B}_b(\mathsf{X})$  the set of all bounded  $\mathcal{B}(\mathsf{X})/\mathcal{B}(\mathbb{R})$ -measurable functions from  $\mathsf{X}$  to  $\mathbb{R}$ . For any measure  $\mu$  on  $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$  and measurable function  $f$  satisfying  $\int_{\mathsf{X}} |f(x)| \mu(dx) < \infty$  we set  $\mu(f) = \int_{\mathsf{X}} f(x) \mu(dx)$ . Moreover, we say that two measures  $\mu$  and  $\nu$  are *proportional* (written  $\mu \propto \nu$ ) if they differ only by a normalization constant.

Let  $\mathsf{X}$  and  $\mathsf{Y}$  be two general state spaces. A kernel  $V$  from  $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$  to  $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$  is a map from  $\mathsf{X} \times \mathcal{B}(\mathsf{Y})$  into  $[0, 1]$  such that, for each  $A \in \mathcal{B}(\mathsf{Y})$ ,  $x \mapsto V(x, A)$  is a nonnegative bounded measurable function on  $\mathsf{X}$  and, for each  $x \in \mathsf{X}$ ,  $A \mapsto V(x, A)$  is a measure on  $\mathcal{B}(\mathsf{Y})$ . The function  $V(\cdot, f)$  belongs to  $\mathbb{B}(\mathsf{X})$  and we sometimes use the abridged notation  $Vf$  instead of  $V(\cdot, f)$ . For a

measure  $\nu$  on  $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ , we denote by  $\nu V$  the measure on  $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$  defined by, for any  $A \in \mathcal{B}(\mathsf{Y})$ ,  $\nu V(A) = \int_{\mathsf{X}} V(x, A) \nu(dx)$ .

For simplicity, we consider a *fully dominated* state space models for which there exists a  $\sigma$ -finite measure  $\nu$  on  $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$  such that, for all  $x \in \mathsf{X}$ ,  $Q(x, \cdot)$  has a transition probability density  $q(x, \cdot)$  with respect to  $\nu$ . For notational simplicity,  $\nu(dx)$  is sometimes replaced by  $dx$ .

For any initial distribution  $\chi$  on  $\mathsf{X}$  and any  $t \leq s \leq s' \leq T$ , denote by  $\phi_{\chi, s: s' | t: T}$  the posterior distribution of the state vector  $X_{s: s'}$  given the observations  $Y_{t: T}$  and knowing that  $X_0 \sim \chi$ . For all  $A \in \mathcal{B}(\mathsf{X})^{\otimes(s' - s + 1)}$ , this distribution may be expressed as

$$\phi_{\chi, s: s' | t: T}(A) = \frac{\int \cdots \int \phi_{\chi, t | t}(dx_t) \prod_{u=t+1}^T g_{u-1}(x_{u-1}) Q(x_{u-1}, dx_u) g_T(x_T) \mathbf{1}_A(x_{s: s'})}{\int \cdots \int \phi_{\chi, t | t}(dx_t) \prod_{u=t+1}^T g_{u-1}(x_{u-1}) Q(x_{u-1}, dx_u) g_T(x_T)},$$

with the convention  $\prod_{u=s}^t = 1$  if  $s > t$ . For simplicity, we will use the shorthand notations:

$$\begin{cases} \phi_{\chi, s | t: T} \stackrel{\text{def}}{=} \phi_{\chi, s: s | t: T}, \\ \phi_{\chi, s: s' | T} \stackrel{\text{def}}{=} \phi_{\chi, s: s' | 0: T}, \\ \phi_{\chi, s | T} \stackrel{\text{def}}{=} \phi_{\chi, s: s | 0: T}. \end{cases} \quad (1)$$

In fully dominated case, the smoothing distributions  $\phi_{\chi, s: s' | t: T}$  have densities (which we will denote similarly) with respect to the product measure  $\nu^{\otimes(s' - s + 1)}$ .

**2. Algorithms.** Conditionally on the observations  $Y_{0: T}$ , the state sequence  $\{X_s\}_{s \geq 0}$  is a time-inhomogeneous Markov chain. This property remains true in the *time-reversed* direction, i.e. given a strictly positive index  $T$ , initial distribution  $\chi$ , and index  $s \in \{0, \dots, T-1\}$ , for any  $f \in \mathbb{B}_b(\mathsf{X})$ ,

$$\mathbb{E}_{\chi} [f(X_s) | X_{s+1: T}, Y_{0: T}] = \mathbb{E}_{\chi} [f(X_s) | X_{s+1}, Y_{s: T}] = B_{\chi, s}(X_{s+1}, f).$$

where  $B_{\chi, s}(x_{s+1}, \cdot)$  is the *backward kernel*. In the fully dominated case, this kernel may be expressed as

$$B_{\chi, s}(x_{s+1}, A) = \int_A \frac{\phi_{\chi, s | s}(x) q(x, x_{s+1})}{\int \phi_{\chi, s | s}(x') q(x', x_{s+1}) dx'} \mathbf{1}_A(x) dx, \quad (2)$$

Using these notations, for any integers  $T > 0$ , index  $s \in \{0, \dots, T-1\}$  and initial probability  $\chi$ , the joint smoothing distribution may, for all  $f \in \mathbb{B}_b(\mathsf{X}^{T-s+1})$ , be recursively expressed as

$$\begin{aligned} \phi_{\chi, s: T | T}(f) &= \mathbb{E}_{\chi} [f(X_{s: T}) | Y_{0: T}] \\ &= \int \cdots \int f(x_{s: T}) B_{\chi, s}(x_{s+1}, dx_s) \phi_{\chi, s+1: T | T}(dx_{s+1: T}), \end{aligned} \quad (3)$$

with  $\phi_{\chi, T: T | T} = \phi_{\chi, T | T}$  being the filtering distribution at time  $T$ . If  $f$  depends on the first component  $x_s$  only, then Eq. (3) yields the marginal smoothing distribution, which is defined recursively by:

$$\phi_{\chi, s | T}(f) = \iint f(x_s) B_{\chi, s}(x_{s+1}, dx_s) \phi_{\chi, s+1 | T}(dx_{s+1}). \quad (4)$$

The method proposed by [12, 16] consists in approximating the smoothing distribution by storing the particles and associated weights obtained in a forward filtering pass and revising the weights in a backward smoothing pass. In the forward pass, particle approximations of the filtering distributions  $\phi_{\chi, s|s}$  are computed recursively for  $s = 0, \dots, T$ . Each approximation is formed by a set of particles  $\{\xi_s^i\}_{i=1}^N$  and associated importance weights  $\{\omega_s^i\}_{i=1}^N$  according to

$$\hat{\phi}_{\chi, s|s}(\mathrm{d}x) = \Omega_s^{-1} \sum_{i=1}^N \omega_s^i \delta_{\xi_s^i}(\mathrm{d}x), \quad (5)$$

where  $\Omega_s = \sum_{i=1}^N \omega_s^i$  and  $\delta_x$  denotes the Dirac mass located at  $x$ . There are several ways of producing such weighted samples  $\{(\xi_s^i, \omega_s^i)\}_{i=1}^N$ ; see [11], [21], [2], and the references therein. Most of these algorithms can be recasted into the common unifying framework of the *auxiliary particle filter*. Let  $\{\xi_0^i\}_{i=1}^N$  be i.i.d. random variables such that  $\xi_0^i \sim \rho_0$  and set  $\omega_0^i = \frac{d\chi}{d\rho_0}(\xi_0^i) g_0(\xi_0^i)$ . By classical importance sampling, the weighted sample  $\{(\xi_0^i, \omega_0^i)\}_{i=1}^N$  targets the distribution  $\phi_{\chi, 0|0}$ . Assume now that the weighted sample  $\{(\xi_{s-1}^i, \omega_{s-1}^i)\}_{i=1}^N$  targets  $\phi_{\chi, s-1|s-1}$ , i.e. for  $h \in \mathbb{B}_b(\mathsf{X})$ ,  $\Omega_{s-1}^{-1} \sum_{i=1}^N \omega_{s-1}^i h(\xi_{s-1}^i)$  is an estimate of  $\int \phi_{\chi, s-1|s-1}(\mathrm{d}x) h(x)$ . We may approximate  $\phi_{\chi, s|s}$  by replacing  $\phi_{\chi, s-1|s-1}$  in the forward filtering recursion

$$\phi_{\chi, s|s}(f) \propto \int \phi_{\chi, s-1|s-1}(\mathrm{d}x_{s-1}) q(x_{s-1}, x_s) g_s(x_s) f(x_s) \mathrm{d}x_s \quad (6)$$

by its particle approximation  $\hat{\phi}_{\chi, s-1|s-1}$ , leading to the target distribution

$$\hat{\phi}_{\chi, s|s}^{\mathrm{tar}}(\mathrm{d}x) \propto \sum_{i=1}^N \omega_{s-1}^i q(\xi_{s-1}^i, x) g_s(x) \mathrm{d}x. \quad (7)$$

To avoid an  $O(N^2)$  algorithm, [23] introduces an *auxiliary variable* corresponding to the selected particle index and target instead the probability density

$$\hat{\phi}_{\chi, s|s}^{\mathrm{aux}}(i, x_s) \propto \omega_{s-1}^i q(\xi_{s-1}^i, x_s) g_s(x_s) \quad (8)$$

on the product space  $\{1, \dots, N\} \times \mathsf{X}$ . Since  $\hat{\phi}_{\chi, s|s}^{\mathrm{tar}}$  is the marginal distribution of  $\hat{\phi}_{\chi, s|s}^{\mathrm{aux}}$  with respect to the particle index, we may sample from  $\hat{\phi}_{\chi, s|s}^{\mathrm{tar}}$  by simulating instead a set  $\{(I_s^i, \xi_s^i)\}_{i=1}^N$  of indices and particle positions from an instrumental distribution having probability density

$$\pi_{s|s}(i, x_s) \propto \omega_{s-1}^i \vartheta_s(\xi_{s-1}^i) p_s(\xi_{s-1}^i, x_s), \quad (9)$$

where  $\{\vartheta_s(\xi_{s-1}^i)\}_{i=1}^N$  are so-called *adjustment multiplier weights* and  $p_s$  is the *proposal transition* density function. Each draw  $(I_s^i, \xi_s^i)$  is assigned to the weight

$$\omega_s^i \stackrel{\mathrm{def}}{=} \frac{q(\xi_{s-1}^{I_s^i}, \xi_s^i) g_s(\xi_s^i)}{\vartheta_s(\xi_{s-1}^{I_s^i}) p_s(\xi_{s-1}^{I_s^i}, \xi_s^i)}, \quad (10)$$

which is proportional to  $\hat{\phi}_{\chi, s|s}^{\mathrm{aux}}(I_s^i, \xi_s^i) / \pi_{s|s}(I_s^i, \xi_s^i)$ . Hereafter, the indices are discarded and  $\{(\xi_s^i, \omega_s^i)\}_{i=1}^N$  is taken as an approximation of the target distribution

$\phi_{\mathcal{X},s|s}$ . The simplest choice, yielding to the so-called *bootstrap particle filter algorithm* proposed by [14], consists in setting, for all  $x \in \mathbf{X}$ ,  $\vartheta_s(x) \equiv 1$  and  $p_s(x, \cdot) = q(x, \cdot)$ . A more appealing choice from a theoretical standpoint –but often computationally costly– consists in setting  $\vartheta_s^*(x) = \int q(x, x_s) g_s(x_s) dx_s$  and

$$p_s^*(x, x_s) = \frac{q(x, x_s) g_s(x_s)}{\vartheta_s^*(x)}.$$

In this case, the importance weights  $\{\omega_s^i\}_{i=1}^N$  are all unity and the auxiliary particle filter is said to be *fully adapted*. Sampling from the fully adapted version of the auxiliary particle filter is in general difficult; the general method, based on the auxiliary accept-reject principle, proposed by [16] and [20] for sampling from these distributions is, with few exceptions, computationally involved. Other choices are discussed in [10] and [4].

2.1. *The Forward Filtering Backward Smoothing algorithm.* Following [13], the smoothing distribution can be approximated by filtering passes in the forward as well as the backward directions. Firstly, the particle filter is executed, while storing the weighted sample  $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$ ,  $1 \leq t \leq T$ ; secondly, starting with the particle approximation of the filtering distribution at time  $T$ , the importance weights are recursively updated backwards in time by combining particle estimates of the fixed interval smoothing distribution  $\phi_{\mathcal{X},s+1:T|T}$  and the filtering distribution estimate  $\phi_{\mathcal{X},s|s}$ . For  $1 \leq s \leq t \leq T$ , define  $\xi_{s:t}^{i_s:t} \stackrel{\text{def}}{=} (\xi_s^{i_s}, \dots, \xi_t^{i_t})$ . An approximation

$$\hat{\mathbb{B}}_{\mathcal{X},s}(x_{s+1}, dx_s) = \sum_{i=1}^N \frac{\omega_s^i q(\xi_s^i, x_{s+1})}{\sum_{\ell=1}^N \omega_s^\ell q(\xi_s^\ell, x_{s+1})} \delta_{\xi_s^i}(dx_s) \quad (11)$$

of the backward kernel can be obtained by revising the weights  $\{\omega_s^i\}_{i=1}^N$  without moving the particles  $\{\xi_s^i\}_{i=1}^N$ . If in addition the joint smoothing distribution  $\phi_{\mathcal{X},s+1:T|T}$  is approximated at time  $s+1$  using the weighted sample  $\{(\xi_{s+1:T}^{j_{s+1:T}}, \omega_{s+1|T}^{j_{s+1:T}})\}$ ,  $j_{s+1:T} \in \{1, \dots, N\}^{T-s}$ , i.e.

$$\hat{\phi}_{\mathcal{X},s+1:T|T}(dx_{s+1:T}) \propto \sum_{j_{s+1:T}=1}^N \omega_{s+1|T}^{j_{s+1:T}} \delta_{\xi_{s+1:T}^{j_{s+1:T}}}(dx_{s+1:T}), \quad (12)$$

then we may substitute this and the approximation (11) of the backward kernel into (3) to obtain

$$\hat{\phi}_{\mathcal{X},s:T|T}(dx_{s:T}) \propto \sum_{j_{s:T}=1}^N \omega_{s|T}^{j_{s:T}} \delta_{\xi_{s:T}^{j_{s:T}}}(dx_{s:T}),$$

where the new weight is recursively updated according to

$$\omega_{s|T}^{j_{s:T}} = \frac{\omega_s^{j_s} q(\xi_s^{j_s}, \xi_{s+1}^{j_{s+1}})}{\sum_{\ell=1}^N \omega_s^\ell q(\xi_s^\ell, \xi_{s+1}^{j_{s+1}})} \omega_{s+1|T}^{j_{s+1:T}}. \quad (13)$$

The estimator of the joint smoothing distribution may be rewritten as

$$\hat{\phi}_{\chi, s:T|T}(h) \stackrel{\text{def}}{=} \sum_{i_{s:T}=1}^N \frac{\varpi_{s:T}^{i_{s:T}} \omega_T^{i_T}}{\sum_{j_{s:T}=1}^N \varpi_{s:T}^{j_{s:T}} \omega_T^{j_T}} h(\xi_{s:T}^{i_{s:T}}) \quad (14)$$

where

$$\varpi_{s:t}^{i_{s:t}} \stackrel{\text{def}}{=} \prod_{u=s+1}^t \frac{\omega_{u-1}^{i_{u-1}} q(\xi_{u-1}^{i_{u-1}}, \xi_u^{i_u})}{\sum_{\ell=1}^N \omega_{u-1}^{\ell} q(\xi_{u-1}^{\ell}, \xi_u^{i_u})}, \quad (15)$$

with the convention  $\prod_a^b = 1$  if  $a > b$  so that  $\varpi_{s:s}^{i_{s:s}} = 1$ . Since  $\sum_{j_{s:T-1}} \varpi_{s:t}^{j_{s:t}} = 1$ ,  $\hat{\phi}_{\chi, s:T|T}(h)$  may be alternatively expressed as

$$\hat{\phi}_{\chi, s:T|T}(h) = \sum_{i_{s:T}=1}^N \frac{\varpi_{s:T}^{i_{s:T}} \omega_T^{i_T}}{\sum_{j_T=1}^N \omega_T^{j_T}} h(\xi_{s:T}^{i_{s:T}}). \quad (16)$$

This estimate of the joint smoothing distribution may be understood as an approximate importance sampling estimator. The estimator  $\hat{\phi}_{\chi, s:T|T}(h)$  is highly impractical, because its support is the set of  $N^{T-s+1}$  possible particle paths  $\{\xi_{s:T}^{j_{s:T}}\}$ . Nevertheless, this estimator plays a key role in the theoretical derivations.

The importance weight of these path particles is computed as if the path particle  $\xi_{s:T}^{j_{s:T}}$  were simulated by drawing forward in time, for  $s < t \leq T$ ,  $\xi_t^{j_t}$  in the set  $\{\xi_t^i\}_{i=1}^N$ , conditionally independently from  $\{\xi_{s:t-1}^{j_{s:t-1}}\}$  from the distribution

$$\Omega_{t-1}^{-1} \sum_{\ell=1}^N \omega_{t-1}^{\ell} q(\xi_{t-1}^{\ell}, \cdot), \quad i = 1, \dots, N, \quad (17)$$

which approximates the predictive distribution  $\phi_{\chi, t|t-1}$ . Of course, the distribution of  $\xi_{s:T}^{j_{s:T}}$  is not exactly the product of the marginal distribution (17), because the particle position are not independent -this approximation would be approximately correct for a finite block of particles selected randomly, using propagation of chaos property; see e.g. [5, chapter 8] -. This is why standard results on importance sampling estimators cannot be applied to that context.

Most often, it is not required to compute the joint smoothing definition but rather the marginal smoothing distribution  $\phi_{\chi, s|T}$  (or more generally some fixed dimensional marginal of the joint smoothing,  $\phi_{\chi, s:s+\Delta|T}$  for a positive integer  $\Delta$ ). Approximations of the marginal smoothing distributions may be obtained by associating to the set of particle  $\{\xi_s^{j_s}\}$ ,  $j_s \in \{1, \dots, N\}^{\Delta+1}$  the weights obtained by marginalizing the joint smoothing weights  $\{\omega_{s|T}^{j_{s:T}}\}$  over the components  $j_{s+1:T} \in \{1, \dots, N\}^{T-s-\Delta+1}$ ,  $\omega_{s|T}^{j_s} = \sum_{j_{s+1:T}=1}^N \omega_{s|T}^{j_{s:T}}$ . It is easily seen that these marginal weights can be recursively updated as follows:

$$\omega_{s|T}^i = \sum_{j=1}^N \frac{\omega_s^i q(\xi_s^i, \xi_{s+1}^j)}{\sum_{\ell=1}^N \omega_s^{\ell} q(\xi_s^{\ell}, \xi_{s+1}^j)} \omega_{s+1|T}^j, \quad i = 1, \dots, N \quad (18)$$

The complexity of this estimator of the marginal smoothing distribution is  $O(N^2T)$ , which is manageable only if the number of particles is moderate. When



the dimension of the input space is not too large, this computational cost can considerably be reduced to  $N \log(N)$ , but at the price of truncating the distribution and therefore introducing some amount of bias (see for example [19]). Note that, in certain specific scenarios (such as discrete Markov chains over large state space with sparse transition matrix), the complexity can even be reduced to  $O(NT)$ .

*2.2. The Forward Filtering Backward Simulation.* Another way of understanding (14) consists in noting that the importance weight (13) is a probability distribution over  $\{1, \dots, N\}^{T-s}$ ; more precisely,  $\omega_{s+1|T}^{j_{s+1:T}} = \mathbb{P}(J_{s+1:T} = j_{s+1:T} | \mathcal{F}_T)$ , where  $\mathcal{F}_s \stackrel{\text{def}}{=} \sigma\{(\xi_t^i, \omega_t^i); 0 \leq t \leq s, 1 \leq i \leq N\}$  and  $\{J_u\}_{u=s}^T$  is a reversed Markov chain with a final distribution  $\omega_T^i / \Omega_T$ ,  $i = 1, \dots, N$  and backward transition matrix  $\{\hat{B}_{i,j}\}_{i,j=1}^N$ ,  $\mathbb{P}(J_s = j | \mathcal{F}_T \vee \sigma(J_{s+1})) = \hat{B}_{J_s, j}$ , with

$$\hat{B}_{i,j} = \frac{\omega_s^j q(\xi_s^j, \xi_{s+1}^i)}{\sum_{u=1}^N \omega_s^u q(\xi_s^u, \xi_{s+1}^i)}, \quad i, j = 1, \dots, N \quad . \quad (19)$$

With these definitions, the joint smoothing distribution may be written as the conditional expectation

$$\hat{\phi}_{\chi, s:T|T}(h) = \mathbb{E} \left[ h \left( \xi_{s:T}^{J_{s:T}} \right) \middle| \mathcal{F}_T \right] . \quad (20)$$

The idea of simulating the indices  $J_{s:T}$  backward in time to draw approximately from the smoothing distribution  $\hat{\phi}_{\chi, s:T|T}$ , has been proposed in [13] (Algorithm 1, pp. 158). This algorithm proceeds recursively backward in time as follows. At time  $T$ , we draw conditionally independently from an  $N$  indices  $\{J_T^\ell\}_{\ell=1}^N$  from the distribution  $\{\omega_T^\ell\}_{\ell=1}^N$  (for ease of notations, we draw the same number of particles in the forward and backward passes, but there is no need to do that). Given now a  $N$  sample  $\{J_{s+1:T}^\ell\}_{\ell=1}^N \in \{1, \dots, N\}^{T-s}$ , we draw conditionally independently  $J_s^\ell \in \{1, \dots, N\}$ ,  $\ell = 1, \dots, N$  from the distributions  $\{\hat{B}_{j, J_{s+1}^\ell}\}_{j=1}^N$ . This algorithm is referred in the sequel to as the forward filtering backward simulation algorithm (FFBSi). This sample yields to the following (practical) estimator of the joint fixed-interval smoothing distribution:

$$\tilde{\phi}_{\chi, s:T|T}(h) = N^{-1} \sum_{\ell=1}^N h \left( \xi_{s:T}^{J_{s:T}^\ell} \right), \quad h \in \mathbb{B}_b(\mathcal{X}^{T-s+1}). \quad (21)$$

The computational complexity for each individual realization is  $O(N)$  at each time step, so the overall computational effort to estimate  $\tilde{\phi}_{\chi, 0:T|T}$  is therefore  $O(N^2T)$ . Using the methods introduced by [19], this complexity can be further reduced to  $O(N \log(N)T)$ , but here again at the price of some additional approximations. It is easy to modify this algorithm to make it linear in  $N$ . Assume that the transition kernel  $q$  is bounded,  $q(x, x') \leq |q|_\infty$ . Since  $\omega_s^j q(\xi_s^j, \xi_{s+1}^i) \leq |q|_\infty \omega_s^j$ , for any  $i, j \in \{1, \dots, N\}$ , we may sample (19) using the accept-reject mechanism. For any  $\ell = 1, \dots, N$ , we sample independently indices  $I_s^{\ell, u}$ ,  $u = 1, 2, \dots$  from the distribution  $\{\Omega_s^{-1} \omega_s^j\}_{j=1}^N$  and uniform random variables  $U_s^{\ell, u}$  on  $[0, 1]$  and let  $J_s^\ell = I_s^{\ell, \tau_s^\ell}$  where  $\tau_s^\ell$  is the first index  $u$  for which  $U_s^{\ell, u} \leq q(\xi_s^{I_s^{\ell, u}}, \xi_{s+1}^{J_{s+1}^\ell}) / |q|_\infty$ . The complexity of the resulting algorithm is linear instead of quadratic in  $N$ .

**3. Exponential deviation inequality.** In this section, we establish the properties of the forward filtering backward smoothing algorithm. We first establish a non asymptotic deviation inequality. For any function  $f : \mathcal{X}^d \rightarrow \mathbb{R}$ , we define  $|f|_\infty = \sup_{x \in \mathcal{X}^d} |f(x)|$  and  $\text{osc}(f) = \sup_{(x, x') \in \mathcal{X}^d \times \mathcal{X}^d} |f(x) - f(x')|$ . Denote  $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$  and consider the following assumptions. We denote by  $T$  the horizon, which can be either a finite integer or infinite.

**A 1.**  $\sup_{0 \leq t \leq T} |g_t|_\infty < \infty$ .

Define for  $t \geq 0$  the importance weight functions:

$$\omega_0(x) = \frac{d\chi}{d\rho_0}(x)g_0(x) \quad \text{and} \quad \omega_t(x, x') \stackrel{\text{def}}{=} \frac{q(x, x')g_t(x')}{\vartheta_t(x)p_t(x, x')}, t \geq 1. \quad (22)$$

**A 2.**  $\sup_{0 \leq t \leq T} |\vartheta_t|_\infty < \infty$  and  $\sup_{0 \leq t \leq T} |\omega_t|_\infty < \infty$ .

The latter assumption is rather mild. It holds in particular under (A1) for the bootstrap filter ( $q = p_t$  and  $\vartheta_t \equiv 1$ ). It automatically holds in the fully adapted case ( $\omega_t \equiv 1$ ).

The first step in our proof consists in obtaining an exponential deviation inequality of Hoeffding type for the auxiliary particle approximations of the forward filtering distribution  $\phi_{\mathcal{X}, t|t}$ . Such results can be adapted from [5, Chapter 7], using the Feynman-Kac representation of the auxiliary filter. For the sake of completeness, we prove these results explicitly. By convention, we set  $\phi_{\mathcal{X}, 0|-1} = \chi$  and  $\vartheta_0 \equiv 1$ .

**PROPOSITION 1.** *Assume that A1–2. Then, for all  $t \in \{0, \dots, T\}$ , there exist  $0 < B, C < \infty$  such that for all  $N, \epsilon > 0$ , and all measurable functions  $h$ ,*

$$\mathbb{P} \left[ \left| N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) - \frac{\phi_{\mathcal{X}, t|t-1}(g_t h)}{\phi_{\mathcal{X}, t-1|t-1}(\vartheta_t)} \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2/|h|_\infty^2}, \quad (23)$$

$$\mathbb{P} \left[ \left| \hat{\phi}_{\mathcal{X}, t|t}(h) - \phi_{\mathcal{X}, t|t}(h) \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2/\text{osc}^2(h)}, \quad (24)$$

where the weighted sample  $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N$  is defined in (10).

**PROOF.** We prove (23) and (24) together by induction on  $t \geq 0$ . First note that, by construction,  $\{(\xi_t^i, \omega_t^i)\}_{1 \leq i \leq N}$  are i.i.d. conditionally to the  $\sigma$ -field  $\mathcal{F}_{t-1} \stackrel{\text{def}}{=} \sigma\{(\xi_s^i, \omega_s^i); 0 \leq s \leq t-1, 1 \leq i \leq N\}$ . Under (A2), we may therefore apply the Hoeffding inequality, which implies,

$$\mathbb{P} \left[ \left| N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) - \mathbb{E} \left[ N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) \middle| \mathcal{F}_{t-1} \right] \right| > \epsilon \right] \leq 2e^{-2N\epsilon^2/(\omega_t|_\infty^2 |h|_\infty^2)}. \quad (25)$$

For  $t = 0$ ,

$$\mathbb{E} \left[ N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) \middle| \mathcal{F}_{t-1} \right] = \mathbb{E} \left[ \omega_t^1 h(\xi_t^1) \middle| \mathcal{F}_{t-1} \right] = \chi(g_0 h) = \phi_{\mathcal{X}, 0|-1}(g_0 h).$$

Thus, (24) follows by Lemma 11 applied with  $a_N = N^{-1} \sum_{i=1}^N \omega_0^i h(\xi_0^i)$ ,  $b_N = N^{-1} \sum_{i=1}^N \omega_0^i$ ,  $c_N = 0$  and  $b = \beta = \chi(g_0)$ , conditions (I), (II) and (III) being obviously satisfied. For  $t \geq 1$ , we prove (23) by deriving an exponential inequality for  $\mathbb{E} \left[ N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) \mid \mathcal{F}_{t-1} \right]$  thanks to the induction assumption. It follows from the definition that

$$\begin{aligned} \mathbb{E} \left[ N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) \mid \mathcal{F}_{t-1} \right] &= \mathbb{E} \left[ \omega_t^1 h(\xi_t^1) \mid \mathcal{F}_{t-1} \right] \\ &= \sum_{i=1}^N \int \frac{\omega_{t-1}^i \vartheta_t(\xi_{t-1}^i) p_t(\xi_{t-1}^i, x)}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} \frac{q(\xi_{t-1}^i, x) g_t(x)}{\vartheta_t(\xi_{t-1}^i) p_t(\xi_{t-1}^i, x)} h(x) dx, \\ &= \frac{\sum_{i=1}^N \omega_{t-1}^i \int Q(\xi_{t-1}^i, dx) g_t(x) h(x)}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} \end{aligned} \quad (26)$$

We apply Lemma 11 by successively checking conditions (I), (II) and (III) with

$$\begin{cases} a_N \stackrel{\text{def}}{=} \Omega_t^{-1} \sum_{i=1}^N \omega_{t-1}^i \int Q(\xi_{t-1}^i, dx) g_t(x) h(x) \\ b_N \stackrel{\text{def}}{=} \Omega_t^{-1} \sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell) \\ c_N \stackrel{\text{def}}{=} \phi_{\chi, t-1|t-1} [\int Q(\cdot, dx) g_t(x) h(x)] = \phi_{\chi, t|t-1}(g_t h) \\ d_N \stackrel{\text{def}}{=} b \stackrel{\text{def}}{=} \beta \stackrel{\text{def}}{=} \phi_{\chi, t-1|t-1}(\vartheta_t) \end{cases}$$

We have that

$$\begin{aligned} \left| \frac{a_N}{b_N} \right| &= \left| \mathbb{E} \left[ \omega_t^1 h(\xi_t^1) \mid \mathcal{F}_{t-1} \right] \right| \leq |\omega_t|_\infty |h|_\infty \\ \left| \frac{c_N}{d_N} \right| &= \left| \phi_{\chi, t-1|t-1} \left[ \vartheta_t(\cdot) \int \frac{q(\cdot, x) g_t(x)}{\vartheta_t(\cdot) p_t(\cdot, x)} p_t(\cdot, x) h(x) dx \right] / \phi_{\chi, t-1|t-1}(\vartheta_t) \right| \\ &\leq |\omega_t|_\infty |h|_\infty. \end{aligned}$$

Thus, condition (I) is satisfied. Now, assume that the induction assumption (24) holds where  $t$  is replaced by  $t-1$ . Then,

$$a_N - \frac{c_N}{d_N} b_N = \Omega_t^{-1} \sum_{i=1}^N \omega_{t-1}^i H(\xi_{t-1}^i)$$

with  $H(\xi_{t-1}^i) \stackrel{\text{def}}{=} \int Q(\xi_{t-1}^i, dx) g_t(x) h(x) - \frac{\phi_{\chi, t-1|t-1} [\int Q(\cdot, dx) g_t(x) h(x)]}{\phi_{\chi, t-1|t-1}(\vartheta_t)} \vartheta_t(\xi_{t-1}^i)$ . And by noting that  $\phi_{\chi, t-1|t-1}(H) = 0$ , exponential inequalities for  $a_N - (c_N/d_N)b_N$  and  $b_N - b$  are then directly derived from the induction assumption under (A1-2). Thus Lemma 11 applies and finally (23) is proved for  $t \geq 1$ .

To conclude, it remains to see why (23) implies (24). Without loss of generality, we assume that  $\phi_{\chi, t|t}(h) = 0$  holds. An exponential inequality for  $\Omega_t^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i)$  is obtained by applying Lemma 1 with

$$\begin{cases} a_N \stackrel{\text{def}}{=} N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) \\ b_N \stackrel{\text{def}}{=} N^{-1} \sum_{i=1}^N \omega_t^i \\ c_N \stackrel{\text{def}}{=} 0 \\ b \stackrel{\text{def}}{=} \beta \stackrel{\text{def}}{=} \frac{\phi_{\chi, t-1|t-1} [\int Q(\cdot, dx) g_t(x)]}{\phi_{\chi, t-1|t-1}(\vartheta_t)} \end{cases}$$

where conditions (I), (II), (III) are obviously derived from the (23) since the condition  $\phi_{\chi,t|t}(h) = 0$  directly implies that

$$\phi_{\chi,t-1|t-1} \left[ \int Q(\cdot, dx) g_t(x) h(x) \right] = 0$$

□

Using the exponential deviation inequality for the auxiliary particle approximation of the filtering density, it is now possible to derive an exponential inequality for the forward filtering backward smoothing approximation of the joint smoothing distribution.

**THEOREM 2.** *Assume A1–2. Let  $1 \leq s \leq T$ . There exist  $0 < B, C < \infty$  such that for all  $N, \epsilon > 0$ , and all measurable functions  $h$ ,*

$$\mathbb{P} \left[ \left| \hat{\phi}_{\chi,s:T|T}(h) - \phi_{\chi,s:T|T}(h) \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2 / \text{osc}^2(h)}, \quad (27)$$

$$\mathbb{P} \left[ \left| \tilde{\phi}_{\chi,s:T|T}(h) - \phi_{\chi,s:T|T}(h) \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2 / \text{osc}^2(h)}. \quad (28)$$

where  $\hat{\phi}_{\chi,s:T|T}(h)$  and  $\tilde{\phi}_{\chi,s:T|T}(h)$  are defined in (14) and (21).

**PROOF.** Using (20) and the definition of  $\hat{\phi}_{\chi,s:T|T}(h)$ , we may write

$$\tilde{\phi}_{\chi,s:T|T}(h) - \hat{\phi}_{\chi,s:T|T}(h) = N^{-1} \sum_{\ell=1}^N \left[ h \left( \xi_{s:T}^{J_{\ell}} \right) - \mathbb{E} \left[ h \left( \xi_{s:T}^{J_{\ell}} \right) \mid \mathcal{F}_T \right] \right],$$

which implies (28) by the Hoeffding inequality and (27). We now prove (27). Let  $s \leq t < T$ . For  $h$  a measurable function defined on  $\mathsf{X}^{T-s+1}$ , define the kernels  $L_{s,t,T}(\xi_{s:t}^{i_{s:t}}, h) \stackrel{\text{def}}{=} h(\xi_{s:t}^{i_{s:t}})$  and, for  $s \leq t < T$ ,

$$L_{s,t,T}(\xi_{s:t}^{i_{s:t}}, h) \stackrel{\text{def}}{=} \int \cdots \int Q(\xi_t^{i_t}, dx_{t+1}) g_{t+1}(x_{t+1}) \times \prod_{u=t+1}^{T-1} Q(x_u, dx_{u+1}) g_{u+1}(x_{u+1}) h([\xi_{s:t}^{i_{s:t}}, x_{t+1:T}]). \quad (29)$$

By construction,  $L_{s,t,T}$  can be obtained recursively *backwards* in time as follows:

$$L_{s,t-1,T}(\xi_{s:t-1}^{i_{s:t-1}}, h) = \int Q(\xi_{t-1}^{i_{t-1}}, dx) g_t(x) L_{s,t,T}([\xi_{s:t-1}^{i_{s:t-1}}, x], h) \quad (30)$$

Denote by  $\mathbf{1}$  the function identically equal to one. Under (A1),  $|L_{s,t,T}(\cdot, h)|_{\infty} \leq |L_{s,t,T}(\cdot, \mathbf{1})|_{\infty} |h|_{\infty}$ , and  $|L_{s,t,T}(\cdot, \mathbf{1})|_{\infty} < \infty$ . Denote

$$F_{s,s,T}(\xi, h) \stackrel{\text{def}}{=} L_{s,s,T}(\xi, h), \quad (31)$$

and for all  $t \in \{s+1, \dots, T\}$ , set

$$F_{s,t,T}(\xi, h) \stackrel{\text{def}}{=} \sum_{i_{s:t-1}=1}^N \varpi_{s:t-1}^{i_{s:t-1}} \frac{\omega_{t-1}^{i_{t-1}} q(\xi_{t-1}^{i_{t-1}}, \xi)}{\sum_{\ell=1}^N \omega_{t-1}^{\ell} q(\xi_{t-1}^{\ell}, \xi)} L_{s,t,T}([\xi_{s:t-1}^{i_{s:t-1}}, \xi], h), \quad (32)$$

where  $\varpi_{s:t-1}^{i:s:t-1}$  is defined in (15). Define, for  $t \in \{s, \dots, T\}$ ,

$$A_{s,t,T}(h) \stackrel{\text{def}}{=} \sum_{\ell=1}^N \omega_t^\ell F_{s,t,T}(\xi_t^\ell, h). \quad (33)$$

Without loss of generality, we assume that  $\phi_{\chi,s:T|T}(h) = 0$ . With the notations introduced above,  $\hat{\phi}_{\chi,s:T|T}(h)$  may be expressed as the sum

$$\hat{\phi}_{\chi,s:T|T}(h) = \frac{A_{s,T,T}(h)}{A_{s,T,T}(\mathbf{1})} = \frac{A_{s,s,T}(h)}{A_{s,s,T}(\mathbf{1})} + \sum_{t=s+1}^T \left\{ \frac{A_{s,t,T}(h)}{A_{s,t,T}(\mathbf{1})} - \frac{A_{s,t-1,T}(h)}{A_{s,t-1,T}(\mathbf{1})} \right\}. \quad (34)$$

We now compute an exponential bound for the terms appearing the RHS of the previous identity. Note that

$$\frac{A_{s,s,T}(h)}{A_{s,s,T}(\mathbf{1})} = \frac{\Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell L_{s,s,T}(\xi_s^\ell, h)}{\Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell L_{s,s,T}(\xi_s^\ell, \mathbf{1})}$$

We apply Lemma 11 by successively checking conditions (I), (II) and (III) with  $a_N = \Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell L_{s,s,T}(\xi_s^\ell, h)$ ,  $b_N = \Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell L_{s,s,T}(\xi_s^\ell, \mathbf{1})$ ,  $c_N = 0$ , and  $b = \beta = \phi_{\chi,s|s}(L_{s,s,T}(\cdot, \mathbf{1}))$ . It follows immediately from the definition that  $\left| \frac{a_N}{b_N} \right| \leq |h|_\infty$ . Moreover,  $\phi_{\chi,s:T|T}(h) = 0$  implies that  $\phi_{\chi,s|s}(L_{s,s,T}(\cdot, h)) = 0$ , conditions (II) and (III) are then directly derived from the exponential inequality for the auxiliary filter (see Proposition 1, Eq. (24)).

We now establish an exponential inequality for

$$A_{s,t,T}(h)/A_{s,t,T}(\mathbf{1}) - A_{s,t-1,T}(h)/A_{s,t-1,T}(\mathbf{1})$$

using again Lemma 11. We take  $a_N = N^{-1} A_{s,t,T}(h)$ ,  $b_N = N^{-1} A_{s,t,T}(\mathbf{1})$ ,  $c_N = N^{-1} A_{s,t-1,T}(h)$ ,  $d_N = N^{-1} A_{s,t-1,T}(\mathbf{1})$  and

$$b = \beta = \frac{\phi_{\chi,t-1|t-1}(L_{s,t-1,T}(\cdot, \mathbf{1}))}{\phi_{\chi,t-1|t-1}(\vartheta_t)}. \quad (35)$$

By definition,  $|a_N/b_N| \leq |h|_\infty$  and  $|c_N/d_N| \leq |h|_\infty$ , showing Lemma 11, condition (I). We now check condition (II). The function  $\xi_{s:t}^{i:s:t} \mapsto L_{s,t,T}(\xi_{s:t}^{i:s:t}, \mathbf{1})$  depends only on  $\xi_t^{it}$ ; with a slight abuse of notation, we set  $L_{s,t,T}(\xi_{s:t}^{i:s:t}, \mathbf{1}) = L_{t,t,T}(\xi_t^{it}, \mathbf{1})$ . It follows from the definition (32) that  $F_{s,t,T}(\xi, \mathbf{1}) = L_{t,t,T}(\xi, \mathbf{1})$ . Plugging this into the definition of  $A_{s,t,T}(\mathbf{1})$  yields:  $A_{s,t,T}(\mathbf{1}) = \sum_{\ell=1}^N \omega_t^\ell L_{t,t,T}(\xi_t^\ell, \mathbf{1})$ . Condition (II) follows from Proposition 1-Eq. (23). Finally, we check condition (III). Write  $a_N - \frac{c_N}{d_N} b_N = N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{s,t,T}(\xi_t^\ell, h)$  where

$$G_{s,t,T}(\xi, h) \stackrel{\text{def}}{=} F_{s,t,T}(\xi, h) - \frac{A_{s,t-1,T}(h)}{A_{s,t-1,T}(\mathbf{1})} F_{s,t,T}(\xi, \mathbf{1}). \quad (36)$$

Since  $\{(\xi_t^\ell, \omega_t^\ell)\}_{\ell=1}^N$  are i.i.d. conditionally to the  $\sigma$ -field  $\mathcal{F}_{t-1}$ , we have that  $\{\omega_t^\ell G_{s,t,T}(\xi_t^\ell, h)\}_{\ell=1}^N$  are also i.i.d. conditionally to  $\mathcal{F}_{t-1}$ . That allows to apply the conditional Hoeffding's inequality to  $N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{s,t,T}(\xi_t^\ell, h)$  provided that

we have first checked that  $\mathbb{E}[\eta^1 | \mathcal{F}_{t-1}] = 0$  and that  $(\omega_t^\ell G_{s,t,T}(\xi_t^\ell, h))_{1 \leq \ell \leq N}$  are bounded random variables. By (26), for any bounded function  $f$ ,

$$\mathbb{E} \left[ \omega_t^1 f(\xi_t^1) \mid \mathcal{F}_{t-1} \right] = \int \frac{\sum_{i=1}^N \omega_{t-1}^i q(\xi_{t-1}^i, x)}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} g_t(x) f(x) dx$$

Applying this relation with  $f(\cdot) = F_{s,t,T}(\cdot, h)$  and using the recursion (30), the previous relation implies by direct calculation

$$\mathbb{E} \left[ \omega_t^1 F_{s,t,T}(\xi_t^1, h) \mid \mathcal{F}_{t-1} \right] = \frac{\sum_{\ell=1}^N \omega_{t-1}^\ell F_{s,t-1,T}(\xi_{t-1}^\ell, h)}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)}. \quad (37)$$

Therefore, since  $A_{s,t-1,T}(h)/A_{s,t-1,T}(\mathbf{1})$  is  $\mathcal{F}_{t-1}$ -measurable,

$$\begin{aligned} & \mathbb{E} \left[ \omega_t^1 G_{s,t,T}(\xi_t^1, h) \mid \mathcal{F}_{t-1} \right] \\ &= \frac{\sum_{\ell=1}^N \omega_{t-1}^\ell F_{s,t-1,T}(\xi_{t-1}^\ell, h)}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} - \frac{A_{s,t-1,T}(h)}{A_{s,t-1,T}(\mathbf{1})} \frac{\sum_{\ell=1}^N \omega_{t-1}^\ell F_{s,t-1,T}(\xi_{t-1}^\ell, \mathbf{1})}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} = 0, \end{aligned} \quad (38)$$

by definition (33). Moreover, since  $|F_{s,t,T}(x, h)| \leq F_{s,t,T}(x, \mathbf{1}) |h|_\infty$ , we have  $|A_{s,t,T}(h)/A_{s,t,T}(\mathbf{1})| \leq |h|_\infty$ , showing that

$$\begin{aligned} |\omega_t(x, x') G_{s,t,T}(x', h)| &= \left| \omega_t(x, x') \left( F_{s,t,T}(x', h) - \frac{A_{s,t-1,T}(h)}{A_{s,t-1,T}(\mathbf{1})} F_{s,t,T}(x', \mathbf{1}) \right) \right| \\ &\leq 2 |\omega_t|_\infty |F_{s,t,T}(\cdot, \mathbf{1})|_\infty |h|_\infty \leq 2 |\omega_t|_\infty |L_{t,t,T}(\cdot, \mathbf{1})|_\infty |h|_\infty < \infty. \end{aligned} \quad (39)$$

The Hoeffding inequality therefore implies:

$$\begin{aligned} \mathbb{P} \left( \left| a_N - \frac{c_N}{d_N} b_N \right| > \epsilon \right) &= \mathbb{P} \left( \left| N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{s,t,T}(\xi_t^\ell, h) \right| > \epsilon \right) \\ &\leq B \exp \left\{ -CN \left( \frac{\epsilon}{|h|_\infty} \right)^2 \right\}, \end{aligned}$$

showing condition (III) and concluding the proof.  $\square$

**4. Asymptotic normality.** We now derive a Central Limit Theorem (CLT) for the forward-filtering backward-smoothing estimator (14). Consider the following assumption.

**A 3.** for all  $t \in \{1, \dots, T\}$  and  $M > 0$ ,  $\int \sup_{|x| \leq M} p_t(x, x') dx' < \infty$ .

We first recall that, under assumption (A1-2) the auxiliary particle filter approximation of the filtering distribution satisfies a CLT (see for example [10, Theorem 3.2]). For any bounded measurable function  $h : \mathbf{X} \rightarrow \mathbb{R}$ , define the kernel

$$L_{s,t}(x, h) \stackrel{\text{def}}{=} L_{s,s,t}(x, \bar{h}) \quad (40)$$

where  $\bar{h}(x_{s:t}) = h(x_s)$ . The quantity  $L_{s,t}(\xi, h)/L_{s,t}(\xi, \mathbf{1})$  may be interpreted as the conditional expectation of  $h(X_t)$  given the observations up to time  $t$  and  $X_s$  evaluated at  $X_s = \xi$ . Moreover, for any distribution  $\nu$ ,

$$\phi_{\nu, t|s:t}(h) = \frac{\int \nu(dx_s) L_{s,t}(x_s, h)}{\int \nu(dx_s) L_{s,t}(x_s, \mathbf{1})}. \quad (41)$$

**PROPOSITION 3.** *Assume A1-2. Then, for all bounded measurable functions  $h : \mathbf{X} \rightarrow \mathbb{R}$ , and  $0 \leq s \leq T$ ,*

$$\sqrt{N} \left( \hat{\phi}_{\chi, s|s}(h) - \phi_{\chi, s|s}(h) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \Gamma_{\chi, s|s} \left[ h - \phi_{\chi, s|s}(h) \right] \right), \quad (42)$$

where

$$\Gamma_{\chi, s|s} [h] = \sum_{r=0}^s V_{\chi, r, s} [h], \quad (43)$$

with

$$V_{\chi, 0, s} [h] = \frac{\rho_0 \left( \omega_0^2(\cdot) L_{0, s}^2(\cdot, h) \right)}{\left( \phi_{\chi, 0| -1} [g_0(\cdot) L_{0, s}(\cdot, \mathbf{1})] \right)^2}, \quad (44)$$

and for  $1 \leq r \leq s$ ,

$$V_{\chi, r, s} [h] = \frac{\phi_{\chi, r-1|r-1} \left[ \vartheta_r(\cdot) \int p_r(\cdot, x) \omega_r^2(\cdot, x) L_{r, s}^2(x, h) dx \right] \phi_{\chi, r-1|r-1}(\vartheta_r)}{\left( \phi_{\chi, r|r-1} [g_r(\cdot) L_{r, s}(\cdot, \mathbf{1})] \right)^2}. \quad (45)$$

Using the CLT for the auxiliary particle approximation of the filtering distribution, we establish a CLT for the auxiliary particle approximation of the fixed interval joint smoothing distribution. The proof is established using this time a recursion going forward in time, i.e. a CLT for  $\hat{\phi}_{\chi, s:t|t}(\cdot)$  is deduced from a CLT for  $\hat{\phi}_{\chi, s:t-1|t-1}(\cdot)$ . The proof is based on the techniques developed in [9] (extending [3] and [20]) which are tailored to the analysis of sequential Monte-Carlo algorithms. Define, for  $0 \leq s < t \leq T$ ,  $q_{s, t-1}(x_{s:t-1}, x) = q(x_{t-1}, x)$

$$v_{s, t, T}(x, h) \stackrel{\text{def}}{=} \vartheta_t(x) \int p_t(x, x') \omega_t^2(x, x') g_{s, t, T}^2(x', h) dx', \quad (46)$$

where

$$g_{s, t, T}(x, h) \stackrel{\text{def}}{=} \frac{\phi_{\chi, s:t-1|t-1}(q_{s:t-1}(\cdot, x) L_{s, t, T}([\cdot, x], h))}{\phi_{\chi, t-1|t-1}(q(\cdot, x))}. \quad (47)$$

**THEOREM 4.** *Assume A1-2. Let  $s \leq T$ . Then, for all bounded measurable functions  $h : \mathbf{X}^{T-s+1} \rightarrow \mathbb{R}$ ,*

$$\sqrt{N} \left( \hat{\phi}_{\chi, s:T|T}(h) - \phi_{\chi, s:T|T}(h) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \Gamma_{\chi, s:T|T} \left[ h - \phi_{\chi, s:T|T}(h) \right] \right), \quad (48)$$

with

$$\begin{aligned} \Gamma_{\chi, s:T|T} [h] &\stackrel{\text{def}}{=} \frac{\Gamma_{\chi, s|s} [L_{s, s, T}(\cdot, h)]}{\phi_{\chi, s|s}^2 [L_{s, s, T}(\cdot, \mathbf{1})]} \\ &\quad + \sum_{t=s+1}^T \frac{\phi_{\chi, t-1|t-1}(v_{s, t, T}(\cdot, h)) \phi_{\chi, t-1|t-1}(\vartheta_t)}{\phi_{\chi, t|t-1}^2 [g_t(\cdot) L_{t, t, T}(\cdot, \mathbf{1})]} \end{aligned} \quad (49)$$

where  $\Gamma_{\chi,s|s}$  is defined in (42). Moreover,

$$\begin{aligned} & \sqrt{N} \left( \tilde{\phi}_{\chi,s:T|T}(h) - \phi_{\chi,s:T|T}(h) \right) \\ & \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \phi_{\chi,s:T|T}^2 \left[ h - \phi_{\chi,s:T|T}(h) \right] + \Gamma_{\chi,s:T|T} \left[ h - \phi_{\chi,s:T|T}(h) \right] \right), \end{aligned} \quad (50)$$

PROOF. Without loss of generality, we assume that  $\phi_{\chi,s:T|T}(h) = 0$ . Denote by  $\langle \cdot, \cdot \rangle$  the scalar product,  $V_{s,T}^N = [V_{s,s,T}^N, \dots, V_{s,T,T}^N]$  and  $W_{s,T}^N = [W_{s,s,T}^N, \dots, W_{s,T,T}^N]$  the vectors given by

$$V_{s,s,T}^N(h) = N^{1/2} \frac{A_{s,s,T}(h)}{\Omega_s}, \quad (51)$$

$$V_{s,t,T}^N(h) \stackrel{\text{def}}{=} N^{-1/2} \left( A_{s,t,T}(h) - \frac{A_{s,t-1,T}(h)}{A_{s,t-1,T}(\mathbf{1})} A_{s,t,T}(\mathbf{1}) \right), \quad t = s+1, \dots, T, \quad (52)$$

$$W_{s,s,T}^N = \frac{\Omega_s}{A_{s,s,T}(\mathbf{1})}, \quad W_{s,t,T}^N = \frac{N}{A_{s,t,T}(\mathbf{1})}, \quad t = s+1, \dots, T. \quad (53)$$

where  $A_{s,t,T}$  is defined in (33). Using these notations, we decompose  $\sqrt{N} \hat{\phi}_{\chi,s:T|T}(h)$  as follows

$$\sqrt{N} \hat{\phi}_{\chi,s:T|T}(h) = \left\langle V_{s,T}^N(h), W_{s,T}^N \right\rangle, \quad (54)$$

Since  $W_{s,s,T}^N = \left( \hat{\phi}_{\chi,s|s}(L_{s,s,T}(\cdot, \mathbf{1})) \right)^{-1}$  and similarly for  $t = s+1, \dots, T$ ,  $W_{s,t,T}^N = \left( N^{-1} \sum_{\ell=1}^N \omega_t^\ell L_{t,t,T}(\xi_t^\ell, \mathbf{1}) \right)^{-1}$ , Proposition 1, (23) and (24) show that

$$W_{s,s,T}^N \xrightarrow{P}_{n \rightarrow \infty} \frac{1}{\phi_{\chi,s|s}(L_{s,s,T}(\cdot, \mathbf{1}))} \quad (55)$$

$$W_{s,t,T}^N \xrightarrow{P}_{N \rightarrow \infty} \frac{\phi_{\chi,t-1|t-1}(\vartheta_t)}{\phi_{\chi,t|t-1}(g_t(\cdot) L_{t,t,T}(\cdot, \mathbf{1}))}. \quad (56)$$

Therefore (48) follows from the application of the Slutsky Lemma provided that we establish a multivariate CLT for the sequence of random vectors  $V_{s,T}^N(h)$ . For that purpose, we show that for any  $t \in \{s, \dots, T\}$  and any scalars  $\alpha_s, \dots, \alpha_t$ ,

$$\sum_{r=s}^t \alpha_r V_{s,r,T}^N(h) \xrightarrow{\mathcal{D}}_{N \rightarrow \infty} \mathcal{N} \left( 0, \sum_{r=1}^t \alpha_r^2 \sigma_{\chi,s:r|T}^2[h] \right), \quad (57)$$

where  $\sigma_{\chi,s|T}^2[h] = \Gamma_{\chi,s|s}[L_{s,s,T}(\cdot, h)]$ , and for  $r \in \{s+1, \dots, T\}$ ,

$$\sigma_{\chi,s:r|T}^2[h] \stackrel{\text{def}}{=} \frac{\phi_{\chi,s:r-1|r-1}[v_{s,r,T}(\cdot, h)]}{\phi_{\chi,r-1|r-1}(\vartheta_r)}. \quad (58)$$

First consider  $t = s$ . Since  $A_{s,s,T}(h) = \sum_{\ell=1}^N \omega_s^\ell L_{s,s,T}(\xi_s^\ell, h)$  and  $\phi_{\chi,s|s}(L_{s,s,T}(\cdot, h)) = 0$ , Proposition 3 shows that  $N^{1/2} \Omega_s^{-1} A_{s,s,T}(h)$  is asymptotically normal with zero mean and variance  $\Sigma_{s,s,T}(h)$ . Assume now that the property holds for some



$t-1 \geq s$ . We apply the results on triangular array of dependent random variables developed in [9];  $V_{s,t,T}^N(h)$  may be expressed as

$$V_{s,t,T}^N(h) = \sum_{\ell=1}^N U_{M,\ell}, \quad U_{M,\ell} = N^{-1/2} \omega_t^\ell G_{s,t,T}(\xi_t^\ell, h), \quad (59)$$

where  $G_{s,t,T}$  is defined in (36). Eq. (38) shows that  $\mathbb{E}[U_{M,\ell} | \mathcal{F}_{t-1}] = 0$ ,  $\ell = 1, \dots, N$ , which implies:

$$\mathbb{E} \left[ \sum_{r=s}^t \alpha_r V_{s,r,T}^N(h) \middle| \mathcal{F}_{t-1} \right] = \sum_{r=s}^{t-1} \alpha_r V_{s,r,T}^N(h).$$

The induction assumption shows that  $\mathbb{E} \left[ \sum_{r=s}^t \alpha_r V_{s,r,T}^N(h) \middle| \mathcal{F}_{t-1} \right]$  converges in distribution to a centered Gaussian distribution with covariance  $\sum_{r=1}^{t-1} \alpha_r^2 \Sigma_{s,r,T}(h)$ . We will now prove that

$$\mathbb{E} \left[ \exp \left( iu \sqrt{N} V_{s,t,T}^N(h) \right) \middle| \mathcal{F}_{t-1} \right] \xrightarrow{P}_{N \rightarrow \infty} \exp \left\{ -\frac{u^2}{2} \Sigma_{s,r,T}(h) \right\}.$$

By [9, Corollary 11], setting it remains to show that

$$\sum_{\ell=1}^N \mathbb{E} \left[ U_{M,\ell}^2 \middle| \mathcal{G}_{N,\ell-1} \right] \xrightarrow{P} \Sigma_{s,r,T}(h), \quad (60)$$

$$\sum_{\ell=1}^N \mathbb{E} \left[ U_{M,\ell}^2 \mathbf{1}_{\{|U_{M,\ell}| \geq \epsilon\}} \middle| \mathcal{G}_{N,\ell-1} \right] \xrightarrow{P} 0, \text{ for any } \epsilon > 0, \quad (61)$$

where, for  $\ell \in \{1, \dots, N\}$ ,  $\mathcal{G}_{N,\ell} = \mathcal{F}_{t-1} \vee \sigma(\omega_t^j, \xi_t^j, j \leq \ell)$ . We first prove (60). It follows from the definitions that

$$\begin{aligned} \sum_{\ell=1}^N \mathbb{E} \left[ U_{N,\ell}^2 \middle| \mathcal{G}_{N,\ell-1} \right] &= \mathbb{E} \left[ \left( \omega_t^1 G_{s,t,T}(\xi_t^1, h) \right)^2 \middle| \mathcal{F}_{t-1} \right] \\ &= \int \sum_{\ell=1}^N \frac{\omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell) p_t(\xi_{t-1}^\ell, x)}{\sum_{j=1}^N \omega_{t-1}^j \vartheta_t(\xi_{t-1}^j)} \left( \omega_t(\xi_{t-1}^\ell, x) G_{s,t,T}(x, h) \right)^2 dx, \\ &= \frac{\Omega_{t-1}}{\sum_{j=1}^N \omega_{t-1}^j \vartheta_t(\xi_{t-1}^j)} \frac{1}{\Omega_{t-1}} \sum_{\ell=1}^N \omega_{t-1}^\ell \Upsilon_{s,t,T}(\xi_{t-1}^\ell, h), \end{aligned}$$

where

$$\Upsilon_{s,t,T}(x, h) \stackrel{\text{def}}{=} \vartheta_t(x) \int p_t(x, x') \omega_t^2(x, x') G_{s,t,T}^2(x', h) dx'. \quad (62)$$

We will show, by applying Lemma 13 that

$$\frac{1}{\Omega_{t-1}} \sum_{\ell=1}^N \omega_{t-1}^\ell \Upsilon_{s,t,T}(\xi_{t-1}^\ell, h) \xrightarrow{P} \phi_{\chi, t-1|t-1}(v_{s,t,T}(\cdot, h)),$$

where  $v_{s,t,T}$  is defined in (46). To that purpose, we need to show that **(i)** there exists a constant  $v_\infty$  such that  $|\Upsilon_{s,t,T}(\cdot, h)|_\infty \leq v_\infty$   $\mathbb{P}$ -a.s.  $|v_{s,t,T}(\cdot, h)|_\infty \leq$

$v_\infty$ , and **(ii)** for any constant  $M \geq 0$ ,  $\sup_{|x| \leq M} |\Upsilon_{s,t,T}(x, h) - v_{s,t,T}(x, h)| \xrightarrow{\mathbb{P}} 0$ . Eq. (39) shows that  $|\omega_t(\cdot)G_{s,t,T}(\cdot, h)|_\infty \leq 2|\omega_t|_\infty |L_{t,t,T}(\cdot, \mathbf{1})|_\infty |h|_\infty$ , which implies that

$$|\Upsilon_{s,t,T}(\cdot, h)|_\infty \leq 2|\vartheta_t|_\infty |\omega_t|_\infty^2 |L_{t,t,T}(\cdot, \mathbf{1})|_\infty^2 |h|_\infty^2, \quad \mathbb{P} - a.s. \quad (63)$$

Similarly, the bound  $|g_{s,t,T}|_\infty \leq |L_{t,t,T}(\cdot, \mathbf{1})|_\infty |h|_\infty$  implies that

$$|v_{s,t,T}(\cdot, h)|_\infty \leq |\vartheta_t|_\infty |\omega_t|_\infty^2 |L_{t,t,T}(\cdot, \mathbf{1})|_\infty^2 |h|_\infty^2. \quad (64)$$

The bounds (63) and (64) imply **(i)**. To prove **(ii)**, first note that for all  $M \geq 0$ ,

$$\begin{aligned} & \sup_{|x| \leq M} |\Upsilon_{s,t,T}(x, h) - v_{s,t,T}(x, h)| \\ & \leq |\vartheta_t|_\infty |\omega_t|_\infty \int \sup_{|x| \leq M} p_t(x, x') \left| G_{s,t,T}^2(x', h) - g_{s,t,T}^2(x', h) \right| dx'. \end{aligned} \quad (65)$$

Under (A3),  $\int \sup_{|x| \leq M} p_t(x, x') dx' < \infty$ , and since  $\left| G_{s,t,T}^2(x', h) - g_{s,t,T}^2(x', h) \right| \leq 2|L_{t,t,T}(\cdot, \mathbf{1})|_\infty^2 |h|_\infty^2$ , Lemma 12 shows that

$$\int \sup_{|x| \leq M} p_t(x, x') \left| G_{s,t,T}^2(x', h) - g_{s,t,T}^2(x', h) \right| dx' \xrightarrow{\mathbb{P}} 0,$$

provided that we can show that, for any given  $x \in \mathbf{X}$ ,  $G_{s,t,T}(x, h) \xrightarrow{\mathbb{P}} g_{s,t,T}(x, h)$ . The definitions (32) of the function  $F_{s,t,T}$  and (16) of the smoothing distribution implies that

$$F_{s,t,T}(x, h) = \frac{\hat{\phi}_{\mathcal{X}, s:t-1|t-1}[q_{s,t-1}(\cdot, x)L_{s,t,T}([\cdot, x], h)]}{\hat{\phi}_{\mathcal{X}, t-1|t-1}(q(\cdot, x))}.$$

Theorem 2 show that  $F_{s,t,T}(x, h) \xrightarrow{\mathbb{P}} g_{s,t,T}(x, h)$ . On the other hand, it follows from the definitions that  $A_{s,t-1,T}(h) = \Omega_{t-1} \hat{\phi}_{\mathcal{X}, s:t|t}(L_{s,t,T}(\cdot, h))$ .

Since  $\phi_{\mathcal{X}, s:t|t}(L_{s,t,T}(\cdot, h)) = 0$ , this decomposition implies that  $A_{s,t-1,T}(h) \xrightarrow{\mathbb{P}} 0$ . Therefore,  $F_{s,t,T}(x, h) \xrightarrow{\mathbb{P}} g_{s,t,T}(x, h)$ .

It remains to check the tightness condition (61). This property is straightforward since  $|U_{M,i}| \leq N^{-1/2} |\omega_t|_\infty |h|_\infty$ .

We now prove (50). Using (20) and the definition of  $\hat{\phi}_{\mathcal{X}, s:T|T}(h)$ , we may write

$$\begin{aligned} \sqrt{N} \left( \tilde{\phi}_{\mathcal{X}, s:T|T}(h) - \phi_{\mathcal{X}, s:T|T}(h) \right) &= N^{-1/2} \sum_{\ell=1}^N \left[ h \left( \xi_{s:T}^{J_{s:T}^\ell} \right) - \mathbb{E} \left[ h \left( \xi_{s:T}^{J_{s:T}} \right) \mid \mathcal{F}_T \right] \right] \\ &+ \sqrt{N} \left( \hat{\phi}_{\mathcal{X}, s:T|T}(h) - \phi_{\mathcal{X}, s:T|T}(h) \right). \end{aligned}$$

Note that since  $(J_{s:T}^\ell)_{1 \leq \ell \leq N}$  are iid conditional to  $\mathcal{F}_T$ , (50) follows from (48) and direct application of Corollary 11 in [9] by noting that

$$\begin{aligned} & N^{-1} \sum_{\ell=1}^N \mathbb{E} \left[ \left\{ h \left( \xi_{s:T}^{J_{s:T}^\ell} \right) - \mathbb{E} \left[ h \left( \xi_{s:T}^{J_{s:T}} \right) \mid \mathcal{F}_T \right] \right\}^2 \mid \mathcal{F}_T \right] \\ &= \hat{\phi}_{\mathcal{X}, s:T|T}^2 \left[ h - \hat{\phi}_{\mathcal{X}, s:T|T}(h) \right] \xrightarrow{\mathbb{P}} \phi_{\mathcal{X}, s:T|T}^2 \left[ h - \phi_{\mathcal{X}, s:T|T}(h) \right] \end{aligned}$$

□

**5. Time-uniform deviation inequality.** In this section, we study the long-term behavior of the marginal fixed-interval smoothing distribution estimator. For that purpose, it is required to impose a form of mixing condition on the Markov transition kernel. For simplicity, we consider conditions which are similar to the ones used in [5, chapter 7.4] or [2, chapter 4]; these conditions can be relaxed, but at the expense of many technical problems. This condition requires that the transition kernel is strongly mixing in the sense that

**A 4.** There exist two constants  $0 < \sigma_- \leq \sigma_+ < \infty$ , such that, for any  $(x, x') \in \mathsf{X} \times \mathsf{X}$ ,

$$\sigma_- \leq q(x, x') \leq \sigma_+ . \quad (66)$$

In addition, there exists a constant  $c_- > 0$  such that,  $\int \chi(dx_0)g_0(x_0) \geq c_-$  and for all  $t \geq 1$ ,

$$\inf_{x \in \mathsf{X}} \int q(x, x')g_t(x')dx' \geq c_- > 0 . \quad (67)$$

Note that assumption 4 implies that  $\nu(\mathsf{X}) < \infty$ ; in the sequel, we will consider without loss of generality that  $\nu(\mathsf{X}) = 1$ . Moreover, the average number of simulations required in the accept-reject mechanism per sample of the FFBSi algorithm is bounded by  $\sigma^+/\sigma^-$ . An important consequence of the uniform ergodicity condition is the forgetting of either the initial or the final conditions. The following Proposition extends some of the results obtained initially in [6] and later extended in [7] (see also [5, Chapter 7] and [2, Chapter 2]). Define

$$\begin{aligned} \ell_{s,t}(x_s, x_t) &\stackrel{\text{def}}{=} \int \cdots \int q(x_s, x_{s+1})g_{s+1}(x_{s+1}) \\ &\quad \times \prod_{u=s+1}^{t-1} q(x_u, x_{u+1})g_{u+1}(x_{u+1})dx_{s+1:t-1} \end{aligned} \quad (68)$$

for  $s < t$ , and  $\ell_{t,t}(x_s, x_t) \stackrel{\text{def}}{=} \delta_{x_s}(x_t)$ , so that

$$L_{s,t}(x_s, h) = \int \ell_{s,t}(x_s, x_t)h(x_t)dx_t .$$

**PROPOSITION 5.** *Assume (A4). Then, for all distributions  $\chi, \chi'$  and for all  $s \leq t$  and any bounded measurable functions  $h$ ,*

$$\left| \frac{\iint \chi(dx_s)\ell_{s,t}(x_s, x_t)h(x_t)dx_t}{\iint \chi(dx_s)\ell_{s,t}(x_s, x_t)dx_t} - \frac{\iint \chi'(dx_s)\ell_{s,t}(x_s, x_t)h(x_t)dx_t}{\iint \chi'(dx_s)\ell_{s,t}(x_s, x_t)dx_t} \right| \leq \rho^{t-s} \text{osc}(h) , \quad (69)$$

where  $\rho \stackrel{\text{def}}{=} 1 - \sigma_-/\sigma_+$ . In addition, for any bounded non-negative measurable functions  $f$  and  $f'$ ,

$$\left| \frac{\iint \chi(dx_s)h(x_s)\ell_{s,t}(x_s, x_t)f(x_t)dx_t}{\iint \chi(dx_s)\ell_{s,t}(x_s, x_t)f(x_t)dx_t} - \frac{\iint \chi(dx_s)h(x_s)\ell_{s,t}(x_s, x_t)f'(x_t)dx_t}{\iint \chi(dx_s)\ell_{s,t}(x_s, x_t)f'(x_t)dx_t} \right| \leq \rho^{t-s} \text{osc}(h) , \quad (70)$$

as soon as the denominators are non-zero.

PROOF. The first statement is well-known; see for example [5] and [2, Proposition 4.3.23]. To check the second statement, denote by  $B_{\chi,s,t}$  the backward smoothing kernel, defined for  $s \leq t$  and for any real-valued, measurable function  $\psi$  on  $\mathbf{X}^2$  by

$$\mathbb{E}_{X_s \sim \chi} [\psi(X_{t-1}, X_t) | Y_{s:t}] = \int \psi(x_{t-1}, x_t) \phi_{\chi,s:t|s:t}(x_t) B_{\chi,s,t}(x_t, dx_{t-1}) ,$$

where  $\phi_{\chi,s:t-1|s:t-1}$  is defined in (1). This kernel is absolutely continuous with respect to the dominating measure  $\lambda$  and its density is given by

$$b_{\chi,s,t}(x_t, x_{t-1}) = \frac{\phi_{\chi,s:t-1|s:t-1}(x_{t-1}) q(x_{t-1}, x_t)}{\int \phi_{\chi,s:t-1|s:t-1}(x) q(x, x_t) dx} .$$

Under (A4), this transition density is lower bounded by

$$b_{\chi,s,t}(x_t, x_{t-1}) \geq \frac{\sigma_-}{\int \phi_{\chi,s:t-1|s:t-1}(x) \sigma_+ dx} = \frac{\sigma_-}{\sigma_+} .$$

Since  $\text{osc}(B_{\chi,s,t}(\cdot, h)) \leq \rho \text{osc}(h)$ , it follows that

$$\text{osc}(B_{\chi,s,t} \dots B_{\chi,s,s+1}(\cdot, h)) \leq \rho^{t-s} \text{osc}(h) . \quad (71)$$

Note that

$$\begin{aligned} & \left| \frac{\iint \chi(dx_s) h(x_s) \ell_{s,t}(x_s, x_t) f(x_t) dx_t}{\iint \chi(dx_s) \ell_{s,t}(x_s, x_t) f(x_t) dx_t} - \frac{\iint \chi(dx_s) h(x_s) \ell_{s,t}(x_s, x_t) f'(x_t) dx_t}{\iint \chi(dx_s) \ell_{s,t}(x_s, x_t) f'(x_t) dx_t} \right| \\ &= \left| \frac{\phi_{\chi,t|s:t}(B_{\chi,s,t} \dots B_{\chi,s,s+1}(\cdot, h) f(\cdot))}{\phi_{\chi,t|s:t}(f(\cdot))} - \frac{\phi_{\chi,t|s:t}(B_{\chi,s,t} \dots B_{\chi,s,s+1}(\cdot, h) f'(\cdot))}{\phi_{\chi,t|s:t}(f'(\cdot))} \right| \\ &= |\mu_f [B_{\chi,s,t} \dots B_{\chi,s,s+1}(\cdot, h)] - \mu_{f'} [B_{\chi,s,t} \dots B_{\chi,s,s+1}(\cdot, h)]| \end{aligned}$$

with, for any  $A \in \mathcal{B}(\mathbf{X})$ ,  $\mu_f(A) \stackrel{\text{def}}{=} \phi_{\chi,t|s:t}(\mathbf{1}_A f) / \phi_{\chi,t|s:t}(f)$ . Therefore, since for any probabilities  $\mu$  and  $\mu'$  on  $\mathcal{B}(\mathbf{X})$  and any measurable function  $\psi$ ,  $|\mu(\psi) - \mu'(\psi)| \leq \text{osc}(\psi)$ , (71) shows that

$$|\mu_f [B_{\chi,s,t} \dots B_{\chi,s,s+1}(\cdot, h)] - \mu_{f'} [B_{\chi,s,t} \dots B_{\chi,s,s+1}(\cdot, h)]| \leq \rho^{t-s} \text{osc}(h) .$$

□

The goal of this section consists in establishing, under the assumptions mentioned above, that the FFBS approximation of the *marginal* fixed interval smoothing probability satisfies an exponential deviation inequality with constants that are uniform in time.

The first step in the proof consists in showing a time-uniform deviation inequality for the auxiliary particle filter. Here again, the proof of this result could be adapted from [5, Section 7.4.3]. For the sake of clarity, we present a self-contained proof, which is valid under assumptions that are weaker than those used in [5, Chapter 7].

PROPOSITION 6. *Assume that A1–4 hold with  $T = \infty$ . Then, the filtering distribution satisfies a time-uniform exponential deviation inequality, i.e. there exist constants  $B$  and  $C$  such that, for all integers  $N$  and  $t \geq 0$ , all measurable functions  $h$  and all  $\epsilon > 0$ ,*

$$\mathbb{P} \left[ \left| N^{-1} \sum_{i=1}^N \omega_t^i h(\xi_t^i) - \frac{\phi_{\mathcal{X},t|t-1}(g_t h)}{\phi_{\mathcal{X},t-1|t-1}(\vartheta_t)} \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2/|h|_\infty^2}, \quad (72)$$

$$\mathbb{P} \left[ \left| \hat{\phi}_{\mathcal{X},t|t}(h) - \phi_{\mathcal{X},t|t}(h) \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2/\text{osc}^2(h)}. \quad (73)$$

PROOF. We first prove (73). Without loss of generality, we will assume that  $\phi_{\mathcal{X},t|t}(h) = 0$ . Similar to [5, Eq. (7.24)], the quantity  $\hat{\phi}_{\mathcal{X},t|t}(h)$  is decomposed as,

$$\hat{\phi}_{\mathcal{X},t|t}(h) = \sum_{s=1}^t \left( \frac{B_{s,t}(h)}{B_{s,t}(\mathbf{1})} - \frac{B_{s-1,t}(h)}{B_{s-1,t}(\mathbf{1})} \right) + \frac{B_{0,t}(h)}{B_{0,t}(\mathbf{1})}, \quad (74)$$

where

$$B_{s,t}(h) = N^{-1} \sum_{i=1}^N \omega_s^i \frac{L_{s,t}(\xi_s^i, h)}{|L_{s,t}(\cdot, \mathbf{1})|_\infty}. \quad (75)$$

We first establish some exponential inequality for  $B_{0,t}(h)/B_{0,t}(\mathbf{1})$  where the dependence in  $t$  will be explicitly expressed. For that purpose, we will apply Lemma 11 by successively checking Conditions (I), (II), and (III), with

$$\begin{cases} a_N \stackrel{\text{def}}{=} B_{0,t}(h) \\ b_N \stackrel{\text{def}}{=} B_{0,t}(\mathbf{1}) \\ c_N \stackrel{\text{def}}{=} 0 \\ b \stackrel{\text{def}}{=} \int \chi(dx_0) g_0(x_0) \frac{L_{0,t}(x_0, \mathbf{1})}{|L_{0,t}(\cdot, \mathbf{1})|_\infty} \\ \beta \stackrel{\text{def}}{=} \frac{\sigma_-}{\sigma_+} \int \chi(dx_0) g_0(x_0). \end{cases}$$

Under the strong mixing condition A4, it may be shown that (see [5, chapter 4] or [2, section 4.3.3])

$$\frac{\sigma_-}{\sigma_+} \leq \frac{L_{s,t}(\xi, \mathbf{1})}{|L_{s,t}(\cdot, \mathbf{1})|_\infty} \leq \frac{\sigma_+}{\sigma_-}, \quad (76)$$

which implies that  $b \geq \beta$ . Since  $\phi_{\mathcal{X},t|t}(h) = 0$ , Eqs. (41) and (69) imply

$$\begin{aligned} \left| \frac{a_N}{b_N} \right| &= \left| \frac{B_{0,t}(h)}{B_{0,t}(\mathbf{1})} - \phi_{\mathcal{X},t|t}(h) \right| \\ &= \left| \frac{\sum_{i=1}^N \omega_0^i L_{0,t}(\xi_0^i, h)}{\sum_{i=1}^N \omega_0^i L_{0,t}(\xi_0^i, \mathbf{1})} - \frac{\int \chi(dx_0) g_0(x_0) L_{0,t}(x_0, h)}{\int \chi(dx_0) g_0(x_0) L_{0,t}(x_0, \mathbf{1})} \right| \leq \rho^t \text{osc}(h). \end{aligned} \quad (77)$$

This shows condition (I) with  $M = \rho^t |h|_\infty$ . We now turn to condition (II). We have

$$b_N - b = N^{-1} \sum_{i=1}^N \omega_0^i \frac{L_{0,t}(\xi_0^i, \mathbf{1})}{|L_{0,t}(\cdot, \mathbf{1})|_\infty} - \int \rho_0(dx_0) \omega_0(x_0) g_0(x_0) \frac{L_{0,t}(x_0, \mathbf{1})}{|L_{0,t}(\cdot, \mathbf{1})|_\infty}.$$

Since  $\left| \omega_0^i \frac{L_{0,t}(\xi_0^i, \mathbf{1})}{|L_{0,t}(\cdot, \mathbf{1})|_\infty} \right| \leq \frac{\sigma_+}{\sigma_-} |\omega_0|_\infty$ , we have by Hoeffding's inequality

$$\mathbb{P}[|b_N - b| \geq \epsilon] \leq B \exp(-CN\epsilon^2),$$

where the constants  $B$  and  $C$  do not depend on  $t$ . This shows condition (II). We now check condition (III). We have

$$a_N - (c_N/d_N)b_N = a_N = N^{-1} \sum_{i=1}^N \omega_0^i \frac{L_{0,t}(\xi_0^i, h)}{|L_{0,t}(\cdot, \mathbf{1})|_\infty}.$$

Now, as  $\phi_{\chi,t|t}(h) = 0$  implies  $\int \chi(dx) L_{0,t}(x, h) = 0$ , it holds that  $\mathbb{E}(a_N) = 0$ . Moreover,

$$\begin{aligned} \left| \omega_0^i \frac{L_{0,t}(\xi_0^i, h)}{|L_{0,t}(\cdot, \mathbf{1})|_\infty} \right| &\leq |\omega_0|_\infty \left| \frac{L_{0,t}(\xi_0^i, \mathbf{1})}{|L_{0,t}(\cdot, \mathbf{1})|_\infty} \left( \frac{L_{0,t}(\xi_0^i, h)}{L_{0,t}(\xi_0^i, \mathbf{1})} - \phi_{\chi,t|t}(h) \right) \right| \\ &\leq |\omega_0|_\infty \frac{\sigma_+}{\sigma_-} \rho^t \text{osc}(h), \end{aligned}$$

using (41) and (69). Condition (III) follows from Hoeffding's inequality. Then, Lemma 11 gives

$$\mathbb{P}[|B_{0,t}(h)/B_{0,t}(\mathbf{1})| > \epsilon] \leq B e^{-CN\epsilon^2/(\rho^t \text{osc}(h))^2},$$

where the constants  $B$  and  $C$  do not depend on  $t$ .

We now consider for  $1 \leq s \leq t$  the difference  $B_{s,t}(h)/B_{s,t}(\mathbf{1}) - B_{s-1,t}(h)/B_{s-1,t}(\mathbf{1})$ , where  $B_{s,t}$  is defined in (75). We again use Lemma 11 where  $\mathbb{P}(\cdot) = \mathbb{P}(\cdot | \mathcal{F}_{s-1})$ ,  $a_N = B_{s,t}(h)$ ,  $b_N = B_{s,t}(\mathbf{1})$ ,  $c_N = B_{s-1,t}(h)$ ,  $d_N = B_{s-1,t}(\mathbf{1})$ ,

$$b = \frac{\sum_{i=1}^N \omega_{s-1}^i \int Q(\xi_{s-1}^i, dx) g_s(x) L_{0,s}(x, \mathbf{1})}{|L_{0,s}(\cdot, \mathbf{1})|_\infty \sum_{\ell=1}^N \omega_{s-1}^\ell \vartheta_s(\xi_{s-1}^\ell)}, \quad \text{and} \quad \beta = \frac{c_- \sigma_-}{\sigma_+ |\vartheta_s|_\infty},$$

where  $\sigma_-$  and  $c_-$  are defined in (66) and (67), respectively. It appears using (76) and (A4) that  $b \geq \beta$ . Moreover,

$$\left| \frac{a_N}{b_N} - \frac{c_N}{d_N} \right| = \left| \frac{\sum_{i=1}^N \omega_s^i L_{s,t}(\xi_s^i, h)}{\sum_{i=1}^N \omega_s^i L_{s,t}(\xi_s^i, \mathbf{1})} - \frac{\sum_{i=1}^N \omega_{s-1}^i L_{s-1,t}(\xi_{s-1}^i, h)}{\sum_{i=1}^N \omega_{s-1}^i L_{s-1,t}(\xi_{s-1}^i, \mathbf{1})} \right| \leq \rho^{t-s} \text{osc}(h), \quad (78)$$

showing condition (I) with  $M = \rho^{t-s} \text{osc}(h)$ . We now check condition (II). By (26), we have

$$b_N - b = N^{-1} \sum_{i=1}^N \omega_s^i \frac{L_{s,t}(\xi_s^i, \mathbf{1})}{|L_{s,t}(\cdot, \mathbf{1})|_\infty} - \mathbb{E} \left[ \omega_s^1 \frac{L_{s,t}(\xi_s^1, \mathbf{1})}{|L_{s,t}(\cdot, \mathbf{1})|_\infty} \middle| \mathcal{F}_{s-1} \right].$$

Thus, since  $|\omega_s^i L_{s,t}(\xi_s^i, \mathbf{1}) / |L_{s,t}(\cdot, \mathbf{1})|_\infty| \leq \sup_t |\omega_t|_\infty \sigma_+ / \sigma_-$ , we have by conditional Hoeffding's inequality

$$\mathbb{P}(|b_N - b| > \epsilon | \mathcal{F}_{s-1}) \leq B e^{-CN\epsilon^2/(\text{osc}(h))^2}$$

showing condition (II) with constants which do not depend on  $s$ . Moreover, write  $a_N - \frac{c_N}{d_N} b_N = N^{-1} \sum_{\ell=1}^N \eta^\ell$  where

$$\eta^\ell \stackrel{\text{def}}{=} \omega_s^\ell \frac{L_{s,t}(\xi_s^\ell, h)}{|L_{s,t}(\cdot, \mathbf{1})|_\infty} - \frac{\sum_{i=1}^N \omega_{s-1}^i L_{s-1,t}(\xi_{s-1}^i, h)}{\sum_{i=1}^N \omega_{s-1}^i L_{s-1,t}(\xi_{s-1}^i, \mathbf{1})} \left( \omega_s^\ell \frac{L_{s,t}(\xi_s^\ell, \mathbf{1})}{|L_{s,t}(\cdot, \mathbf{1})|_\infty} \right).$$

Since  $\{(\xi_t^\ell, \omega_t^\ell)\}_{\ell=1}^N$  are i.i.d. conditionally to the  $\sigma$ -field  $\mathcal{F}_{t-1}$ , we have that  $\{\eta^\ell\}_{\ell=1}^N$  are also i.i.d. conditionally to  $\mathcal{F}_{t-1}$ . Moreover, it can be easily checked using (26) that  $\mathbb{E}[\eta^1 | \mathcal{F}_{t-1}] = 0$ . In order to apply the conditional Hoeffding inequality, we need to check that  $\eta^\ell$  is bounded. In fact, using (41) and (69),

$$|\eta^\ell| = \frac{L_{s,t}(\xi_s^i, \mathbf{1})}{|L_{s,t}(\cdot, \mathbf{1})|_\infty} \left| \frac{L_{s,t}(\xi_s^\ell, h)}{L_{s,t}(\xi_s^\ell, \mathbf{1})} - \frac{\sum_{i=1}^N \omega_{s-1}^i L_{s-1,t}(\xi_{s-1}^i, h)}{\sum_{i=1}^N \omega_{s-1}^i L_{s-1,t}(\xi_{s-1}^i, \mathbf{1})} \right| \leq \frac{\sigma_+}{\sigma_-} \rho^{t-s} \text{osc}(h)$$

Consequently,

$$\begin{aligned} \mathbb{P} \left( \left| a_N - \frac{c_N}{d_N} b_N \right| > \epsilon \mid \mathcal{F}_{s-1} \right) &= \mathbb{P} \left( \left| N^{-1} \sum_{\ell=1}^N \eta^\ell \right| > \epsilon \mid \mathcal{F}_{s-1} \right) \\ &\leq B \exp \left\{ -CN \left( \frac{\epsilon}{\rho^{t-s} \text{osc}(h)} \right)^2 \right\}, \end{aligned}$$

where the constants  $B$  and  $C$  do not depend on  $s$ . This shows condition (III). Finally by Lemma 11,

$$\mathbb{P} \left( \left| \frac{B_{s,t}(h)}{B_{s,t}(\mathbf{1})} - \frac{B_{s-1,t}(h)}{B_{s-1,t}(\mathbf{1})} \right| > \epsilon \mid \mathcal{F}_{s-1} \right) \leq B \exp \left\{ -CN \left( \frac{\epsilon}{\rho^{t-s} \text{osc}(h)} \right)^2 \right\}$$

The proof is concluded by using Lemma 14.  $\square$

We now show that the time uniform deviation inequality for the filtering estimator extends, under the mixing assumption (A4) on the Markov kernel  $Q$ , to the FFBS smoothing estimator. The key result to establish a time uniform bound for the FFBS smoothing estimator is the following Proposition, which establishes the uniform ergodicity of the particle approximation of the backward kernel.

LEMMA 7. *Assume A4. Then, for any probability distributions  $\mu$  and  $\mu'$  on the set  $\{1, \dots, N\}$ , any integers  $0 \leq s < t$  and any function  $h$  on  $\{1, \dots, N\}$ ,*

$$\left| \sum_{i_{s:t}=1}^N h(i_s) \varpi_{s:t}^{i_{s:t}} \{ \mu(i_t) - \mu'(i_t) \} \right| \leq \text{osc}(h) \rho^{t-s},$$

where  $\varpi_{s:t}^{i_{s:t}}$  is defined in (15).

PROOF. For  $u \in \{s+1, \dots, t\}$ , define  $W_u$  the  $N \times N$  matrix with entries

$$W_u^{i,j} = \frac{\omega_{u-1}^j q(\xi_{u-1}^j, \xi_u^i)}{\sum_{\ell=1}^N \omega_{u-1}^\ell q(\xi_{u-1}^\ell, \xi_u^i)}, \quad 1 \leq i, j \leq N.$$

The matrix  $W_u$  can be interpreted as the Markov transition matrix of a non-homogeneous Markov chain on the state-space  $\{1, \dots, N\}$  (which may be seen as the particle approximation of the backward kernel (2)). Using this notation, for any probability distribution  $\mu$  on  $\{1, \dots, N\}$  and any function  $h$  on  $\{1, \dots, N\}$ , the sum  $\sum_{i_s:t=1}^N h(i_s) \varpi_{s:t}^{i_s:t} \mu(i_t)$  may be interpreted as the expectation of the function  $h$  under the marginal distribution at time  $t - s$  of a non-homogeneous Markov chain started at time 0 from the initial distribution  $\mu$  and driven the transition matrix  $W_t, W_{t-1}, \dots$ :

$$\sum_{i_s:t=1}^N h(i_s) \varpi_{s:t}^{i_s:t} \mu(i_t) = \sum_{i_s:t=1}^N \mu(i_t) W_t^{i_t, i_{t-1}} \dots W_{s+1}^{i_{s+1}, i_s} h(i_s).$$

Under (A4), the entries of these transition kernels are lower-bounded by  $\sigma_-/\sigma_+$ . Therefore, the Dobrushin coefficient of each transition matrix  $W_u$ ,  $u \in \{s+1, \dots, t\}$  is upper bounded by  $\rho$  (see [8]). The result follows.  $\square$

We then show that the existence of time-uniform exponential deviation inequality for the auxiliary particle filter approximation of the filtering distribution extends to the FFBS smoothing estimator.

**THEOREM 8.** *Assume A1–4 hold with  $T = \infty$ . Then, there exist constants  $0 \leq B, C < \infty$  such that for all integers  $N, s$ , and  $T, s \leq T$ , all  $\epsilon > 0$ ,*

$$\mathbb{P} \left[ \left| \hat{\phi}_{\chi, s|T}(h) - \phi_{\chi, s|T}(h) \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2 / \text{osc}^2(h)}, \quad (79)$$

$$\mathbb{P} \left[ \left| \tilde{\phi}_{\chi, s|T}(h) - \phi_{\chi, s|T}(h) \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2 / \text{osc}^2(h)}, \quad (80)$$

where  $\hat{\phi}_{\chi, s|T}(h)$  and  $\tilde{\phi}_{\chi, s|T}(h)$  are defined in (14) and (21).

**PROOF.** (80) follows from (79) along the same lines as in Theorem 2. We use the notations of Theorem 2. Let  $h$  be a function defined on  $\mathbf{X}$  and  $s, T$  be positive integers such that  $s \leq T$ . Without loss of generality, we assume that  $\phi_{\chi, s|T}(h) = 0$ . We will denote by

$$\bar{h} : (x_s, \dots, x_T) \mapsto h(x_s). \quad (81)$$

For  $s \in \{0, \dots, T\}$ , consider again the following decomposition

$$\hat{\phi}_{\chi, s|T}(h) = \frac{A_{s, T, T}(\bar{h})}{A_{s, T, T}(\mathbf{1})} = \frac{A_{s, s, T}(\bar{h})}{A_{s, s, T}(\mathbf{1})} + \sum_{t=s+1}^T \left\{ \frac{A_{s, t, T}(\bar{h})}{A_{s, t, T}(\mathbf{1})} - \frac{A_{s, t-1, T}(\bar{h})}{A_{s, t-1, T}(\mathbf{1})} \right\},$$

where  $\bar{h}$ ,  $A_{s, s, T}$ , and  $A_{s, t, T}$  are defined in (81), (31) and (33), respectively. Note that  $\bar{h}$  depends of  $\xi_{s:t}^{i_s:t}$  only through its first component  $\xi_s^{i_s}$ ; therefore, it follows from the definition (29) of  $L_{s, t, T}$  that  $L_{s, t, T}(\xi_{s:t}^{i_s:t}, \bar{h}) = h(\xi_s^{i_s}) L_{t, t, T}(\xi_t^{i_t}, \mathbf{1})$ . In particular,  $L_{s, t, T}(\xi_{s:t}^{i_s:t}, \mathbf{1}) = L_{t, t, T}(\xi_t^{i_t}, \mathbf{1})$ .

We first consider the term  $A_{s, s, T}(\bar{h})/A_{s, s, T}(\mathbf{1})$ . It follows from the definition that

$$A_{s, s, T}(\bar{h}) = \sum_{\ell=1}^N \omega_s^\ell F_{s, s, T}(\xi_s^\ell, \bar{h}) = \sum_{\ell=1}^N \omega_s^\ell h(\xi_s^\ell) L_{s, s, T}(\xi_s^\ell, \mathbf{1}).$$



We apply Lemma 11 with

$$\begin{cases} a_N = \Omega_s^{-1} A_{s,s,T}(\bar{h}) / |L_{s,s,T}(\cdot, \mathbf{1})|_\infty \\ b_N = \Omega_s^{-1} A_{s,s,T}(\mathbf{1}) / |L_{s,s,T}(\cdot, \mathbf{1})|_\infty \\ c_N = 0 \\ b = \phi_{\chi,s|s}[L_{s,s,T}(\cdot, \mathbf{1})] / |L_{s,s,T}(\cdot, \mathbf{1})|_\infty \\ \beta = \sigma_- / \sigma_+ . \end{cases}$$

Using the definition (29) and (A4), for any  $s \in \{0, \dots, T\}$ ,

$$\frac{\sigma_-}{\sigma_+} \leq \frac{L_{s,s,T}(\xi, \mathbf{1})}{|L_{s,s,T}(\cdot, \mathbf{1})|_\infty} \leq \frac{\sigma_+}{\sigma_-} . \quad (82)$$

Therefore  $b \geq \beta$ . Then, note that  $|a_N/b_N| \leq |h|_\infty$ ; therefore condition (I) is satisfied with  $M = |h|_\infty$ . We now check condition (II). We have

$$b_N - b = \hat{\phi}_{\chi,t|t} \left[ \frac{L_{s,s,T}(\cdot, \mathbf{1})}{|L_{s,s,T}(\cdot, \mathbf{1})|_\infty} \right] - \phi_{\chi,t|t} \left[ \frac{L_{s,s,T}(\cdot, \mathbf{1})}{|L_{s,s,T}(\cdot, \mathbf{1})|_\infty} \right] .$$

Inequalities (73) and (82) show that there exists constants  $B$  and  $C$  such that for any  $\epsilon > 0$  and all positive integers  $s \leq T$ ,

$$\mathbb{P}(|b_N - b| \geq \epsilon) \leq B e^{-CN\epsilon^2} .$$

Hence, condition (II) is satisfied. Moreover,

$$a_N - \frac{c_N}{d_N} b_N = a_N = \Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell G_s(\xi_s^\ell) , \quad \text{where } G_s(\xi) = h(\xi) \frac{L_{s,s,T}(\xi, \mathbf{1})}{|L_{s,s,T}(\cdot, \mathbf{1})|_\infty} .$$

Using the definition (29) of  $L_{s,s,T}$ ,

$$\phi_{\chi,s|T}(h) = \frac{\phi_{\chi,s|s}[h(\cdot)L_{s,s,T}(\cdot, \mathbf{1})]}{\phi_{\chi,s|s}[L_{s,s,T}(\cdot, \mathbf{1})]} .$$

The condition  $\phi_{\chi,s|T}(h) = 0$  therefore implies that  $\phi_{\chi,s|s}(G_s) = 0$ . On the other hand, using (82),  $|G_s|_\infty \leq |h|_\infty \sigma_+ / \sigma_-$ . Hence, by (73),

$$\mathbb{P} \left[ \left| a_N - \frac{c_N}{d_N} b_N \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2 / \text{osc}^2(h)} ,$$

for some  $B$  and  $C$  which do not depend on  $s$  nor  $T$ . Hence condition (III) is satisfied. Combining the result above, Lemma 11 therefore shows that,

$$\mathbb{P} \left[ \left| \frac{A_{s,s,T}(\bar{h})}{A_{s,s,T}(\mathbf{1})} \right| \geq \epsilon \right] \leq 2B \exp \left( -CN\epsilon^2 / \text{osc}^2(h) \right) .$$

We now consider the term  $A_{s,t,T}(\bar{h})/A_{s,t,T}(\mathbf{1}) - A_{s,t-1,T}(\bar{h})/A_{s,t-1,T}(\mathbf{1})$  for  $t > s$ . For that purpose, we use Lemma 11 with

$$\begin{cases} a_N = N^{-1} A_{s,t,T}(\bar{h}) / |L_{t,t,T}(\cdot, \mathbf{1})|_\infty \\ b_N = N^{-1} A_{s,t,T}(\mathbf{1}) / |L_{t,t,T}(\cdot, \mathbf{1})|_\infty \\ c_N = A_{s,t-1,T}(\bar{h}) \\ d_N = A_{s,t-1,T}(\mathbf{1}) \\ b = \phi_{\chi,t-1|t-1}[L_{t-1,t-1,T}(\cdot, \mathbf{1})] / (|L_{t,t,T}(\cdot, \mathbf{1})|_\infty \phi_{\chi,t-1|t-1}(\vartheta_t)) \\ \beta = c_- \sigma_- / (\sigma_+ |\vartheta_s|_\infty) \end{cases}$$

Inequality (82) and (67) directly imply that  $b \geq \beta$ . Moreover,

$$\begin{aligned} \frac{a_N}{b_N} - \frac{c_N}{d_N} &= \sum_{i_{s:t}=1}^N h(\xi_s^{i_s}) \varpi_{s:t-1}^{i_{s:t-1}} \frac{\omega_{t-1}^{i_{t-1}} q(\xi_{t-1}^{i_{t-1}}, \xi_t^{i_t})}{\sum_{\ell=1}^N \omega_{t-1}^\ell q(\xi_{t-1}^\ell, \xi_t^{i_t})} \frac{\omega_t^{i_t} L_{t,t,T}(\xi_t^{i_t}, \mathbf{1})}{\sum_{\ell=1}^N \omega_t^\ell L_{t,t,T}(\xi_t^\ell, \mathbf{1})} \\ &\quad - \sum_{i_{s:t-1}=1}^N h(\xi_s^{i_s}) \varpi_{s:t-1}^{i_{s:t-1}} \frac{\omega_{t-1}^{i_{t-1}} L_{t-1,t-1,T}(\xi_{t-1}^{i_{t-1}}, \mathbf{1})}{\sum_{\ell=1}^N \omega_{t-1}^\ell L_{t-1,t-1,T}(\xi_{t-1}^\ell, \mathbf{1})} \\ &= \sum_{i_{s:t}=1}^N h(\xi_s^{i_s}) \varpi_{s:t-1}^{i_{s:t-1}} \left\{ \mu_{t-1}(i_{t-1}, \xi_t^{i_t}) - \mu'_{t-1}(i_{t-1}) \right\} \frac{\omega_t^{i_t} L_{t,t,T}(\xi_t^{i_t}, \mathbf{1})}{\sum_{\ell=1}^N \omega_t^\ell L_{t,t,T}(\xi_t^\ell, \mathbf{1})}, \end{aligned} \quad (83)$$

where  $\mu_{t-1}(\cdot, \xi)$  and  $\mu'_{t-1}(\cdot)$  are two probability distributions on the set  $\{1, \dots, N\}$  defined as

$$\mu_{t-1}(i, \xi) = \frac{\omega_{t-1}^i q(\xi_{t-1}^i, \xi)}{\sum_{\ell=1}^N \omega_{t-1}^\ell q(\xi_{t-1}^\ell, \xi)} \quad \text{and} \quad \mu'_{t-1}(i) = \frac{\omega_{t-1}^i L_{t-1,t-1,T}(\xi_{t-1}^i, \mathbf{1})}{\sum_{\ell=1}^N \omega_{t-1}^\ell L_{t-1,t-1,T}(\xi_{t-1}^\ell, \mathbf{1})}. \quad (84)$$

It follows from Lemma 7 that, for all  $\xi$

$$\left| \sum_{i_{s:t-1}=1}^N h(\xi_s^{i_s}) \varpi_{s:t-1}^{i_{s:t-1}} \left\{ \mu_{t-1}(i_{t-1}, \xi) - \mu'_{t-1}(i_{t-1}) \right\} \right| \leq \text{osc}(h) \rho^{t-1-s},$$

which in turn implies that

$$\left| \frac{a_N}{b_N} - \frac{c_N}{d_N} \right| \leq \text{osc}(h) \rho^{t-1-s}, \quad (85)$$

showing condition (I) with  $M = \text{osc}(h) \rho^{t-s-1}$ . We now consider the condition (II). It follows from the definition of  $b_N$  that:

$$b_N = N^{-1} \sum_{i_t=1}^N \omega_t^{i_t} \frac{L_{t,t,T}(\xi_t^{i_t}, \mathbf{1})}{|L_{t,t,T}(\cdot, \mathbf{1})|_\infty}.$$

By (72) and (82),

$$\mathbb{P}[|b_N - b| \geq \epsilon] \leq B e^{-CN\epsilon^2}$$

where the constants  $B$  and  $C$  do not depend on the time indexes  $t$  and  $T$ . This relation shows condition (II). We finally consider condition (III). Using (83),  $a_N - \frac{c_N}{d_N} b_N = \Omega_t^{-1} \sum_{i=1}^N \omega_t^i G_t(\xi_t^i)$ , where

$$\begin{aligned} G_t(\xi) &= \sum_{i_{s:t-1}=1}^N h(\xi_s^{i_s}) \varpi_{s:t-1}^{i_{s:t-1}} \\ &\times \left\{ \frac{\omega_{t-1}^{i_{t-1}} q(\xi_{t-1}^{i_{t-1}}, \xi)}{\sum_{\ell=1}^N \omega_{t-1}^\ell q(\xi_{t-1}^\ell, \xi)} - \frac{\omega_{t-1}^{i_{t-1}} L_{t-1,t-1,T}(\xi_{t-1}^{i_{t-1}}, \mathbf{1})}{\sum_{\ell=1}^N \omega_{t-1}^\ell L_{t-1,t-1,T}(\xi_{t-1}^\ell, \mathbf{1})} \right\} \frac{L_{t,t,T}(\xi, \mathbf{1})}{|L_{t,t,T}(\cdot, \mathbf{1})|_\infty}. \end{aligned}$$

Using that

$$\begin{aligned} \mathbb{E} \left[ \frac{\omega_{t-1}^{i_{t-1}} q(\xi_{t-1}^{i_{t-1}}, \xi_t^1) \omega_t^1 L_{t,t,T}(\xi_t^1, \mathbf{1})}{\sum_{\ell=1}^N \omega_{t-1}^\ell q(\xi_{t-1}^\ell, \xi_t^1)} \middle| \mathcal{F}_{t-1} \right] \\ = \frac{\omega_{t-1}^{i_{t-1}} \int q(\xi_{t-1}^{i_{t-1}}, x) g_t(x) L_{t,t,T}(x, \mathbf{1}) dx}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} = \frac{\omega_{t-1}^{i_{t-1}} L_{t-1,t-1,T}(\xi_{t-1}^{i_{t-1}}, \mathbf{1})}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} \end{aligned}$$

and

$$\mathbb{E} \left[ \omega_t^1 L_{t,t,T}(\xi_t^1, \mathbf{1}) \middle| \mathcal{F}_{t-1} \right] = \frac{\sum_{\ell=1}^N \omega_{t-1}^\ell L_{t-1,t-1,T}(\xi_{t-1}^\ell, \mathbf{1})}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)}$$

it follows that  $\mathbb{E} [\omega_t^1 G_t(\xi_t^1) \mid \mathcal{F}_{t-1}] = 0$ . On the other hand, using Lemma 7 (with  $\mu$  and  $\mu'$  defined in (84)) and (82),  $|G_t(\xi)| \leq \rho^{t-s-1} \text{osc}(h)$ . We may therefore apply the Hoeffding inequality to show that

$$\mathbb{P} \left[ \left| N^{-1} \sum_{i=1}^N \omega_t^i G_t(\xi_t^i) \right| \geq \epsilon \right] \leq 2 \exp \left( - \frac{2N\epsilon^2}{\text{osc}^2(h) \rho^{2(t-s-1)}} \right)$$

showing that condition (III) is satisfied with constants that do not depend on  $t$ . Combining these results, Lemma 11 shows that, there exists some constants  $B$  and  $C$ , such that, for all  $s < t$ ,

$$\mathbb{P} \left[ \left| \frac{A_{s,t,T}(h)}{A_{s,t,T}(\mathbf{1})} - \frac{A_{s,t-1,T}(h)}{A_{s,t-1,T}(\mathbf{1})} \right| \geq \epsilon \right] \leq B \exp \left( -CN \frac{\epsilon^2}{\rho^{2(t-s-1)} \text{osc}^2(h)} \right).$$

The proof is concluded by applying Lemma 14.  $\square$

**6. A limiting expression of the variance of the marginal smoothing distribution.** In this section, we study the expression of the variance (49) for the FFBS approximation of the marginal smoothing distribution. In particular, we show that under the strong mixing condition (A4) the asymptotic variance of the marginal smoothing estimator  $\Gamma_{\chi,s:T|T}[\bar{h}]$ , where  $\bar{h}$  is defined in (81) has a finite limiting value as  $T \rightarrow \infty$  for a given value of  $s$ . We will also show that this variance is upper bounded uniformly in time, allowing to construct uniform confidence intervals.

The first step consists in showing that the asymptotic variance of the auxiliary particle filter has a finite limiting value as  $T \rightarrow \infty$ , and deriving an upper-bound for this limit.

**PROPOSITION 9.** *Assume (A1-4) hold for  $T = \infty$ . Then, with the notations of Proposition 3,*

$$\Gamma_{\chi,s|s}[h] \leq \frac{\sigma_+}{c_- \sigma_-} \sup_{r \geq 0} |\omega_r|_\infty |\vartheta_r|_\infty \frac{1}{1 - \rho^2} \text{osc}^2(h).$$

PROOF. Without loss of generality, we assume that  $\phi_{\chi,s|s}(h) = 0$ . Note that, for any  $r \in \{0, \dots, s\}$ , under (A4),

$$\frac{\int Q(x, dx') g_r(x') L_{r,s}(x', \mathbf{1})}{\phi_{\chi,r|r-1}[g_r(\cdot) L_{r,s}(\cdot, \mathbf{1})]} \leq \frac{\sigma_+}{\sigma_-}, \quad (86)$$

$$\left| \frac{L_{r,s}(x, h)}{L_{r,s}(x, \mathbf{1})} - \frac{\phi_{\chi,r|r}[L_{r,s}(\cdot, h)]}{\phi_{\chi,r|r}[L_{r,s}(\cdot, \mathbf{1})]} \right| \leq \rho^{s-r}. \quad (87)$$

We now bound  $V_{\chi,r,s}[h]$ ,  $r = 0 \dots s$ , defined (44) and (45). First, for  $r = 0$ , using that  $\chi(L_{0,s}(\cdot, h)) = 0$ , Proposition 5 and inequality (76) show that

$$\begin{aligned} V_{\chi,0,s}[h] &= \frac{\rho_0 \left( \omega_0^2(\cdot) L_{0,s}^2(\cdot, h) \right)}{\left( \phi_{\chi,0|-1}[g_0(\cdot) L_{0,s}(\cdot, \mathbf{1})] \right)^2} = \frac{\chi \left( \frac{d\chi}{d\rho_0}(\cdot) [g_0(\cdot) L_{0,s}(\cdot, h)]^2 \right)}{\left( \phi_{\chi,0|-1}[g_0(\cdot) L_{0,s}(\cdot, \mathbf{1})] \right)^2} \\ &= \chi \left[ \frac{d\chi}{d\rho_0}(\cdot) \left\{ \frac{g_0(\cdot) L_{0,s}(\cdot, \mathbf{1})}{\chi[g_0(\cdot) L_{0,s}(\cdot, \mathbf{1})]} \left[ \frac{L_{0,s}(\cdot, h)}{L_{0,s}(\cdot, \mathbf{1})} - \frac{\chi(L_{0,s}(\cdot, h))}{\chi(L_{0,s}(\cdot, \mathbf{1}))} \right] \right\}^2 \right] \\ &\leq \frac{|\omega_0|_\infty \sigma_+}{\chi(g_0) \sigma_-} \rho^{2s} \text{osc}^2(h) \leq \frac{|\omega_0|_\infty \sigma_+}{c_- \sigma_-} \rho^{2s} \text{osc}^2(h). \end{aligned}$$

Similarly, for  $r > 0$ , using that  $\phi_{\chi,r|r}[L_{r,s}(\cdot, h)] = 0$ , Eqs. (86) and (87) show that

$$\begin{aligned} &\frac{\phi_{\chi,r-1|r-1} \left[ \vartheta_r(\cdot) \int p_r(\cdot, x) \omega_r^2(\cdot, x) L_{r,s}^2(x, h) dx \right]}{\left( \phi_{\chi,r|r-1}[g_r(\cdot) L_{r,s}(\cdot, \mathbf{1})] \right)^2} \\ &= \phi_{\chi,r-1|r-1} \left[ \int Q(\cdot, dx) g_r(x) \omega_r(\cdot, x) \times \dots \right. \\ &\quad \left. \left\{ \frac{L_{r,s}(x, \mathbf{1})}{\phi_{\chi,r|r-1}[g_r(\cdot) L_{r,s}(\cdot, \mathbf{1})]} \left[ \frac{L_{r,s}(x, h)}{L_{r,s}(x, \mathbf{1})} - \frac{\phi_{\chi,r|r}[L_{r,s}(\cdot, h)]}{\phi_{\chi,r|r}[L_{r,s}(\cdot, \mathbf{1})]} \right] \right\}^2 \right] \\ &\leq \frac{|\omega_r|_\infty \sigma_+}{\phi_{\chi,r|r-1}[g_r(\cdot)] \sigma_-} \rho^{2(r-s)} \text{osc}^2(h) \leq \frac{|\omega_r|_\infty \sigma_+}{c_- \sigma_-} \rho^{2(r-s)} \text{osc}^2(h) \end{aligned}$$

which implies that  $V_{\chi,r,s}[h] \leq |\vartheta_r|_\infty (\sigma_+/c_- \sigma_-) \rho^{2(r-s)} \text{osc}^2(h) |\omega_r|_\infty$ . The result follows.  $\square$

We are now in position to state and prove the main result of this section, which provides a uniform bound for the variance of the particle estimator of the marginal smoothing distribution.

**THEOREM 10.** *Assume (A1-4) hold for  $T = \infty$ . Then, for any  $s \leq T$ ,*

$$\Gamma_{\chi,s|T}[\bar{h}] \leq \frac{\text{osc}^2(h)}{1 - \rho^2} \left( \frac{1}{c_-} \left( \frac{\sigma_+}{\sigma_-} \right)^2 \sup_{r \geq 0} |\omega_r|_\infty |\vartheta_r|_\infty + \frac{\sigma_+^4}{c_-^2 \sigma_-^3} \sup_{r \geq 0} |\omega_r|_\infty |\vartheta_r|_\infty |g_r|_\infty \right),$$

where the function  $\bar{h}$  and the covariance  $\Gamma_{\chi,s|T}[\bar{h}]$  are defined in (81) and (49), respectively.

PROOF. We bound the summands appearing in (49). Consider first the term  $\Gamma_{\chi,s|s} [L_{s,s,T}(\cdot, h)] / \phi_{\chi,s|s}^2 [L_{s,s,T}(\cdot, \mathbf{1})]$ . We first apply Proposition 9 to the function  $L_{s,s,T}(\cdot, \bar{h}) / \phi_{\chi,s|s} [L_{s,s,T}(\cdot, \mathbf{1})]$ . Using (76) and  $L_{s,s,T}(\cdot, \bar{h}) = h(\cdot) L_{s,s,T}(\cdot, \mathbf{1})$ ,

$$\frac{1}{\int \phi_{\chi,s|s}(dx_s) \left[ \frac{L_{s,s,T}(x_s, \mathbf{1})}{L_{s,s,T}(\mathbf{1})} \right]} \leq \frac{\sigma_+}{\sigma_-},$$

hence  $\text{osc} \left( L_{s,s,T}(\cdot, \bar{h}) / \phi_{\chi,s|s} [L_{s,s,T}(\cdot, \mathbf{1})] \right) \leq \frac{\sigma_+}{\sigma_-} \text{osc}(h)$ , and

$$\frac{\Gamma_{\chi,s|s} [L_{s,s,T}(\cdot, h)]}{\phi_{\chi,s|s}^2 [L_{s,s,T}(\cdot, \mathbf{1})]} \leq \frac{1}{c_-} \left( \frac{\sigma_+}{\sigma_-} \right)^2 \sup_{r \geq 0} |\omega_r|_\infty |\vartheta_r|_\infty \frac{\text{osc}^2(h)}{1 - \rho^2}. \quad (88)$$

Now, we write

$$\begin{aligned} & \frac{\phi_{\chi,t-1|t-1}(v_{s,t,T}(\cdot, h))}{\phi_{\chi,t|t-1}^2 [g_t(\cdot) L_{t,t,T}(\cdot, \mathbf{1})]} \\ &= \phi_{\chi,t-1|t-1} \left( \vartheta_t(\cdot) \int \frac{p_t(\cdot, x') \omega_t^2(\cdot, x') \phi_{\chi,s:t-1|t-1}^2 [\bar{h}(\cdot) q_{s,t-1}(\cdot, x')] L_{t,t,T}^2(x', \mathbf{1})}{\phi_{\chi,t-1|t-1}^2 [q(\cdot, x')] \phi_{\chi,t|t-1}^2 [g_t(\cdot) L_{t,t,T}(\cdot, \mathbf{1})]} dx' \right). \end{aligned}$$

We will show that

$$\left| \frac{\phi_{\chi,s:t-1|t-1} [\bar{h}(\cdot) q_{s,t-1}(\cdot, x')] L_{t,t,T}(x', \mathbf{1})}{\phi_{\chi,t|t-1} [g_t(\cdot) L_{t,t,T}(\cdot, \mathbf{1})]} \right| \leq \frac{\sigma_+^2}{\sigma_- c_-} \rho^{t-s} \text{osc}(h). \quad (89)$$

Using this inequality,

$$\begin{aligned} & \frac{\phi_{\chi,t-1|t-1}(v_{s,t,T}(\cdot, h)) \phi_{\chi,t-1|t-1}(\vartheta_t)}{\phi_{\chi,t|t-1}^2 [g_t(\cdot) L_{t,t,T}(\cdot, \mathbf{1})]} \\ & \leq \left( \frac{\sigma_+^2}{\sigma_- c_-} \right)^2 \rho^{2(t-s)} \text{osc}(h)^2 \phi_{\chi,t-1|t-1} \left( \vartheta_t(\cdot) \int \frac{p_t(\cdot, x') \omega_t^2(\cdot, x')}{\phi_{\chi,t-1|t-1}^2 [q(\cdot, x')]} dx' \right) |\vartheta_t|_\infty \\ & \leq \left( \frac{\sigma_+^2}{\sigma_- c_-} \right)^2 \rho^{2(t-s)} \text{osc}(h)^2 \phi_{\chi,t-1|t-1} \left( \int \frac{q(\cdot, x') g_t(x') \omega_t(\cdot, x')}{\phi_{\chi,t-1|t-1}^2 [q(\cdot, x')]} dx' \right) |\vartheta_t|_\infty \\ & \leq \left( \frac{\sigma_+^2}{\sigma_- c_-} \right)^2 \rho^{2(t-s)} \text{osc}(h)^2 \frac{|\omega_t|_\infty |\vartheta_t|_\infty |g_t|_\infty}{\sigma_-}, \end{aligned}$$

where we used the Fubini Theorem in the last step. Let us finally turn to the proof of the inequality (89).

$$\begin{aligned} & \frac{\phi_{\chi,s:t-1|t-1} [\bar{h}(\cdot) q_{s,t-1}(\cdot, x')] L_{t,t,T}(x', \mathbf{1})}{\phi_{\chi,t|t-1} [g_t(\cdot) L_{t,t,T}(\cdot, \mathbf{1})]} \\ &= \frac{\int \phi_{\chi,s|s}(dx_s) h(x_s) l_{s,t-1}(x_s, x_{t-1}) q(x_{t-1}, x') L_{t,t,T}(x', \mathbf{1}) dx_{t-1}}{\int \phi_{\chi,s|s}(dx_s) l_{s,t-1}(x_s, x_{t-1}) \left\{ \int g_t(x_t) q(x_{t-1}, x_t) L_{t,t,T}(x_t, \mathbf{1}) dx_t \right\} dx_{t-1}}. \end{aligned}$$

The last expression can be written  $A \times B$  with

$$A = \frac{\int \phi_{\chi,s|s}(dx_s) h(x_s) l_{s,t-1}(x_s, x_{t-1}) q(x_{t-1}, x') dx_{t-1}}{\int \phi_{\chi,s|s}(dx_s) l_{s,t-1}(x_s, x_{t-1}) q(x_{t-1}, x') dx_{t-1}}$$

and

$$B = \frac{\int \phi_{\chi,s|s}(\mathrm{d}x_s) l_{s,t-1}(x_s, x_{t-1}) q(x_{t-1}, x') L_{t,t,T}(x', \mathbf{1}) \mathrm{d}x_{t-1}}{\int \phi_{\chi,s|s}(\mathrm{d}x_s) l_{s,t-1}(x_s, x_{t-1}) \left\{ \int g_t(x_t) q(x_{t-1}, x_t) L_{t,t,T}(x_t, \mathbf{1}) \mathrm{d}x_t \right\} \mathrm{d}x_{t-1}}.$$

We will bound these two terms separately. Since  $\phi_{\chi,s|T}(h) = 0$ ,

$$A' = \frac{\int \phi_{\chi,s|s}(\mathrm{d}x_s) h(x_s) l_{s,t-1}(x_s, x_{t-1}) L_{t-1,t-1,T}(x_{t-1}, \mathbf{1}) \mathrm{d}x_{t-1}}{\int \phi_{\chi,s|s}(\mathrm{d}x_s) l_{s,t-1}(x_s, x_{t-1}) L_{t-1,t-1,T}(x_{t-1}, \mathbf{1}) \mathrm{d}x_{t-1}} = 0.$$

Thus, by Proposition 5,

$$|A| = |A - A'| \leq \rho^{t-s} \operatorname{osc}(h).$$

On the other hand, as  $q(x_{t-1}, x') \leq \sigma_+$ , as  $\int g_t(x_t) q(x_{t-1}, x_t) \mathrm{d}x_t \geq c_-$  and as for every  $x_t$  it holds that

$$\frac{L_{t,t,T}(x_t, \mathbf{1})}{L_{t,t,T}(x', \mathbf{1})} \geq \frac{\sigma_-}{\sigma_+},$$

$B$  is upper-bounded by  $\sigma_+^2 / (\sigma_- c_-)$ .  $\square$

## APPENDIX A: TECHNICAL RESULTS

LEMMA 11. *Assume that  $a_N, b_N, c_N, d_N$  and  $b$  are random variables such that there exist positive constants  $\beta, B_1, C_1, B_2, C_2, M$  such that*

- (I)  $|a_N/b_N - c_N/d_N| \leq M$ ,  $\mathbb{P}$ -a.s. and  $b \geq \beta$ ,  $\mathbb{P}$ -a.s.
- (II) For all  $\epsilon > 0$  and all  $N \geq 1$ ,  $\mathbb{P}(|b_N - b| > \epsilon) \leq B_1 e^{-C_1 N \epsilon^2}$ ,
- (III) For all  $\epsilon > 0$  and all  $N \geq 1$ ,  $\mathbb{P}(|a_N - (c_N/d_N)b_N| > \epsilon) \leq B_2 e^{-C_2 N (\frac{\epsilon}{M})^2}$

Then,

$$\mathbb{P}\left(\left|\frac{a_N}{b_N} - \frac{c_N}{d_N}\right| > \epsilon\right) \leq B_1 e^{-C_1 N (\frac{\epsilon\beta}{2M})^2} + B_2 e^{-C_2 N (\frac{\epsilon\beta}{2M})^2}$$

PROOF. Write

$$\begin{aligned} \left|\frac{a_N}{b_N} - \frac{c_N}{d_N}\right| &\leq b^{-1} \left|\frac{a_N}{b_N} - \frac{c_N}{d_N}\right| |b - b_N| + b^{-1} \left|a_N - \frac{c_N}{d_N} b_N\right| \\ &\leq \beta^{-1} M |b - b_N| + \beta^{-1} \left|a_N - \frac{c_N}{d_N} b_N\right| \quad a.s. \end{aligned}$$

Thus,

$$\left\{ \left|\frac{a_N}{b_N} - \frac{c_N}{d_N}\right| > \epsilon \right\} \subset \left\{ |b - b_N| > \frac{\epsilon\beta}{2M} \right\} \cup \left\{ \left|a_N - \frac{c_N}{d_N} b_N\right| > \frac{\epsilon\beta}{2M} \right\}$$

and the proof follows.  $\square$

LEMMA 12. *Let  $\nu$  be a measure and  $\{A_N(x)\}$  be a sequence of stochastic processes such that,*

1. for  $\nu$ -almost every  $x$ ,  $A_N(x) \xrightarrow{\mathbb{P}} a(x)$ ,

2. there exists a constant  $C$  and a  $\nu$ -integrable function  $h$ , for  $\nu$ -almost every  $x$ ,  $|A_N(x)| \leq Ch(x)$ .

Then,  $\int |A_N(x) - a(x)|\nu(dx) \xrightarrow{N \rightarrow \infty} 0$  in  $L^1(\nu)$ .

PROOF. Under the stated assumptions  $\mathbb{E}|A_N(x) - a(x)| \rightarrow 0$  as  $N$  goes to infinity  $\nu$ -a.e.. On the other hand,  $\nu$ -a.e.,  $|A_N(x) - a(x)| \leq 2Ch(x)$ . The proof follows from the Fubini Theorem and the dominated convergence Theorem,

$$\mathbb{E} \int |A_N(x) - a(x)|\nu(dx) = \int \mathbb{E} |A_N(x) - a(x)| \nu(dx) \xrightarrow{N \rightarrow \infty} 0.$$

□

LEMMA 13. Assume that (A1-2) hold for some  $T$ . Let  $\{\Upsilon_N(x)\}$  be a sequence of stochastic processes and  $v$  a function such that (i) there exists a constant  $C < \infty$  such that, for all  $N$ ,  $|\Upsilon_N|_\infty \leq v_\infty$  and  $|v|_\infty \leq C$  and (ii) for all  $M \geq 0$ ,  $\sup_{|x| \leq M} |\Upsilon_N(x) - v(x)| \xrightarrow{P}_{N \rightarrow \infty} 0$ . Then,  $t \leq T$ ,  $\Omega_t^{-1} \sum_{\ell=1}^N \omega_t^\ell \Upsilon_N(\xi_t^\ell) \xrightarrow{P}_{N \rightarrow \infty} \phi_{\chi, t|t}(v)$ .

PROOF. Write  $\Omega_N^{-1} \sum_{t=1}^N \omega_t^\ell \{\Upsilon_N(\xi_t^\ell) - v(\xi_t^\ell)\} = S_{N,1} + S_{N,2}$ , with  $S_{N,1} \stackrel{\text{def}}{=} \Omega_N^{-1} \sum_{t=1}^N \omega_t^\ell \{\Upsilon_N(\xi_t^\ell) - v(\xi_t^\ell)\} \mathbf{1}\{|\xi_t^\ell| \leq M\}$  and  $S_{N,2} \stackrel{\text{def}}{=} \Omega_N^{-1} \sum_{t=1}^N \omega_t^\ell \{\Upsilon_N(\xi_t^\ell) - v(\xi_t^\ell)\} \mathbf{1}\{|\xi_t^\ell| > M\}$ . Since  $S_{N,1} \leq \sup_{|x| \leq M} |\Upsilon_N(x) - v(x)|$ , assumption (ii) implies that  $S_{N,1} \xrightarrow{P}_{N \rightarrow \infty} 0$ . On the other hand,  $S_{N,2} \leq 2C \Omega_t^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathbf{1}\{|\xi_t^\ell| > M\}$ . By Proposition 1,  $\Omega_t^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathbf{1}\{|\xi_t^\ell| > M\} \xrightarrow{P}_{N \rightarrow \infty} \phi_{\chi, t|t}(\mathbf{1}\{|\cdot| > M\})$ . the proof follows since  $\lim_{M \rightarrow \infty} \phi_{\chi, t|t}(\mathbf{1}\{|\cdot| > M\}) = 0$ . □

LEMMA 14. Let  $\{Y_{n,i}\}_{i=1}^n$  be a triangular array of random variables such that there exist constants  $B > 0$ ,  $C > 0$  and  $\rho$ ,  $0 < \rho < 1$  such that, for all  $n$ ,  $i \in \{1, \dots, n\}$  and  $\epsilon > 0$ ,

$$\mathbb{P}(|Y_{n,i}| \geq \epsilon) \leq B e^{-C\epsilon^2 \rho^{-2i}}.$$

Then, there exists  $\bar{B}$  and  $\bar{C}$  such that, for any  $n$  and  $\epsilon > 0$ ,

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_{n,i}\right| \geq \epsilon\right) \leq \bar{B} e^{-\bar{C}\epsilon^2}.$$

PROOF. Denote by  $S \stackrel{\text{def}}{=} \sum_{i=1}^\infty \sqrt{i} \rho^i$ . It is plain to see that

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_{n,i}\right| \geq \epsilon\right) \leq \sum_{i=1}^n \mathbb{P}\left(|Y_{n,i}| \geq \epsilon S^{-1} \sqrt{i} \rho^i\right) \leq B \sum_{i=1}^n e^{-CS^{-1}\epsilon^2 i}.$$

Set  $\epsilon_0 > 0$ . The proof follows by noting that, for any  $\epsilon \geq \epsilon_0$ ,

$$\sum_{i=1}^n e^{-CS^{-1}\epsilon^2 i} \leq (1 - e^{CS^{-1}\epsilon_0^2})^{-1} e^{CS^{-1}\epsilon_0^2} e^{-CS^{-1}\epsilon^2}.$$

□

## REFERENCES

- [1] ANDERSON, B. D. O. AND MOORE, J. B. (1979). *Optimal Filtering*. Prentice-Hall.
- [2] CAPPÉ, O., MOULINES, E., AND RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer. <http://perso.telecom-paristech.fr/cappe/ihmm/>.
- [3] CHOPIN, N. (2004). Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *Ann. Statist.* **32**, 6, 2385–2411.
- [4] CORNEBISE, J., MOULINES, E., AND OLSSON, J. (2008). Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **18**, 4, 461–480.
- [5] DEL MORAL, P. (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer.
- [6] DEL MORAL, P. AND GUIONNET, A. (1998). Large deviations for interacting particle systems: applications to non-linear filtering. *Stoch. Proc. App.* **78**, 69–95.
- [7] DEL MORAL, P., LEDOUX, M., AND MICLO, L. (2003). On contraction properties of Markov kernels. *Probab. Theory Related Fields* **126**, 3, 395–420.
- [8] DOBRUSHIN, R. (1956). Central limit theorem for non-stationary Markov chains. I. *Teor. Veroyatnost. i Primenen.* **1**, 72–89.
- [9] DOUC, R. AND MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36**, 5, 2344–2376.
- [10] DOUC, R., MOULINES, E., AND OLSSON, J. (2008). Optimality of the auxiliary particle filter. *Probab. Math. Statist.* **28**, 2. To appear.
- [11] DOUCET, A., DE FREITAS, N., AND GORDON, N., Eds. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- [12] DOUCET, A., GODSILL, S., AND ANDRIEU, C. (2000). On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10**, 197–208.
- [13] GODSILL, S. J., DOUCET, A., AND WEST, M. (2004). Monte carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.* **50**, 438–449.
- [14] GORDON, N., SALMOND, D., AND SMITH, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar Signal Process.* **140**, 107–113.
- [15] HANDSCHIN, J. AND MAYNE, D. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. In *Int. J. Control.* Vol. **9**. 547–559.
- [16] HÜRZELER, M. AND KÜNSCH, H. R. (1998). Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.* **7**, 175–193.
- [17] KAILATH, T., SAYED, A., AND HASSIBI, B. (2000). *Linear Estimation*. Prentice-Hall.
- [18] KITAGAWA, G. (1996). Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **1**, 1–25.
- [19] KLAAS, M., BRIERS, M., DE FREITAS, N., DOUCET, A., MASKELL, S., AND LANG, D. (2006). Fast particle smoothing: If i had a million particles. In *23rd Int. Conf. Machine Learning (ICML)*. Pittsburgh, Pennsylvania.
- [20] KÜNSCH, H. R. (2005). Recursive Monte-Carlo filters: algorithms and theoretical analysis. *Ann. Statist.* **33**, 5, 1983–2021.
- [21] LIU, J. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York.
- [22] LIU, J. AND CHEN, R. (1998). Sequential Monte-Carlo methods for dynamic systems. *J. Am. Statist. Assoc.* **93**, 443, 1032–1044.
- [23] PITT, M. K. AND SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.* **94**, 446, 590–599.
- [24] RISTIC, B., ARULAMPALAM, M., AND GORDON, A. (2004). *Beyond Kalman Filters: Particle Filters for Target Tracking*. Artech House.
- [25] RUBIN, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm (discussion of Tanner and Wong). *J. Am. Statist. Assoc.* **82**, 543–546.

TÉLÉCOM SUDPARIS,  
 9 RUE CHARLES FOURIER,  
 91011 EVRY-CEDEX, FRANCE  
 E-MAIL: randal.douc@it-sudparis.eu

TÉLÉCOM PARISTECH,  
 46 RUE BARRAULT,  
 75634 PARIS, CÉDEX 13, FRANCE  
 E-MAIL: eric.moulines@telecom-paristech.fr



CENTER OF MATHEMATICAL SCIENCES,  
LUND UNIVERSITY,  
BOX 118 SE-22100 LUND, SWEDEN  
E-MAIL: jimmy@maths.lth.se