On the two-filter approximations of marginal smoothing distributions in general state space models

Thi Ngoc Minh Nguyen, Sylvain Le Corff, Éric Moulines

On the two-filter approximations of marginal smoothing distributions in general state space models

Thi Ngoc Minh Nguyen[*] Sylvain Le Corff[†] Eric Moulines[‡]

Abstract

A prevalent problem in general state space models is the approximation of the smoothing distribution of a state conditional on the observations from the past, the present, and the future. The aim of this paper is to provide a rigorous analysis of such approximations of smoothed distributions provided by the two-filter algorithms. We extend the results available for the approximation of smoothing distributions to these two-filter approaches which combine a forward filter approximating the filtering distributions with a backward information filter approximating a quantity proportional to the posterior distribution of the state given future observations.

1 Introduction

State-space models play a key role in a large variety of disciplines such as engineering, econometrics, computational biology or signal processing, see [9, 8] and references therein. This paper provides a nonasymptotic analysis of a Sequential Monte Carlo Method (SMC) which aims at performing optimal smoothing in nonlinear and non Gaussian state space models. Given two measurable spaces $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$, consider a bivariate stochastic process $\{(X_t, Y_t)\}_{t \geq 0}$ taking values in the product space $(\mathsf{X} \times \mathsf{Y}, \mathcal{X} \otimes \mathcal{Y})$, where the hidden state sequence $\{X_t\}_{t \geq 0}$ is observed only through the observation process $\{Y_t\}_{t \geq 0}$. Statistical inference in general state space models usually involves the computation of conditional distributions of some unobserved states given a set of observations. These posterior distributions are crucial to compute smoothed expectations of additive functionals which appear naturally for maximum likelihood parameter inference in hidden Markov models (computation of the Fisher score or of the intermediate quantity of the Expectation Maximization algorithm), see [3, Chapter 10 and 11], [17, 25, 20, 21].

Nevertheless, exact computation of the filtering and smoothing distributions is possible only for linear and Gaussian state spaces or when the state space X is finite. This paper focuses on particular instances of Sequential Monte Carlo methods which approximate sequences of distributions in a general state space X with random samples, named particles, associated with nonnegative importance weights. Those particle filters and smoothers rely on the combination of sequential importance sampling steps to propagate particles in the state space and importance resampling steps to duplicate or discard particles according to their importance weights. The first implementation of these SMC methods, introduced in

[*]LTCI, CNRS and Télécom ParisTech.
[†]Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay.
[‡]Centre de Mathématiques Appliquées, Ecole Polytechnique.

[14, 18], propagates the particles using the Markov kernel of the hidden process $\{X_t\}_{t\geq 0}$ and uses a multinomial resampling step based on the importance weights to select particles at each time step. An interesting feature of this *Poor man's smoother* is that it provides an approximation of the joint smoothing distribution by storing the ancestral line of each particle with a complexity growing only linearly with the number N of particles, see for instance [4]. However, this smoothing algorithm has a major shortcoming since the successive resampling steps induce an important depletion of the particle trajectories. This degeneracy of the particle sequences leads to trajectories sharing a common ancestor path; see [25, 16] for a discussion.

Approximations of the smoothing distributions may also be obtained using the forward filtering backward smoothing decomposition in general state space models. The Forward Filtering Backward Smoothing algorithm (FFBS) and the Forward Filtering Backward Simulation algorithm (FFBSi) developed respectively in [18, 15, 10] and [13] avoid the path degeneracy issue of the *Poor man's smoother* at the cost of a computational complexity growing with N^2. Both algorithms rely on a forward pass which produces a set of particles and weights approximating the sequence of filtering distributions up to time T. Then, the backward pass of the FFBS algorithm modifies all the weights computed in the forward pass according to the so-called backward decomposition of the smoothing distribution keeping all the particles fixed. On the other hand, the FFBSi algorithm samples independently particle trajectories among all the possible paths produced by the forward pass. It is shown in [22, 2, 5] that the FFBS algorithm can be implemented using only a forward pass when approximating smoothed expectations of additive functionals but with a complexity still growing quadratically with N. Under the mild assumption that the transition density of the hidden chain $\{X_t\}_{t\geq 0}$ is uniformly bounded above, [6] proposed an accept-reject mechanism to implement the FFBSi algorithm with a complexity growing only linearly with N. Concentration inequalities, controls of the L_q-norm of the deviation between smoothed functionals and their approximations and Central Limit Theorems (CLT) for the FFBS and the FFBSi algorithms have been established in [5, 6, 11].

Recently, [23] proposed a new SMC algorithm, the particle-based rapid incremental smoother (PaRIS), to approximate online, using only a forward pass, smoothed expectations of additive functionals. The crucial feature of this algorithm is that its complexity grows only linearly with N as it samples on-the-fly particles distributed according to the backward dynamics of the hidden chain conditionally on the observations $Y_0, \ldots, Y_T$. The authors show concentration inequalities and CLT for the estimators provided by the PaRIS algorithm.

In this paper, we extend the theoretical results available for the SMC approximations of smoothing distributions to the estimators given by the two-filter algorithms. These methods were first introduced in the particle filter literature by [18] and developed further by [1] and [12]. The two-filter approach combines the output of two independent filters, one that evolves forward in time and approximates the filtering distributions and another that evolves backward in time approximating a quantity proportional to the posterior distribution of a state given future observations. In [12], the authors introduced a proposal mechanism leading to algorithms whose complexity grows linearly with the number of particles. An algorithm similar to the algorithm of [1] may also be implemented with an $\mathcal{O}(N)$ computational complexity following the same idea. We analyze all these algorithms which approximate the marginal smoothing distributions (smoothing distributions of one state given all the observations) and provide concentration inequalities as well as CLT.

This paper is organized as follows. Section 2 introduces the different particle approximations of the marginal smoothing distributions given by the two-filter algorithms. Sections 3 and 4 provide exponential deviation inequalities and CLT for the particle approximations under mild assumptions on the hidden Markov chain. Under additional *strong mixing assumptions*, it is shown that the results of Section 3 are uniform in time and that the asymptotic variance in Section 4 may be uniformly bounded in time. All proofs are postponed to Section 5.

Notations and conventions

Let X and Y be two general state-spaces endowed with countably generated σ-fields $\mathcal{X}$ and $\mathcal{Y}$. $\mathbb{F}_b(\mathsf{X},\mathcal{X})$ is the set of all real valued bounded measurable functions on $(\mathsf{X},\mathcal{X})$. Q is a Markov transition kernel defined on $\mathsf{X}\times\mathcal{X}$ and $\{g_t\}_{t\geq 0}$ a family of positive functions defined on X. For any $x\in\mathsf{X}$, $Q(x,\cdot)$ has a density $q(x,\cdot)$ with respect to a measure λ on $(\mathsf{X},\mathcal{X})$. The oscillation of a real valued function defined on a space Z is given by: $\mathrm{osc}\,(h) := \sup_{z,z'\in\mathsf{Z}} |h(z)-h(z')|$.

2 The two-filter algorithms

For any measurable function h on X^{t-s+1}, probability distribution χ on $(\mathsf{X},\mathcal{X})$, $T\geq 0$ and $0\leq s\leq t\leq T$, define the joint smoothing distribution by:

$$\phi_{\chi,s:t|T}[h] := \frac{\int \chi(\mathrm{d}x_0)g_0(x_0)\prod_{u=1}^{T} Q(x_{u-1},\mathrm{d}x_u)g_u(x_u)h(x_{s:t})}{\int \chi(\mathrm{d}x_0)g_0(x_0)\prod_{u=1}^{T} Q(x_{u-1},\mathrm{d}x_u)g_u(x_u)}\;, \tag{1}$$

where $a_{u:v}$ is a short-hand notation for $\{a_s\}_{s=u}^{v}$. In the following we use the notations $\phi_{\chi,s|T} := \phi_{\chi,s:s|T}$ and $\phi_{\chi,t} := \phi_{\chi,t:t|t}$. The aim of this paper is to provide a rigorous analysis of the performance of SMC algorithms approximating the sequence $\phi_{\chi,s|T}$ for $0\leq s\leq T$. The algorithms analyzed in this paper are based on the *two-filter formula* introduced in [1, 12], which we now detail.

2.1 Forward filter

Let $\{\xi_0^\ell\}_{\ell=1}^N$ be i.i.d. and distributed according to the instrumental distribution ρ_0 and define the importance weights

$$\omega_0^\ell := \frac{\mathrm{d}\chi}{\mathrm{d}\rho_0}(\xi_0^\ell)\,g_0(\xi_0^\ell)\;.$$

For any $h\in\mathbb{F}_b(\mathsf{X},\mathcal{X})$,

$$\phi^N_{\chi,0}[h] := \Omega_0^{-1}\sum_{\ell=1}^{N}\omega_0^\ell h(\xi_0^\ell)\;,\quad \text{where}\quad \Omega_0 := \sum_{\ell=1}^{N}\omega_0^\ell\;,$$

is a consistent estimator of $\phi_{\chi,0}[h]$, see for instance [4]. Then, based on $\{(\xi_{s-1}^\ell,\omega_{s-1}^\ell)\}_{\ell=1}^N$ a new set of particles and importance weights is obtained using the auxiliary sampler introduced in [24]. Pairs $\{(I_s^\ell,\xi_s^\ell)\}_{\ell=1}^N$ of indices and particles are simulated independently from the instrumental distribution with density on $\{1,\ldots,N\}\times\mathsf{X}$:

$$\pi_{s|s}(\ell,x) \propto \omega_{s-1}^\ell\vartheta_s(\xi_{s-1}^\ell)p_s(\xi_{s-1}^\ell,x)\;, \tag{2}$$

where ϑ_s is the adjustment multiplier weight function and p_s is a Markovian transition density. For any $\ell \in \{1, \dots, N\}$, ξ_s^ℓ is associated with the importance weight defined by:

$$\omega_s^\ell := \frac{q(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell) g_s(\xi_s^\ell)}{\vartheta_s(\xi_{s-1}^{I_s^\ell}) p_s(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell)} \tag{3}$$

to produce the following approximation of $\phi_{\chi,s}[h]$:

$$\phi_{\chi,s}^N[h] := \Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell h(\xi_s^\ell) \;, \quad \text{where} \quad \Omega_s := \sum_{\ell=1}^N \omega_s^\ell \;.$$

2.2 Backward filter

Let $\{\gamma_t\}_{t\geq 0}$ be a family of positive measurable functions such that, for all $t \in \{0, \dots, T\}$,

$$\int \gamma_t(x_t)\, \mathrm{d}x_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1})\, Q(x_{u-1}, \mathrm{d}x_u) \right] g_T(x_T) < \infty \;. \tag{4}$$

Following [1], for any $0 \leq t \leq T$ we introduce the backward filtering distribution $\psi_{\gamma,t|T}$ on $\mathcal{X}$ (referred to as the *backward information filter* in [18] and [1]) defined, for any $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$, by:

$$\psi_{\gamma,t|T}[h] := \frac{\int \gamma_t(x_t)\, \mathrm{d}x_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1})\, Q(x_{u-1}, \mathrm{d}x_u) \right] g_T(x_T) h(x_t)}{\int \gamma_t(x_t)\, \mathrm{d}x_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1})\, Q(x_{u-1}, \mathrm{d}x_u) \right] g_T(x_T)} \;.$$

If the distribution of X_t has probability density function γ_t, then $\psi_{\gamma,t|T}$ is the conditional distribution of X_t given $Y_{t:T}$. Contrary to [1] or [12], $\int \gamma_t(x_t) \mathrm{d}x_t$ may be infinite. The only requirement about the nonnegative functions $\{\gamma_t\}_{t\geq 0}$ is the condition (4) and the fact that γ_t should be available in closed form. Here γ_t is a possibly improper prior introduced to make $\psi_{\gamma,t|T}$ a proper posterior distribution, which is of key importance when producing particle approximations of such quantities. For $0 \leq t \leq T-1$, the backward information filter is computed by the recursion

$$\psi_{\gamma,t|T}[h] \propto \int \psi_{\gamma,t+1|T}(\mathrm{d}x_{t+1}) \left[\gamma_t(x_t) g_t(x_t) \frac{q(x_t, x_{t+1})}{\gamma_{t+1}(x_{t+1})} \right] h(x_t)\, \mathrm{d}x_t \;, \tag{5}$$

in the backward time direction. (5) is analogous to the forward filter recursion and particle approximations of the backward information filter can be obtained similarly. Using the definition of the forward filtering distribution at time $s-1$ and the backward information filter at time $s+1$, the marginal smoothing distribution may be expressed as

$$\phi_{\chi,s|T}[h] \propto \int \phi_{\chi,s-1}(\mathrm{d}x_{s-1}) \psi_{\gamma,s+1|T}(\mathrm{d}x_{s+1}) q(x_{s-1}, x_s) g_s(x_s) \frac{q(x_s, x_{s+1})}{\gamma_{s+1}(x_{s+1})} h(x_s) \mathrm{d}x_s \;. \tag{6}$$

We now describe the Sequential Monte Carlo methods used to approximate the recursion (5) in [1], [12]. Let $\check{\rho}_T$ be an instrumental probability density on X and $\{\check{\xi}_{T|T}^i\}_{i=1}^N$ be i.i.d. random variables such that $\check{\xi}_{T|T}^i \sim \check{\rho}_T$ and define

$$\check{\omega}_{T|T}^i := \frac{g_T(\check{\xi}_{T|T}^i) \gamma_T(\check{\xi}_{T|T}^i)}{\check{\rho}_T(\check{\xi}_{T|T}^i)} \;.$$

Let now $\{(\breve{\xi}^i_{t+1|T}, \breve{\omega}^i_{t+1|T})\}_{i=1}^N$ be a weighted sample targeting the backward information filter distribution $\psi_{\gamma,t+1|T}[h]$ at time $t+1$:

$$\psi^N_{\gamma,t+1|T}[h] := \breve{\Omega}^{-1}_{t+1|T} \sum_{i=1}^N \breve{\omega}^i_{t+1|T} h(\breve{\xi}^i_{t+1|T}) \;, \quad \text{where} \quad \breve{\Omega}_{t+1|T} := \sum_{i=1}^N \breve{\omega}^i_{t+1|T} \;.$$

Plugging this approximation into (5) yields the target probability density

$$\hat{\psi}^{\text{tar}}_{\gamma,t|T}(x_t) \propto \sum_{i=1}^N \breve{\omega}^i_{t+1|T} \left[\gamma_t(x_t) g_t(x_t) \frac{q(x_t, \breve{\xi}^i_{t+1|T})}{\gamma_{t+1}(\breve{\xi}^i_{t+1|T})} \right] \;,$$

which is the marginal probability density function of x_t of the joint density

$$\hat{\psi}^{\text{aux}}_{\gamma,t|T}(i, x_t) \propto \frac{\breve{\omega}^i_{t+1|T}}{\gamma_{t+1}(\breve{\xi}^i_{t+1|T})} \gamma_t(x_t) g_t(x_t) q(x_t, \breve{\xi}^i_{t+1|T}) \;.$$

A particle approximation of the backward information filter at time t can be derived by choosing an adjustment weight function $\vartheta_{t|T}$ and an instrumental density kernel $r_{t|T}$, and simulating $\{(\breve{I}^i_t, \breve{\xi}^i_{t|T})\}_{i=1}^N$ from the instrumental probability density on $\{1, \ldots, N\} \times \mathsf{X}$ given by

$$\pi_{t|T}(i, x_t) \propto \frac{\breve{\omega}^i_{t+1|T} \vartheta_{t|T}(\breve{\xi}^i_{t+1|T})}{\gamma_{t+1}(\breve{\xi}^i_{t+1|T})} r_{t|T}(\breve{\xi}^i_{t+1|T}, x_t) \;. \tag{7}$$

Subsequently, the particles are associated with the importance weights

$$\breve{\omega}^i_{t|T} := \frac{\gamma_t(\breve{\xi}^i_{t|T}) g_t(\breve{\xi}^i_{t|T}) q(\breve{\xi}^i_{t|T}, \breve{\xi}^{\breve{I}^i_t}_{t+1|T})}{\vartheta_{t|T}(\breve{\xi}^{\breve{I}^i_t}_{t+1|T}) r_{t|T}(\breve{\xi}^{\breve{I}^i_t}_{t+1|T}, \breve{\xi}^i_{t|T})} \;. \tag{8}$$

Ideally, a fully adapted version of the auxiliary backward information filter is obtained by using the adjustment weights $\vartheta^\star_{t|T}(x) = \int \gamma_t(x_t) g_t(x_t) q(x_t, x) \, \mathrm{d}x_t$ and the proposal kernel density

$$r^\star_{t|T}(x, x_t) = \gamma_t(x_t) g_t(x_t) \frac{q(x_t, x)}{\vartheta^\star_{t|T}(x)} \;,$$

yielding uniform importance weights. Such a solution is most likely to be cumbersome from a computational perspective.

2.3 Two-filter approximations of the marginal smoothing distributions

Plugging the particle approximations of the forward and backward filter distributions into (6) provides the following mixture approximation of the smoothing distribution:

$$\hat{\phi}^{\text{tar}}_{\chi,s|T}(x_s) \propto \sum_{i=1}^N \sum_{j=1}^N \frac{\omega^i_{s-1} \breve{\omega}^j_{s+1|T}}{\gamma_{s+1}(\breve{\xi}^j_{s+1|T})} q(\xi^i_{s-1}, x_s) g_s(x_s) q(x_s, \breve{\xi}^j_{s+1|T}) \;. \tag{9}$$

Following the $\mathsf{TwoFilt}_{fwt}$ algorithm of Fearnhead, Wyncoll and Tawn [12], the probability density (9) might be seen as the marginal density of x_s obtained from the joint density on the product space $\{1,\ldots,N\}^2 \times \mathsf{X}$ given by

$$\hat{\phi}^{\text{aux}}_{\chi,s|T}(i,j,x_s) \propto \frac{\omega^i_{s-1}\breve{\omega}^j_{s+1|T}}{\gamma_{s+1}(\breve{\xi}^j_{s+1|T})} q(\xi^i_{s-1},x_s)g_s(x_s)q(x_s,\breve{\xi}^j_{s+1|T}) \; . \tag{10}$$

The $\mathsf{TwoFilt}_{fwt}$ algorithm draws a set $\{(I^\ell_s,\breve{I}^\ell_s,\tilde{\xi}^\ell_{s|T})\}_{\ell=1}^N$ of indices and particle positions from the instrumental density

$$\pi_{s|T}(i,j,x_s) \propto \frac{\omega^i_{s-1}\tilde{\vartheta}_{s|T}(\xi^i_{s-1},\breve{\xi}^j_{s+1|T})\breve{\omega}^j_{s+1|T}}{\gamma_{s+1}(\breve{\xi}^j_{s+1|T})} \tilde{r}_{s|T}(\xi^i_{s-1},\breve{\xi}^j_{s+1|T};x_s) \; , \tag{11}$$

where, as above, $\tilde{\vartheta}_{s|T}(x,x')$ is an adjustment multiplier weight function (which now depends on the forward and backward particles) and $\tilde{r}_{s|T}$ is an instrumental kernel. We then associate with each draw $(I^\ell_s,\breve{I}^\ell_s,\tilde{\xi}^\ell_{s|T})$ the importance weight

$$\tilde{\omega}^\ell_{s|T} := \frac{q(\xi^{I^\ell_s}_{s-1},\tilde{\xi}^\ell_{s|T})g_s(\tilde{\xi}^\ell_{s|T})q(\tilde{\xi}^\ell_{s|T},\breve{\xi}^{\breve{I}^\ell_s}_{s+1|T})}{\tilde{\vartheta}_{s|T}(\xi^{I^\ell_s}_{s-1},\breve{\xi}^{\breve{I}^\ell_s}_{s+1|T})\tilde{r}_{s|T}(\xi^{I^\ell_s}_{s-1},\breve{\xi}^{\breve{I}^\ell_s}_{s+1|T};\tilde{\xi}^\ell_{s|T})} \; , \quad \tilde{\Omega}_{s|T} := \sum_{\ell=1}^N \tilde{\omega}^\ell_{s|T} \; . \tag{12}$$

Then, the auxiliary indices $\{(I^\ell_s,\breve{I}^\ell_s)\}_{\ell=1}^N$ are discarded and $\{(\tilde{\omega}^\ell_{s|T},\tilde{\xi}^\ell_{s|T})\}_{\ell=1}^N$ approximate the target smoothing density $\hat{\phi}^{\text{tar}}_{\chi,s|T}$. Mimicking the arguments in [15] and further developed in [19], the auxiliary particle filter is fully adapted if the adjustment weight function is $\vartheta^\star_{s|T}(x,x') = \int q(x,x_s)g_s(x_s)q(x_s,x')\,\mathrm{d}x_s$ and the instrumental kernel is

$$r^\star_{s|T}\,(x,x';x_s) = q(x,x_s)g_s(x_s)q(x_s,x')/\vartheta^\star_{s|T}(x,x') \; .$$

Except in simple scenarios, simulating from the fully adapted auxiliary filter is computationally intractable.

Instead of considering the target distribution (9) as the marginal of the auxiliary distribution (10) over pairs of indices, the $\mathsf{TwoFilt}_{bdm}$ algorithm of [1] uses the following *partial* auxiliary distributions having densities,

$$\hat{\phi}^{\text{aux,f}}_{s|T}(i,x_s) \propto \omega^i_{s-1}q(\xi^i_{s-1},x_s)g_s(x_s)\sum_{j=1}^N \frac{\breve{\omega}^j_{s+1|T}}{\gamma_{s+1}(\breve{\xi}^j_{s+1|T})} q(x_s,\breve{\xi}^j_{s+1|T}) \; ,$$
$$\hat{\phi}^{\text{aux,b}}_{s|T}(j,x_s) \propto \frac{\breve{\omega}^j_{s+1|T}}{\gamma_{s+1}(\breve{\xi}^j_{s+1|T})} q(x_s,\breve{\xi}^j_{s+1|T})g_s(x_s)\sum_{i=1}^N \omega^i_{s-1}q(\xi^i_{s-1},x_s) \; .$$

Since $\hat{\phi}^{\text{tar}}_{\chi,s|T}$ is the marginal probability density of the partial auxiliary distributions $\hat{\phi}^{\text{aux,f}}_{s|T}$ and $\hat{\phi}^{\text{aux,b}}_{s|T}$ with respect to the forward and the backward particle indices, respectively, we may sample from $\hat{\phi}^{\text{tar}}_{\chi,s|T}$ by simulating instead $\{(I^\ell_s,\xi^\ell_s)\}_{\ell=1}^N$ or $\{(\breve{I}^\ell_s,\breve{\xi}^\ell_{s|T})\}_{\ell=1}^N$ from the instrumental probability density functions

$$\pi^{\text{f}}_{s|T}(i,x_s) \propto \omega^i_{s-1}\vartheta_s(\xi^i_{s-1})p_s(\xi^i_{s-1},x_s) \; ,$$
$$\pi^{\text{b}}_{s|T}(j,x_s) \propto \vartheta_{s|T}(\breve{\xi}^j_{s+1|T})\breve{\omega}^j_{s+1|T}r_{s|T}(\breve{\xi}^j_{s+1|T},x_s)/\gamma_{s+1}(\breve{\xi}^j_{s+1|T}) \; ,$$

where (ϑ_s, p_s) and $(\vartheta_{s+1|T}, r_{s|T})$ are the adjustment multiplier weight functions and the instrumental kernels used in the forward and backward passes. In this case the algorithm uses the particles obtained when approximating the forward filter and backward information filter to provide two different weighted samples $\{(\tilde{\omega}_{s|T}^{i,\mathrm{f}}, \xi_s^i)\}_{i=1}^N$ and $\{(\tilde{\omega}_{s|T}^{i,\mathrm{b}}, \check{\xi}_{s|T}^i)\}_{i=1}^N$ targeting the marginal smoothing distribution, where the *forward* $\{\tilde{\omega}_{s|T}^{i,\mathrm{f}}\}_{i=1}^N$ and *backward* $\{\tilde{\omega}_{s|T}^{i,\mathrm{b}}\}_{i=1}^N$ importance weights are given by

$$\tilde{\omega}_{s|T}^{i,\mathrm{f}} := \omega_s^i \sum_{j=1}^N \check{\omega}_{s+1|T}^j q(\xi_s^i, \check{\xi}_{s+1|T}^j)/\gamma_{s+1}(\check{\xi}_{s+1|T}^j) , \quad \tilde{\Omega}_{s|T}^{\mathrm{f}} := \sum_{j=1}^N \tilde{\omega}_{s|T}^{j,\mathrm{f}} , \tag{13}$$

$$\tilde{\omega}_{s|T}^{j,\mathrm{b}} := \check{\omega}_{s|T}^j \sum_{i=1}^N \omega_{s-1}^i q(\xi_{s-1}^i, \check{\xi}_{s|T}^j)/\gamma_s(\check{\xi}_{s|T}^j) , \qquad \tilde{\Omega}_{s|T}^{\mathrm{b}} := \sum_{j=1}^N \tilde{\omega}_{s|T}^{j,\mathrm{b}} . \tag{14}$$

An important drawback of these algorithms is that the computation of the forward and backward importance weights grows quadratically with the number N of particles.

2.4 $\mathcal{O}(N)$ approximations of the marginal smoothing distributions

In [12], the authors introduced a proposal mechanism in (11) such that the indices $(I_s, \check{I}_s)$ of the forward and backward particles chosen at time $s-1$ and $s+1$ are sampled independently. Such choices lead to algorithms whose complexity grows linearly with the number of particles. The $\mathcal{O}(N)$ algorithm displayed in [12] suggests to use an adjustment multiplier weight function in (11) such that I_s and $\check{I}_s$ are chosen according to the same distributions as the indices sampled in the forward filter and in the backward information filter. It is done in [12] by choosing $\tilde{\vartheta}_{s|T}(x, x') = \vartheta_s(x)\vartheta_{s|T}(x')$ so that (11) becomes

$$\pi_{s|T}(i, j, x_s) \propto \omega_{s-1}^i \vartheta_s(\xi_{s-1}^i) \frac{\vartheta_{s|T}(\check{\xi}_{s+1|T}^j)\check{\omega}_{s+1|T}^j}{\gamma_{s+1}(\check{\xi}_{s+1|T}^j)} \tilde{r}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x_s) . \tag{15}$$

In this case, the importance weight (12) associated with each draw $(I_s^\ell, \check{I}_s^\ell, \tilde{\xi}_{s|T}^\ell)$ is given by

$$\tilde{\omega}_{s|T}^\ell := \frac{q(\xi_{s-1}^{I_s^\ell}, \tilde{\xi}_{s|T}^\ell) g_s(\tilde{\xi}_{s|T}^\ell) q(\tilde{\xi}_{s|T}^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_s(\xi_{s-1}^{I_s^\ell}) \vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}) \tilde{r}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}; \tilde{\xi}_{s|T}^\ell)} . \tag{16}$$

Instead of sampling new particles at time s, an algorithm similar to the $\mathsf{TwoFilt}_{bdm}$ algorithm of [1] which uses the forward particles $\{\xi_s^\ell\}_{\ell=1}^N$ or backward particles $\{\check{\xi}_{s|T}^\ell\}_{\ell=1}^N$ may also be implemented with an $\mathcal{O}(N)$ computational complexity.

(a) Choosing $\tilde{r}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}; x_s) = r_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}, x_s)$ in (15), the smoothing distribution approximation is obtained by reweighting the particles obtained in the backward pass. The backward particles $\{\check{\xi}_{s|T}^\ell\}_{\ell=1}^N$ are associated with the importance weights:

$$\begin{aligned}\tilde{\omega}_{s|T}^\ell &:= \frac{\gamma_s(\check{\xi}_{s|T}^\ell) g_s(\check{\xi}_{s|T}^\ell) q(\check{\xi}_{s|T}^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}) r_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}, \check{\xi}_{s|T}^\ell)} \frac{q(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s|T}^\ell)}{\gamma_s(\check{\xi}_{s|T}^\ell)\vartheta_s(\xi_{s-1}^{I_s^\ell})} , \\ &= \check{\omega}_{s|T}^\ell \frac{q(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s|T}^\ell)}{\gamma_s(\check{\xi}_{s|T}^\ell)\vartheta_s(\xi_{s-1}^{I_s^\ell})} .\end{aligned} \tag{17}$$

(b) Choosing $\tilde{r}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}; x_s) = p_s(\xi_{s-1}^{I_s^\ell}, x_s)$ in (15), the smoothing distribution approximation is obtained by reweighting the particles obtained in the forward filtering pass. The forward particles $\{\xi_s^\ell\}_{\ell=1}^N$ are associated with the importance weights:

$$\tilde{\omega}_{s|T}^\ell := \frac{q(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell) g_s(\xi_s^\ell)}{\vartheta_s(\xi_{s-1}^{I_s^\ell}) p_s(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell)} \frac{q(\xi_s^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell})} = \omega_s^\ell \frac{q(\xi_s^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell})} . \tag{18}$$

3 Exponential deviation inequality for the two-filter algorithms

In this section, we establish exponential deviation inequalities for the two-filter algorithms introduced in Section 2. Before stating the results, some additional notations are required. Define, for all $(x, x', x'') \in \mathsf{X}^3$,

$$q^{[2]}(x, x'; x'') = q(x, x'') q(x'', x')$$

and for any functions $f : \mathsf{X}^2 \to \mathbb{R}$ and $g : \mathsf{X} \to \mathbb{R}$,

$$f \odot g(x, x') := f(x, x') g(x') .$$

Consider the following assumptions:

A1. *$|q|_\infty < \infty$ and for all $0 \le t \le T$, g_t is positive and $|g_t|_\infty < \infty$.*

A2. *For all $0 \le t \le T$, $|\vartheta_t|_\infty < \infty$, $|p_t|_\infty < \infty$ and $|\omega_t|_\infty < \infty$ where*

$$\omega_0(x) := \frac{\mathrm{d}\chi}{\mathrm{d}\rho_0}(x) g_0(x) \quad \textit{and for all } t \ge 1 \quad \omega_t(x, x') := \frac{q(x, x') g_t(x')}{\vartheta_t(x) p_t(x, x')} .$$

A3. *- For all $0 \le t \le T-1$, $\left|\vartheta_{t|T}/\gamma_{t+1}\right|_\infty < \infty$ and $|r_{t|T}|_\infty < \infty$. For all $0 \le t \le T$ $\left|\check{\omega}_{t|T}\right|_\infty < \infty$, where*

$$\check{\omega}_{T|T}(x) := \frac{g_T(x)\gamma_T(x)}{\check{\rho}_T(x)} \textit{ and for all } 0 \le t < T,\ \check{\omega}_{t|T}(x, x') := \frac{\gamma_t(x) g_t(x) q(x, x')}{\vartheta_{t|T}(x') r_{t|T}(x', x)} .$$

- For all $1 \le t \le T-1$, $\left|\tilde{\vartheta}_{t|T} \odot \gamma_{t+1}^{-1}\right|_\infty < \infty$, $\left|q \odot \gamma_{t+1}^{-1}\right|_\infty < \infty$, $\left|\tilde{\omega}_{t|T}\right|_\infty < \infty$ and $\left|\tilde{r}_{t|T}\right|_\infty < \infty$ where

$$\tilde{\omega}_{t|T}(x, x'; x'') := \frac{q^{[2]}(x, x'; x'') g_s(x'')}{\tilde{\vartheta}_{t|T}(x, x'') \tilde{r}_{t|T}(x, x'; x'')} .$$

We first show that the weighted sample $\{(\omega_s^i \check{\omega}_{t|T}^j), (\xi_s^i, \check{\xi}_{t|T}^j)\}_{i,j=1}^N$ targets the product distribution $\phi_{\chi,s} \otimes \psi_{\gamma,t|T}$.

Theorem 1. *Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $0 \le s < t \le T$, there exist $0 < B_{s,t|T}, C_{s,t|T} < \infty$ such that for all $N \ge 1$, $\epsilon > 0$ and all $h \in \mathbb{F}_b(\mathsf{X} \times \mathsf{X}, \mathcal{X} \otimes \mathcal{X})$,*

$$\mathbb{P}\left(\left|\sum_{i,j=1}^N \frac{\omega_s^i}{\Omega_s} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]\right| > \epsilon\right) \le B_{s,t|T} \mathrm{e}^{-C_{s,t|T} N \epsilon^2 / \operatorname{osc}^2(h)} .$$

Proof. The proof is postponed to Section 5.1. □

We now study the weighted sample $\{(\tilde{\omega}^i_{s|T}, \tilde{\xi}^\ell_{s|T})\}_{\ell=1}^N$ produced by the $\mathsf{TwoFilt}_{fwt}$ algorithm of Fearnhead, Wyncoll and Tawn [12] defined in (11) and (12) and targeting the marginal smoothing distribution $\phi_{\chi,s|T}$.

Theorem 2 (deviation inequality for $\mathsf{TwoFilt}_{fwt}$ of [12])**.** *Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $s < T$, there exist $0 < B_{s|T}, C_{s|T} < \infty$ such that for all $N \geq 1$, $\varepsilon > 0$ and all $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^N \frac{\tilde{\omega}^i_{s|T}}{\tilde{\Omega}_{s|T}} h(\tilde{\xi}^i_{s|T}) - \phi_{\chi,s|T}[h]\right| > \epsilon\right) \leq B_{s|T} \mathrm{e}^{-C_{s|T} N \epsilon^2 / \operatorname{osc}^2(h)} \ .$$

Proof. The proof is postponed to Section 5.2. □

Using Theorem 1 and Lemma 7, we may derive an exponential inequality for the weighted samples $\{(\xi^i_s, \tilde{\omega}^{i,\mathrm{f}}_{s|T})\}_{i=1}^N$ and $\{(\check{\xi}^i_{s|T}, \tilde{\omega}^{i,\mathrm{b}}_{s|T})\}_{i=1}^N$ produced by the $\mathsf{TwoFilt}_{bdm}$ algorithm of [1], where $\tilde{\omega}^{i,\mathrm{f}}_{s|T}$ and $\tilde{\omega}^{i,\mathrm{b}}_{s|T}$ are defined in (13) and (14). Therefore, both the forward and the backward particle approximations of the smoothing distribution converge to the marginal smoothing distribution, and these two approximations satisfy an exponential inequality.

Theorem 3 (deviation inequality for the $\mathsf{TwoFilt}_{bdm}$ algorithm of [1])**.** *Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $1 \leq s \leq T-1$, there exist $0 < B_{s|T}, C_{s|T} < \infty$ such that for all $N \geq 1$, $\varepsilon > 0$ and all $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^N \frac{\tilde{\omega}^{i,\mathrm{f}}_{s|T}}{\tilde{\Omega}^{\mathrm{f}}_{s|T}} h(\xi^i_s) - \phi_{\chi,s|T}[h]\right| > \epsilon\right) \leq B_{s|T} \mathrm{e}^{-C_{s|T} N \epsilon^2 / \operatorname{osc}^2(h)} \ , \tag{19}$$

$$\mathbb{P}\left(\left|\sum_{i=1}^N \frac{\tilde{\omega}^{i,\mathrm{b}}_{s|T}}{\tilde{\Omega}^{\mathrm{b}}_{s|T}} h(\check{\xi}^i_{s|T}) - \phi_{\chi,s|T}[h]\right| > \epsilon\right) \leq B_{s|T} \mathrm{e}^{-C_{s|T} N \epsilon^2 / \operatorname{osc}^2(h)} \ . \tag{20}$$

Proof. The proof is postponed to Section 5.3. □

Remark 1. Following [23, 6, 11], time uniform exponential inequalities for the two-filter approximations of the marginal smoothing distributions may be obtained using *strong mixing* assumptions which are standard in the SMC literature:

H1. *There exist $0 < \sigma_- < \sigma_+ < \infty$ and $c_- > 0$ such that for all $x, x' \in \mathsf{X}$, $\sigma_- \leq q(x, x') \leq \sigma_+$ and for all $t \geq 0$,*

$$\int \chi(\mathrm{d}x_0) g_0(x_0) \geq c_- \quad \textit{and} \quad \inf_{x \in \mathsf{X}} \int Q(x, \mathrm{d}x') g_t(x') \geq c_- \ .$$

H2. *There exist $0 < \gamma_- < \gamma_+ < \infty$ and $\check{c}_- > 0$ such that for all $x \in \mathsf{X}$ and all $t \geq 0$, $\gamma_- \leq \gamma_t(x) \leq \gamma_+$ and for all $t \geq 0$,*

$$\int \gamma_T(x_T) g_T(x_T) \mathrm{d}x_T \geq \check{c}_- \quad \textit{and} \quad \inf_{x \in \mathsf{X}} \int \gamma_t(x_t) g_t(x_t) q(x_t, x) \gamma^{-1}_{t+1}(x) \mathrm{d}x_t \geq \check{c}_- \ .$$

(i) If A1 and A2 hold uniformly in T and if H1 holds, then, it is proved in [6] that Proposition 8 holds with constants that are uniform in time : there exist $0 < B, C < \infty$ such that for all $s \geq 0$, $N > 0$, $\epsilon > 0$ and all $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,

$$\mathbb{P}\left(\left|\Omega_s^{-1} \sum_{i=1}^{N} \omega_s^i h(\xi_s^i) - \phi_{\chi,s}[h]\right| \geq \epsilon\right) \leq B\mathrm{e}^{-CN\epsilon^2/\operatorname{osc}(h)^2} \ .$$

(ii) It can be shown following the exact same steps that if A1 and A3 hold uniformly in T and if H1 and H2 hold then Proposition 9 holds with constants that are uniform in time: there exist $0 < B, C < \infty$ such that for all $t \geq 0$, $N \geq 1$, $\epsilon > 0$, and all $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,

$$\mathbb{P}\left[\left|\check{\Omega}_{t|T}^{-1} \sum_{i=1}^{N} \check{\omega}_{t|T}^i h(\check{\xi}_{t|T}^i) - \psi_{\gamma,t|T}[h]\right| \geq \epsilon\right] \leq B\mathrm{e}^{-CN\epsilon^2/\operatorname{osc}(h)^2} \ .$$

(iii) Therefore, if A1, A2 and A3 hold uniformly in T and if H1 and H2 hold, then Theorem 1 holds with constants that are uniform in time. As a direct consequence, Theorems 2 and 3 hold also with constants that are uniform in time.

4 Asymptotic normality of the two-filter algorithms

We now establish CLT for the two-filter algorithms. Note first that under assumptions A1, A2 and A3, for all $0 \leq s, t \leq T$ a CLT may be derived for the weighted samples $\{(\xi_s^\ell, \omega_s^\ell)\}_{\ell=1}^N$ and $\{(\check{\xi}_{t|T}^i, \check{\omega}_{t|T}^i)\}_{i=1}^N$ which target respectively the filtering distribution $\phi_{\chi,s}$ and the backward information filter $\psi_{\gamma,t|T}$. By Propositions 10 and 11, there exist $\Gamma_{\chi,s}$ and $\check{\Gamma}_{\gamma,t|T}$ such that for any $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,

$$N^{1/2} \sum_{i=1}^{N} \frac{\omega_s^i}{\Omega_s} \left(h(\xi_s^i) - \phi_{\chi,s}[h]\right) \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \Gamma_{\chi,s}\left[h - \phi_{\chi,s}[h]\right]\right) \ , \tag{21}$$

$$N^{1/2} \sum_{j=1}^{N} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \left(h(\check{\xi}_{t|T}^j) - \psi_{\gamma,t|T}[h]\right) \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \check{\Gamma}_{\gamma,t|T}\left[h - \psi_{\gamma,t|T}[h]\right]\right) \ . \tag{22}$$

Theorem 4 establishes a CLT for the weighted sample $\{\omega_s^i \check{\omega}_{t|T}^j, (\xi_s^i, \check{\xi}_{t|T}^j)\}_{i,j=1}^N$ which targets the product distribution $\phi_{\chi,s} \otimes \psi_{\gamma,t|T}$. As an important consequence, the asymptotic variance of the weighted sample $\{\omega_s^i \check{\omega}_{t|T}^j, (\xi_s^i, \check{\xi}_{t|T}^j)\}_{i,j=1}^N$ is the sum of two contributions, the first one involves $\Gamma_{\chi,s}$ and the second one $\check{\Gamma}_{\gamma,t|T}$. Intuitively, this may be explained by the fact that the estimator $\phi_{\chi,s}^N \otimes \psi_{\gamma,t|T}^N[h]$ is obtained by mixing two independent weighted samples which suggests the following decomposition:

$$\sqrt{N} \sum_{i,j=1}^{N} \frac{\omega_s^i}{\Omega_s} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \tilde{h}_{s,t}(\xi_s^i, \check{\xi}_{t|T}^j) = \sqrt{N} \sum_{j=1}^{N} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \phi_{\chi,s}[\tilde{h}_{s,t}(\cdot, \check{\xi}_{t|T}^j)] + \sqrt{N} \sum_{i=1}^{N} \frac{\omega_s^i}{\Omega_s} \psi_{\gamma,t|T}[\tilde{h}_{s,t}(\xi_s^i, \cdot)] + \mathcal{E}_{s,T|t}^N(\tilde{h}_{s,t}) \ ,$$

where $\tilde{h}_{s,t} = h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]$ and

$$\mathcal{E}_{s,T|t}^N(h) := \sqrt{N} \sum_{i,j=1}^N \frac{\omega_s^i}{\Omega_s} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \left\{ h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s}[h(\cdot, \check{\xi}_{t|T}^j)] - \psi_{\gamma,t|T}[h(\xi_s^i, \cdot)] \right\} .$$

A CLT for the two independent first terms is obtained by (21) and (22). It remains then to prove that $\mathcal{E}_{s,T|t}^N(h)$ converges in probability to 0. However, this cannot be obtained directly from the exponential deviation inequality derived in Theorem 1 and requires sharper controls of the smoothing error (for instance nonasymptotic L^p-mean error bounds). Theorem 4 provides a direct proof following the asymptotic theory of weighted system of particles developed in [7].

Theorem 4. *Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $0 \le s < t \le T$ and all $h \in \mathbb{F}_b(\mathsf{X} \times \mathsf{X}, \mathcal{X} \times \mathcal{X})$,*

$$\sqrt{N} \left(\sum_{i,j=1}^N \frac{\omega_s^i}{\Omega_s} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h] \right) \xrightarrow{\mathcal{D}}_{N \to \infty} \mathcal{N}\left(0, \tilde{\Gamma}_{s,t|T}\left[h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]\right]\right) ,$$

where $\tilde{\Gamma}_{s,t|T}[h]$ is defined by:

$$\tilde{\Gamma}_{s,t|T}[h] := \Gamma_{\chi,s}\left[\int \psi_{\gamma,t|T}(\mathrm{d}x_t) h(\cdot, x_t)\right] + \check{\Gamma}_{\gamma,t|T}\left[\int \phi_{\chi,s}(\mathrm{d}x_s) h(x_s, \cdot)\right] , \tag{23}$$

with $\Gamma_{\chi,s}$ and $\check{\Gamma}_{\gamma,t|T}$ are given in Proposition 10 and Proposition 11.

Proof. The proof is postponed to Section 5.4. □

Define

$$\sigma_s := \phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T}\left[\int q^{[2]}(\cdot, x) g_s(x) \mathrm{d}x \odot \gamma_{s+1}^{-1}\right] ,$$

$$\Sigma_s[h] := \tilde{\Gamma}_{s-1,s+1|T}\left[\int q^{[2]}(\cdot; x) g_s(x) h(x) \mathrm{d}x \odot \gamma_{s+1}^{-1}\right] .$$

Theorem 5 provides a CLT for the $\mathsf{TwoFilt}_{fwt}$ algorithm of [12]

Theorem 5 (CLT for the $\mathsf{TwoFilt}_{fwt}$ algorithm of [12])**.** *Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $1 \le s \le T-1$ and all $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,*

$$\sqrt{N}\left(\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^i}{\tilde{\Omega}_{s|T}} h(\tilde{\xi}_{s|T}^i) - \phi_{\chi,s|T}[h]\right) \xrightarrow{\mathcal{D}}_{N \to \infty} \mathcal{N}\left(0, \Upsilon_{\chi,s|T}\left[h - \phi_{\chi,s|T}[h]\right]\right) .$$

where

$$\Upsilon_{\chi,s|T}[h] = \sigma_s^{-2} \left\{ \Sigma_s[h] + \phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T}\left[\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}\right] \right.$$
$$\left. \times\ \phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T}\left[\int \tilde{\omega}_{s|T}(\cdot; x) q^{[2]}(\cdot, x) g_s(x) h^2(x) \mathrm{d}x \odot \gamma_{s+1}^{-1}\right] \right\} . \tag{24}$$

Proof. The proof is postponed to Section 5.5. □

The decompositions (27) and (28) together with Theorem 4 allow to prove a CLT form the forward and the backward approximations of the marginal smoothing distribution. Theorem 6 is a direct consequence of Proposition 11, Theorem 4 and Slutsky Lemma.

Theorem 6 (CLT for the $\mathsf{TwoFilt}_{bdm}$ algorithm of [1])**.** *Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $1 \leq s \leq T-1$ and all $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,*

$$\sqrt{N}\left(\sum_{i=1}^{N} \frac{\tilde{\omega}_{s|T}^{i,\mathrm{f}}}{\tilde{\Omega}_{s|T}^{\mathrm{f}}} h(\xi_s^i) - \phi_{\chi,s|T}[h]\right) \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \Delta_{\chi,s|T}^{\mathrm{f}}\left[h - \phi_{\chi,s|T}[h]\right]\right) ,$$

where

$$\Delta_{\chi,s|T}^{\mathrm{f}}[h] := \tilde{\Gamma}_{s,s+1|T}\left[H_s^{\mathrm{f}}\right] / \{\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}[q \odot \gamma_{s+1}^{-1}]\}^2 ,$$
$$H_s^{\mathrm{f}}(x,x') := h(x) q(x,x') \gamma_{s+1}^{-1}(x') .$$

Similarly,

$$\sqrt{N}\left(\sum_{i=1}^{N} \frac{\tilde{\omega}_{s|T}^{i,\mathrm{b}}}{\tilde{\Omega}_{s|T}^{\mathrm{b}}} h(\xi_s^i) - \phi_{\chi,s|T}[h]\right) \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \Delta_{\chi,s|T}^{\mathrm{b}}\left[h - \phi_{\chi,s|T}[h]\right]\right) ,$$

where

$$\Delta_{\chi,s|T}^{\mathrm{b}}[h] := \tilde{\Gamma}_{s-1,s|T}\left[H_s^{\mathrm{b}}\right] / \{\phi_{\chi,s-1} \otimes \psi_{\gamma,s|T}[q \odot \gamma_{s}^{-1}]\}^2 ,$$
$$H_s^{\mathrm{b}}(x,x') := q(x,x') \gamma_{s}^{-1}(x') h(x') .$$

Note that σ_s and $\Sigma_s[h]$ may be written as:

$$\sigma_s = \phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}\left[q \odot \gamma_{s+1}^{-1}\right] \times \phi_{\chi,s-1}\left[\int q(\cdot,x) g_s(x) \mathrm{d}x\right]$$

and by Theorem 4,

$$\Sigma_s[h] = \Gamma_{\chi,s-1}\left[\int q(\cdot,x) g_s(x) h_{s+1}^1(x) \mathrm{d}x\right] + \phi_{\chi,s-1}^2\left[\int q(\cdot,x) g_s(x) \mathrm{d}x\right] \check{\Gamma}_{\gamma,s+1|T}\left[h_{s+1}^2\right] ,$$

with $h_{s+1}^1(x) := h(x)\psi_{\gamma,s+1|T}[q(x,\cdot)\gamma_{s+1}^{-1}]$ and $h_{s+1}^2(x) := \gamma_{s+1}^{-1}(x)\phi_{\chi,s}[h(\cdot)q(\cdot,x)]$. In the case where $\tilde{r}_{s|T}(x_s, x_{s+1}; x_s) = p_s(x_{s-1}, x_s)$ in (15) and $\tilde{\vartheta}_{s|T}(x,x') = \vartheta_s(x)\vartheta_{s|T}(x')$, the smoothing distribution approximation given by the $\mathsf{TwoFilt}_{fwt}$ algorithm is obtained by reweighting the particles obtained in the forward filtering pass and $\Upsilon_{\chi,s|T}[h]$ may be compared to $\Delta_{\chi,s|T}^{\mathrm{f}}[h]$ as both approximations of $\phi_{\chi,s|T}[h]$ are based on the same particles (associated with different importance weights). In this case, the two last terms in (24) are easily interpreted in the case $\vartheta_{s|T} = \gamma_{s+1}$:

$$\phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T}\left[\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}\right] = \phi_{\chi,s-1}[\vartheta_s]\psi_{\gamma,s+1|T}[\vartheta_{s|T}\gamma_{s+1}^{-1}] = \phi_{\chi,s-1}[\vartheta_s]$$

and by Jensen's inequality,

$$
\begin{aligned}
&\phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T}\left[\int \tilde{\omega}_{s|T}(\cdot;x)q^{[2]}(\cdot,x)g_s(x)h^2(x)\mathrm{d}x \odot \gamma_{s+1}^{-1}\right] \\
&\quad = \int \phi_{\chi,s-1}(\mathrm{d}x_{s-1})\omega_s(x_{s-1},x)g_s(x)q(x_{s-1},x)\psi_{\gamma,s+1|T}[q^2(x,\cdot)\gamma_{s+1}^{-2}]h^2(x)\mathrm{d}x\ , \\
&\quad \geq \int \phi_{\chi,s-1}(\mathrm{d}x_{s-1})\omega_s(x_{s-1},x)g_s(x)q(x_{s-1},x)(h_{s+1}^1(x))^2\mathrm{d}x\ .
\end{aligned}
$$

Therefore, by Proposition 11 and Theorem 6

$$
\Upsilon_{\chi,s|T}\left[h\right] \geq \frac{\Gamma_{\chi,s}\left[h_{s+1}^1\right] + \check{\Gamma}_{\gamma,s+1|T}\left[h_{s+1}^2\right]}{\left(\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}\left[q \odot \gamma_{s+1}^{-1}\right]\right)^2} = \Delta_{\chi,s|T}^{\mathrm{f}}\left[h\right]\ ,
$$

where the last inequality comes from Theorem 4. The same inequality holds for $\Delta_{\chi,s|T}^{\mathrm{b}}\left[h\right]$ when $\tilde{r}_{s|T}(x_{s-1},x_{s+1};x_s) = r_{s|T}(x_{s+1},x_s)$ in (15).

Remark 2. Under the strong mixing assumptions H1 and H2, time uniform bounds for the asymptotic variances of the two-filter approximations of the marginal smoothing distributions may be obtained.

(i) If A1 and A2 hold uniformly in T and if H1 holds, then it is proved in [6] that there exists $C > 0$ such that for all $s \geq 0$ and all $h \in \mathbb{F}_b(\mathsf{X},\mathcal{X})$, the asymptotic variance $\Gamma_{\chi,s}\left[h\right]$ defined in Proposition 10 satisfies:

$$
\Gamma_{\chi,s}\left[h\right] \leq C\left|h\right|_\infty^2\ .
$$

(ii) Following the same steps, if A1 and A3 hold uniformly in T and if H1 and H2 hold, there exists $C > 0$ such that for all $0 \leq t \leq T$ and all $h \in \mathbb{F}_b(\mathsf{X},\mathcal{X})$, the asymptotic variance $\check{\Gamma}_{t|T}\left[h\right]$ defined in Proposition 11 satisfies:

$$
\check{\Gamma}_{\gamma,t|T}\left[h\right] \leq C\left|h\right|_\infty^2\ .
$$

(iii) As a consequence, if A1, A2 and A3 hold uniformly in T and if H1 and H2 hold, the asymptotic variances $\tilde{\Gamma}_{s,t|T}\left[h\right]$, $\Delta_{\chi,s|T}^{\mathrm{f}}\left[h\right]$, $\Delta_{\chi,s|T}^{\mathrm{b}}\left[h\right]$ and $\Upsilon_{\chi,s|T}\left[h\right]$ defined in Theorem 4, Theorem 5 and Theorem 6 are all uniformly bounded.

5 Proofs

5.1 Proof of Theorem 1

Define $\mathcal{G}_{t|T}^N := \sigma(\check{\xi}_{t|T}^j, \check{\omega}_{t|T}^j, 1 \leq j \leq N)$ and

$$
f_{t|T}(x) := \check{\Omega}_{t|T}^{-1}\sum_{j=1}^N \check{\omega}_{t|T}^j h(x,\check{\xi}_{t|T}^j)
$$

whose oscillation is bounded by $\operatorname{osc}(h)$. By the exponential inequality for the auxiliary particle filter (Proposition 8), there exist constants B_s and C_s such that

$$\mathbb{P}\left(\left|\sum_{i,j=1}^{N} \frac{\omega_s^i}{\Omega_s} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \sum_{j=1}^{N} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \int \phi_{\chi,s}(\mathrm{d}x_s) h(x_s, \check{\xi}_{t|T}^j)\right| > \epsilon\right)$$
$$= \mathbb{E}\left[\mathbb{P}\left(\left|\sum_{i=1}^{N} \frac{\omega_s^i}{\Omega_s} f_{t|T}(\xi_s^i) - \phi_{\chi,s}(f_{t|T})\right| > \epsilon \,\middle|\, \mathcal{G}_{t|T}^N\right)\right] \le B_s \mathrm{e}^{-C_s N \epsilon^2 / \operatorname{osc}^2(h)} . \quad (25)$$

Since the oscillation of the function $x \mapsto \int \phi_{\chi,s}(\mathrm{d}x_s) h(x_s, x)$ is bounded by $\operatorname{osc}(h)$, by Proposition 9 there exist constants $B_{t|T}$ and $C_{t|T}$ such that

$$\mathbb{P}\left(\left|\sum_{j=1}^{N} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \int \phi_{\chi,s}(\mathrm{d}x_s) h(x_s, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]\right| > \epsilon\right)$$
$$\le B_{t|T} \mathrm{e}^{-C_{t|T} N \epsilon^2 / \operatorname{osc}^2(h)} , \quad (26)$$

which concludes the proof.

5.2 Proof of Theorem 2

Define $\tilde{h}_{s|T} := h - \phi_{\chi,s|T}[h]$. Lemma 7 is used with

$$a_N := N^{-1} \sum_{i=1}^{N} \tilde{\omega}_{s|T}^i \tilde{h}_{s|T}(\tilde{\xi}_{s|T}^i) , \quad b_N := N^{-1} \tilde{\Omega}_{s|T} ,$$
$$b := \frac{\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}\left[\int q^{[2]}(\cdot; x_s) g_s(x_s) \mathrm{d}x_s \odot \gamma_{s+1}^{-1}\right]}{\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}\left[\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}\right]} .$$

Lemma 7-(i) is satisfied using $\beta := b$ and $|a_N|/|b_N| \le \operatorname{osc}(h)$. To prove Lemma 7-(ii) for a_N, note that Hoeffding inequality implies that, for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|a_N - \mathbb{E}\left[\tilde{\omega}_{s|T}^1 \tilde{h}_{s|T}(\tilde{\xi}_{s|T}^1) \,\middle|\, \mathcal{G}_{s,T}^N\right]\right| \ge \epsilon \,\middle|\, \mathcal{G}_{s,T}^N\right) \le 2 \exp\left\{-\frac{N\epsilon^2}{8 \left|\tilde{\omega}_{s|T}\right|_\infty^2 \operatorname{osc}^2(h)}\right\} ,$$

where $\mathcal{G}_{s,T}^N := \mathcal{G}_{s-1}^{N,+} \vee \mathcal{G}_{s+1,T}^{N,-}$ and

$$\mathcal{G}_s^{N,+} := \sigma\left\{\{(\omega_u^i, \xi_u^i)\}_{i=1}^N, u = 1, \ldots, s-1\right\} ,$$
$$\mathcal{G}_{s,T}^{N,-} := \sigma\left\{\{(\check{\omega}_{u|T}^i, \check{\xi}_{u|T}^i)\}_{i=1}^N, u = s+1, \ldots, T\right\} .$$

On the other hand, for all $\ell \in \{1, \ldots, N\}$,

$$\mathbb{E}\left[\tilde{\omega}_{s|T}^\ell \tilde{h}_{s|T}(\tilde{\xi}_{s|T}^\ell) \,\middle|\, \mathcal{G}_{s,T}^N\right]$$
$$= \frac{\sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s+1|T}^j \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \int q^{[2]}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x_s) g_s(x_s) \tilde{h}_{s|T}(x_s) \mathrm{d}x_s}{\sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s+1|T}^j \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \tilde{\vartheta}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j)} .$$

The proof of Lemma 7-(ii) is then completed by applying Lemma 7 to a'_N, b'_N and b' defined by:

$$
\begin{aligned}
a'_N &:= \sum_{i,j=1}^{N} \frac{\omega_{s-1}^i \, \check{\omega}_{s+1|T}^j}{\Omega_{s-1} \, \check{\Omega}_{s+1|T}} \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \int q^{[2]}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x_s) g_s(x_s) \tilde{h}_{s|T}(x_s) \mathrm{d}x_s \;, \\
b'_N &:= \sum_{i,j=1}^{N} \frac{\omega_{s-1}^i \, \check{\omega}_{s+1|T}^j}{\Omega_{s-1} \, \check{\Omega}_{s+1|T}} \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \tilde{\vartheta}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j) \;, \\
b' &:= \phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}[\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}] \;.
\end{aligned}
$$

Note first that Lemma 7-(i) is satisfied using $\beta' := b'$ and $|a'_N/b'_N| \leq \left|\tilde{\omega}_{s|T}\right|_\infty \operatorname{osc}(h)$. In addition, by (6),

$$
\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}[\bar{h}_{s|T}] \propto \phi_{\chi,s|T}[\bar{h}_{s|T}] = 0 \;,
$$

where

$$
\bar{h}_{s|T}(x, x') := \int q^{[2]}(\cdot; x_s) g_s(x_s) \tilde{h}_{s|T}(x_s) \mathrm{d}x_s \odot \gamma_{s+1}^{-1}(x, x') \;.
$$

Theorem 1 ensures that Lemma 7-(ii) is satisfied for a'_N as

$$
\operatorname{osc}\left(\bar{h}_{s|T}\right) \leq 2 \left|\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}\right|_\infty \left|\tilde{\omega}_{s|T}\right|_\infty \operatorname{osc}(h) \;.
$$

Similarly, Theorem 1 yields:

$$
\mathbb{P}\left(|b'_N - b'| \geq \epsilon\right) \leq B_s \mathrm{e}^{-C_s N \epsilon^2 / \operatorname{osc}^2\left(\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}\right)} \;,
$$

which proves Lemma 7-(iii) for b'_N and concludes the proof of Lemma 7-(ii) for a_N. The proof of Lemma 7-(iii) for b_N is along the same lines.

5.3 Proof of Theorem 3

Define

$$
\overline{h}_s(x, x') := \gamma_{s+1}^{-1}(x') h(x) q(x, x') \quad \text{and} \quad \underline{h}_s(x, x') := \gamma_s^{-1}(x') q(x, x') h(x') \;.
$$

It follows from the definition of the forward and backward smoothing weights (13) and (14) that,

$$
\sum_{i=1}^{N} \frac{\tilde{\omega}_{s|T}^{i,\mathrm{f}}}{\tilde{\Omega}_{s|T}^{\mathrm{f}}} h(\xi_s^i) = \frac{\Omega_s^{-1} \check{\Omega}_{s+1|T}^{-1} \sum_{i,j=1}^{N} \omega_s^i \check{\omega}_{s+1|T}^j \overline{h}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)}{\Omega_s^{-1} \check{\Omega}_{s+1|T}^{-1} \sum_{i,j=1}^{N} \omega_s^i \check{\omega}_{s+1|T}^j \overline{\mathbf{1}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)} \;, \tag{27}
$$

$$
\sum_{i=1}^{N} \frac{\tilde{\omega}_{s|T}^{i,\mathrm{b}}}{\tilde{\Omega}_{s|T}^{\mathrm{b}}} h(\check{\xi}_{s|T}^i) = \frac{\Omega_{s-1}^{-1} \check{\Omega}_{s|T}^{-1} \sum_{i,j=1}^{N} \omega_{s-1}^i \check{\omega}_{s|T}^j \underline{h}_s(\xi_{s-1}^i, \check{\xi}_{s|T}^j)}{\Omega_{s-1}^{-1} \check{\Omega}_{s|T}^{-1} \sum_{i,j=1}^{N} \omega_{s-1}^i \check{\omega}_{s|T}^j \underline{\mathbf{1}}_s(\xi_{s-1}^i, \check{\xi}_{s|T}^j)} \;. \tag{28}
$$

On the other hand, from the definition of the filtering distribution and of the backward information filter

$$
\begin{aligned}
\phi_{\chi,s|T}[h] &= \phi_{\chi,s} \otimes \psi_{s+1|T}\left[\overline{h}_s\right] / \phi_{\chi,s} \otimes \psi_{s+1|T}\left[\overline{\mathbf{1}}_s\right] \;, \\
\phi_{\chi,s|T}[h] &= \phi_{\chi,s-1} \otimes \psi_{s|T}\left[\underline{h}_s\right] / \phi_{\chi,s-1} \otimes \psi_{s|T}\left[\underline{\mathbf{1}}_s\right] \;.
\end{aligned}
$$

Then, (19) is established by writing:

$$\sum_{i=1}^{N} \frac{\tilde{\omega}_{s|T}^{i,\mathrm{f}}}{\tilde{\Omega}_{s|T}^{\mathrm{f}}} h(\xi_s^i) - \phi_{\chi,s|T}[h] = a_N^{i,\mathrm{f}} / b_N^{i,\mathrm{f}} \;,$$

where

$$\begin{aligned} a_N^{i,\mathrm{f}} &:= \sum_{i,j=1}^{N} \frac{\omega_s^i \check{\omega}_{s+1|T}^j}{\Omega_s \check{\Omega}_{s+1|T}} \overline{\mathbf{1}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j) \left\{ \frac{\overline{h}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)}{\overline{\mathbf{1}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)} - \frac{\phi_{\chi,s} \otimes \psi_{s+1|T}\left[\overline{h}_s\right]}{\phi_{\chi,s} \otimes \psi_{s+1|T}\left[\overline{\mathbf{1}}_s\right]} \right\} \;, \\ b_N^{i,\mathrm{f}} &:= \sum_{i,j=1}^{N} \frac{\omega_s^i \check{\omega}_{s+1|T}^j}{\Omega_s \check{\Omega}_{s+1|T}} \overline{\mathbf{1}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j) \;, \quad b := \phi_{\chi,s} \otimes \psi_{s+1|T}\left[\overline{\mathbf{1}}_s\right] \;. \end{aligned}$$

Lemma 7 may then be applied with $\beta := b$. Note that

$$\frac{\overline{h}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)}{\overline{\mathbf{1}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)} - \frac{\phi_{\chi,s} \otimes \psi_{s+1|T}\left[\overline{h}_s\right]}{\phi_{\chi,s} \otimes \psi_{s+1|T}\left[\overline{\mathbf{1}}_s\right]} = h(\xi_s^i) - \phi_{\chi,s|T}[h] \;,$$

which ensures that $\left| a_N^{i,\mathrm{f}} / b_N^{i,\mathrm{f}} \right| \le \operatorname{osc}(h)$ and that Lemma 7-(i) is satisfied. By

$$\begin{aligned} &\operatorname{osc}\left(\overline{\mathbf{1}}_s\right) = \operatorname{osc}\left(q \odot \gamma_{s+1}^{-1}\right) \;, \\ &\operatorname{osc}\left(\overline{\mathbf{1}}_s \odot \left\{h(\xi_s^i) - \phi_{\chi,s|T}[h]\right\}\right) \le 2 \left| q \odot \gamma_{s+1}^{-1} \right|_\infty \operatorname{osc}(h) \;, \end{aligned}$$

Theorem 1 shows that Lemma 7-(ii) and (iii) are satisfied. The proof of (20) follows the exact same lines.

5.4 Proof of Theorem 4

For all $1 \le t \le T$, the result is shown by induction on s where $s \in \{0, \dots, t-1\}$. Write $\tilde{h}_{0,t} := h - \phi_{\chi,0} \otimes \psi_{\gamma,t|T}[h]$ and set, for $i \in \{1, \dots, N\}$,

$$U_{N,i} := N^{-1/2} \omega_0^i \sum_{j=1}^{N} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \tilde{h}_{0,t}(\xi_0^i, \check{\xi}_{t|T}^j) \;.$$

Then,

$$\sqrt{N} \left(\sum_{i,j=1}^{N} \frac{\omega_0^i}{\Omega_0} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\xi_0^i, \check{\xi}_{t|T}^j) - \phi_{\chi,0} \otimes \psi_{\gamma,t|T}[h] \right) = (\Omega_0/N)^{-1} \sum_{i=1}^{N} U_{N,i} \;.$$

Define $\mathcal{G}_{N,i} := \sigma\left(\{\xi_0^\ell\}_{\ell \le i}, \{\check{\xi}_{u|T}^j\}_{t \le u \le T}, j = 1, \dots, N \right)$. Then,

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i} \,|\, \mathcal{G}_{N,i-1} \right] = N^{1/2} \sum_{j=1}^{N} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \rho_0 \left[\omega_0 \tilde{h}_{0,t}(\cdot, \check{\xi}_{t|T}^j) \right] \;.$$

As $\int \psi_{\gamma,t|T}(\mathrm{d}x_t)\rho_0(\mathrm{d}x_0)\omega_0(x_0)\tilde{h}_{0,t}(x_0,x_t) = 0$, by the CLT for the backward information filter (Proposition 11),

$$\sum_{i=1}^N \mathbb{E}\left[U_{N,i}\,\middle|\,\mathcal{G}_{N,i-1}\right] \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \check{\Gamma}_{t|T}\left[H_{0,t}\right]\right) ,$$

where $H_{0,t}(x_t) := \int \rho_0(\mathrm{d}x_0)\omega_0(x_0)\tilde{h}_{0,t}(x_0,x_t)$. We now prove that

$$\mathbb{E}\left[\exp\left(\mathrm{i}u\sum_{i=1}^N\{U_{N,i} - \mathbb{E}\left[U_{N,i}\,\middle|\,\mathcal{G}_{N,i-1}\right]\}\right)\middle|\,\mathcal{G}_{N,0}\right] \xrightarrow{\mathrm{P}}_{N\to\infty} \exp\left(-\frac{u^2\sigma^2_{0,t|T}[h]}{2}\right) ,$$

where

$$\sigma^2_{0,t|T}[h] := \int \rho_0(\mathrm{d}x)\omega_0^2(x)\psi^2_{\gamma,t|T}[\tilde{h}_{0,t}(x,\cdot)] .$$

This is done by applying [7, Theorem A.3] which requires to show that

$$\sum_{i=1}^N \left(\mathbb{E}\left[U^2_{N,i}\,\middle|\,\mathcal{G}_{N,i-1}\right] - \mathbb{E}\left[U_{N,i}\,\middle|\,\mathcal{G}_{N,i-1}\right]^2\right) \xrightarrow{\mathrm{P}}_{N\to\infty} \sigma^2_{0,t|T}[h] , \tag{29}$$

$$\sum_{i=1}^N \mathbb{E}\left[U^2_{N,i}\mathbf{1}\{|U_{N,i}| > \varepsilon\}\,\middle|\,\mathcal{G}_{N,i-1}\right] \xrightarrow{\mathrm{P}}_{N\to\infty} 0 . \tag{30}$$

By Proposition 9,

$$\sum_{i=1}^N \mathbb{E}\left[U_{N,i}\,\middle|\,\mathcal{G}_{N,i-1}\right]^2 = \left(\sum_{j=1}^N \frac{\check{\omega}^j_{t|T}}{\check{\Omega}_{t|T}} H_{0,t}(\check{\xi}^j_{t|T})\right)^2 \xrightarrow{\mathrm{P}}_{N\to\infty} \psi^2_{\gamma,t|T}[H_{0,t}] = 0 .$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{i=1}^N \mathbb{E}\left[U^2_{N,i}\,\middle|\,\mathcal{G}_{N,i-1}\right] - \sigma^2_{0,t|T}[h]\right|\right] \\
&\quad = \int \rho_0(\mathrm{d}x)\omega_0^2(x)\mathbb{E}\left[\left|\left(\sum_{j=1}^N \frac{\check{\omega}^j_{t|T}}{\check{\Omega}_{t|T}}\tilde{h}_{0,t}(x,\check{\xi}^j_{t|T})\right)^2 - \psi^2_{\gamma,t|T}[\tilde{h}_{0,t}(x,\cdot)]\right|\right] , \\
&\quad \le 2\operatorname{osc}(h)\int \rho_0(\mathrm{d}x)\omega_0^2(x)\mathbb{E}\left[A_N(x)\right] ,
\end{aligned}$$

where

$$A_N(x) := \left|\sum_{j=1}^N \frac{\check{\omega}^j_{t|T}}{\check{\Omega}_{t|T}}\tilde{h}_{0,t}(x,\check{\xi}^j_{t|T}) - \psi_{\gamma,t|T}[\tilde{h}_{0,t}(x,\cdot)]\right| .$$

By Proposition 9, there exist $B_{t|T}$ and $C_{t|T}$ such that for all $x \in \mathsf{X}$,

$$\begin{aligned}
\mathbb{E}\left[A_N(x)\right] &= \int_0^\infty \mathbb{P}\left(A_N(x) \ge \varepsilon\right)\mathrm{d}\varepsilon \\
&\le B_{t|T}\int_0^\infty \mathrm{e}^{-C_{t|T}N\epsilon^2/\operatorname{osc}(h)^2}\mathrm{d}\varepsilon \le D_{t|T}\operatorname{osc}(h)N^{-1/2} ,
\end{aligned} \tag{31}$$

which shows that

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i}^2 \,\middle|\, \mathcal{G}_{N,i-1}\right] \xrightarrow{\mathrm{P}}_{N\to\infty} \sigma^2_{0,t|T}[h]$$

and concludes the proof of (29). For all $N \geq 1$,

$$\{|U_{N,i}| \geq \varepsilon\} \subseteq \left\{\omega_0^i \geq \varepsilon N^{1/2} \operatorname{osc}(h)^{-1}\right\} ,$$

which implies that

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i}^2 \mathbf{1}\{|U_{N,i}| \geq \varepsilon\} \,\middle|\, \mathcal{G}_{N,i-1}\right] \leq \operatorname{osc}(h)^2 \int \rho_0(\mathrm{d}x)\omega_0^2(x)\mathbf{1}\left\{\omega_0(x) \geq N^{1/2}\operatorname{osc}(h)^{-1}\right\}$$

and (30) follows by letting $N \to \infty$. Note that

$$N^{-1}\Omega_0 \xrightarrow{\mathrm{P}}_{N\to\infty} \int \chi(\mathrm{d}x_0)g_0(x_0) ,$$

which shows (23) since

$$\begin{aligned}
\tilde{\Gamma}_{0,t|T}\left[\tilde{h}_{0,t}\right] &= \left(\int \chi(\mathrm{d}x_0)g_0(x_0)\right)^{-2} \\
&\qquad \times \left(\check{\Gamma}_{\gamma,t|T}\left[H_{0,t}\right] + \int \rho_0(\mathrm{d}x)\omega_0^2(x)\psi^2_{\gamma,t|T}[\tilde{h}_{0,t}(x,\cdot)]\right) , \\
&= \check{\Gamma}_{\gamma,t|T}\left[\int \phi_{\chi,0}(\mathrm{d}x_0)\tilde{h}_{0,t}(x_0,\cdot)\right] + \Gamma_{\chi,0}\left[\int \psi_{\gamma,t|T}(\mathrm{d}x_t)\tilde{h}_{0,t}(\cdot,x_t)\right] .
\end{aligned}$$

Assume now that the result holds for some $s-1$. Write $\tilde{h}_{s,t} := h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]$ and set, for $i \in \{1,\ldots,N\}$,

$$U_{N,i} := N^{-1/2}\omega_s^i \sum_{j=1}^{N} \frac{\breve{\omega}^j_{t|T}}{\check{\Omega}_{t|T}} \tilde{h}_{s,t}(\xi_s^i, \check{\xi}^j_{t|T}) .$$

Then,

$$\sqrt{N}\left(\sum_{i,j=1}^{N} \frac{\omega_s^i}{\Omega_s}\frac{\breve{\omega}^j_{t|T}}{\check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}^j_{t|T}) - \phi_{\chi,0} \otimes \psi_{\gamma,t|T}[h]\right) = (\Omega_s/N)^{-1}\sum_{i=1}^{N} U_{N,i} .$$

Define, for $1 \leq i \leq N$,

$$\mathcal{G}_{N,i} := \sigma\left(\left\{\xi_s^j\right\}_{j=1}^{i}, \left\{\xi_u^\ell\right\}_{\ell=1}^{N}, \left\{\check{\xi}^j_{v|T}\right\}_{j=1}^{N}, 1 \leq u < s, t \leq v \leq T\right) .$$

Then,

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i} \,\middle|\, \mathcal{G}_{N,i-1}\right] = \left(\phi^N_{\chi,s-1}[\vartheta_s]\right)^{-1} N^{1/2} \sum_{i,j=1}^{N} \frac{\omega^i_{s-1}}{\Omega_{s-1}}\frac{\breve{\omega}^j_{t|T}}{\check{\Omega}_{t|T}} H_s(\xi^i_{s-1}, \check{\xi}^j_{t|T}) ,$$

where

$$H_{s,t}(x_{s-1}, x_t) := \int q(x_{s-1}, x)g_s(x)\tilde{h}_{s,t}(x, x_t)\mathrm{d}x . \tag{32}$$

Since $\phi_{\chi,s-1} \otimes \psi_{\gamma,t|T}[H_{s,t}] = 0$, by the induction assumption,

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i} \,\middle|\, \mathcal{G}_{N,i-1}\right] \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \tilde{\Gamma}_{s-1,t|T}\left[H_{s,t}\right] / \phi^2_{\chi,s-1}[\vartheta_s]\right) .$$

We will now prove that

$$\mathbb{E}\left[\exp\left(\mathrm{i}u \sum_{i=1}^{N} \{U_{N,i} - \mathbb{E}\left[U_{N,i} \,\middle|\, \mathcal{G}_{N,i-1}\right]\}\right) \middle|\, \mathcal{G}_{N,0}\right] \xrightarrow{\mathrm{P}}_{N\to\infty} \exp\left(-\frac{u^2 \sigma^2_{s,t|T}[h]}{2}\right) ,$$

where

$$\sigma^2_{s,t|T}[h] := \phi_{\chi,s-1}[\vartheta_s]^{-1} \phi_{\chi,s-1}\left[f_{s-1,t}\right] ,$$
$$f_{s-1,t}(x_{s-1}) := \int q(x_{s-1}, x_s) \omega_s(x_{s-1}, x_s) \psi^2_{\gamma,t|T}[\tilde{h}_{s,t}(x_s, \cdot)] g_s(x_s) \mathrm{d}x_s .$$

This is done using again [7, Theorem A.3] and proving that (29) and (30) hold with $\sigma^2_{0,t|T}[h]$ replaced by $\sigma^2_{s,t|T}[h]$. Note that

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i} \,\middle|\, \mathcal{G}_{N,i-1}\right]^2 = \left(\sum_{i,j=1}^{N} \frac{\omega^i_{s-1}}{\Omega_{s-1}} \frac{\breve{\omega}^j_{t|T}}{\breve{\Omega}_{t|T}} H_{s,t}(\xi^i_{s-1}, \breve{\xi}^j_{t|T})\right)^2 / (\phi^N_{\chi,s-1}[\vartheta_s])^2 ,$$

which converges in probability to 0 by Theorem 1 and the fact that $\phi_{\chi,s-1} \otimes \psi_{\gamma,t|T}[H_{s,t}] = 0$. In addition,

$$\begin{aligned}
\sum_{i=1}^{N} \mathbb{E}\left[U^2_{N,i} \,\middle|\, \mathcal{G}_{N,i-1}\right] &= \left(\phi^N_{\chi,s-1}[\vartheta_s]\right)^{-1} \sum_{i=1}^{N} \frac{\omega^i_{s-1}}{\Omega_{s-1}} \int \frac{q^2(\xi^i_{s-1}, x_s) g_s(x_s)}{\vartheta_s(\xi^i_{s-1}) p_s(\xi^i_{s-1}, x_s)} \\
&\qquad \times \left(\psi^N_{\gamma,t|T}[\tilde{h}_{s,t}(x_s, \cdot)]\right)^2 g_s(x_s) \mathrm{d}x_s , \\
&= \left(\phi^N_{\chi,s-1}[\vartheta_s]\right)^{-1} \sum_{i=1}^{N} \frac{\omega^i_{s-1}}{\Omega_{s-1}} \int \omega_s(\xi^i_{s-1}, x_s) q(\xi^i_{s-1}, x_s) \\
&\qquad \times \left(\psi^N_{\gamma,t|T}[\tilde{h}_{s,t}(x_s, \cdot)]\right)^2 g_s(x_s) \mathrm{d}x_s , \\
&= \left(\phi^N_{\chi,s-1}[\vartheta_s]\right)^{-1} \phi^N_{\chi,s-1}\left[f^N_{s-1,t}\right] ,
\end{aligned}$$

where

$$f^N_{s-1,t}(x_{s-1}) := \int q(x_{s-1}, x_s) \omega_s(x_{s-1}, x_s) \left(\psi^N_{\gamma,t|T}[\tilde{h}_{s,t}(x_s, \cdot)]\right)^2 g_s(x_s) \mathrm{d}x_s .$$

First note that $\phi^N_{\chi,s-1}[\vartheta_s] \xrightarrow{\mathrm{P}}_{N\to\infty} \phi_{\chi,s-1}[\vartheta_s]$ and write

$$\left|\phi^N_{\chi,s-1}\left[f^N_{s-1,t}\right] - \phi_{\chi,s-1}\left[f_{s-1,t}\right]\right| \le A^N_{s,t} + B^N_{s,t} ,$$

where $A^N_{s,t} := |\phi^N_{\chi,s-1}[f^N_{s-1,t}] - \phi^N_{\chi,s-1}[f_{s-1,t}]|$ and $B^N_{s,t} := |\phi^N_{\chi,s-1}[f_{s-1,t}] - \phi_{\chi,s-1}[f_{s-1,t}]|$. As $(\omega^i_{s-1}, \xi^i_{s-1})_{i=1}^N$ and $(\breve{\omega}^j_{t|T}, \breve{\xi}^j_{t|T})_{j=1}^N$ are independent,

$$\mathbb{E}\left[A^N_{s,t}\right] \le |\omega_s|_\infty |g_s|_\infty \mathbb{E}\left[\sum_{i=1}^{N} \frac{\omega^i_{s-1}}{\Omega_{s-1}} \int q(\xi^i_{s-1}, x_s) \mathbb{E}\left[\left|\Delta\psi^N_{\gamma,t|T}[\tilde{h}_{s,t}](x_s)\right|\right] \mathrm{d}x_s\right] ,$$

where $\Delta\psi^N_{\gamma,t|T}[\tilde{h}_{s,t}](x_s) := (\psi^N_{\gamma,t|T}[\tilde{h}_{s,t}(x_s,\cdot)])^2 - \psi^2_{\gamma,t|T}[\tilde{h}_{s,t}(x_s,\cdot)]$. Following the same steps as in (31), there exists $D_{t|T}$ such that

$$\mathbb{E}\left[\left|\Delta\psi^N_{\gamma,t|T}[\tilde{h}_{s,t}](x_s)\right|\right] \le 2D_{t|T}\operatorname{osc}(h)^2/\sqrt{N} ,$$

which yields

$$\begin{aligned}\mathbb{E}\left[A^N_{s,t}\right] &\le 2\operatorname{osc}(h)^2 |\omega_s|_\infty |g_s|_\infty D_{T|t}\mathbb{E}\left[\sum_{i=1}^N \frac{\omega^i_{s-1}}{\Omega_{s-1}}\int q(\xi^i_{s-1},x_s)\mathrm{d}x_s\right]/\sqrt{N} ,\\ &\le 2\operatorname{osc}(h)^2 |\omega_s|_\infty |g_s|_\infty D_{T|t}/\sqrt{N}\end{aligned}$$

and $\mathbb{E}\left[A^N_{s,t}\right] \longrightarrow_{N\to\infty} 0$. On the other hand, as $\operatorname{osc}(f_{s-1,t}) \le \operatorname{osc}(h)^2 |\omega_s|_\infty |g_s|_\infty$, $B^N_{s,t} \xrightarrow{\mathrm{P}}_{N\to\infty} 0$ by Proposition 8. Finally, the tightness condition (30) holds since $|U_{N,i}| \le N^{-1/2}|\omega_s|_\infty \operatorname{osc}(h)$. Note that,

$$N^{-1}\Omega_s \xrightarrow{\mathrm{P}}_{N\to\infty} \phi_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]/\phi_{\chi,s-1}[\vartheta_s] .$$

Therefore (23) holds with

$$\begin{aligned}\tilde{\Gamma}_{s,t|T}\left[\tilde{h}_{s,t}\right] &= \frac{\phi^2_{\chi,s-1}[\vartheta_s]}{\phi^2_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]}\left\{\frac{\tilde{\Gamma}_{s-1,t|T}\left[H_{s,t}\right]}{\phi^2_{\chi,s-1}[\vartheta_s]} + \frac{\phi_{\chi,s-1}\left[f_{s-1,t}\right]}{\phi_{\chi,s-1}[\vartheta_s]}\right\} ,\\ &= \frac{\tilde{\Gamma}_{s-1,t|T}\left[H_{s,t}\right]}{\phi^2_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]} + \frac{\phi_{\chi,s-1}\left[f_{s-1,t}\right]\phi_{\chi,s-1}[\vartheta_s]}{\phi^2_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]} ,\end{aligned}$$

where, by induction assumption,

$$\begin{aligned}\tilde{\Gamma}_{s-1,t|T}\left[H_{s,t}\right] = \Gamma_{\chi,s-1}&\left[\int \psi_{\gamma,t|T}(\mathrm{d}x_t)H_{s,t}(\cdot,x_t)\right]\\ &+ \check{\Gamma}_{\gamma,t|T}\left[\int \phi_{\chi,s-1}(\mathrm{d}x_{s-1})H_{s,t}(x_{s-1},\cdot)\right] .\end{aligned}$$

The proof is completed upon noting that

$$\begin{aligned}\frac{\int \phi_{\chi,s-1}(\mathrm{d}x_{s-1})H_s(x_{s-1},\cdot)}{\phi_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]} &= \frac{\int \phi_{\chi,s-1}(\mathrm{d}x_{s-1})q(x_{s-1},x_s)g_s(x_s)\tilde{h}_{s,t}(x_s,\cdot)\mathrm{d}x_s}{\phi_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]} ,\\ &= \int \phi_{\chi,s}(\mathrm{d}x_s)\tilde{h}_{s,t}(x_s,\cdot)\end{aligned}$$

and, by Proposition 10,

$$\begin{aligned}\Gamma_{\chi,s-1}\left[\frac{\int \psi_{\gamma,t|T}(\mathrm{d}x_t)H_{s,t}(\cdot,x_t)}{\phi_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]}\right] &+ \frac{\phi_{\chi,s-1}\left[f_{s-1,t}\right]\phi_{\chi,s-1}[\vartheta_s]}{\phi^2_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]}\\ &= \Gamma_{\chi,s}\left[\int \psi_{\gamma,t|T}(\mathrm{d}x_t)\tilde{h}_{s,t}(\cdot,x_t)\right] .\end{aligned}$$

5.5 Proof of Theorem 5

Write $\tilde{h}_{s,T} = h - \phi_{\chi,s|T}(h)$. Note that

$$\sqrt{N}\sum_{i=1}^{N}\frac{\tilde{\omega}^i_{s|T}}{\tilde{\Omega}_{s|T}}\tilde{h}_{s,T}(\tilde{\xi}^i_{s|T}) = \left(\tilde{\Omega}_{s|T}/N\right)^{-1}\sum_{i=1}^{N}U_{N,i} \;,$$

where $U_{N,\ell} := N^{-1/2}\tilde{\omega}^\ell_{s|T}\tilde{h}_{s,T}(\tilde{\xi}^\ell_{s|T})$. Set, for $i \in \{1,\dots,N\}$,

$$\mathcal{G}_{N,i} := \sigma\left\{\{(\tilde{\omega}^\ell_{s|T},\tilde{\xi}^\ell_{s|T})\}_{\ell=1}^{i}\;,\;\{(\omega^\ell_u,\xi^\ell_u)\}_{\ell=1}^{N}, u=0,\dots,s-1,\right.$$
$$\left.\{(\check{\omega}^\ell_{v|T},\check{\xi}^\ell_{v|T})\}_{\ell=1}^{N}, v=s+1,\dots,T\right\}\;.$$

By the proof of Theorem 2,

$$N^{-1}\tilde{\Omega}_{s|T} \xrightarrow{\mathrm{P}}_{N\to\infty} \frac{\phi_{\chi,s-1}\otimes\psi_{s+1|T}\left[\int q^{[2]}(\cdot,x)g_s(x)\mathrm{d}x \odot \gamma^{-1}_{s+1}\right]}{\phi_{\chi,s-1}\otimes\psi_{s+1|T}\left[\tilde{\vartheta}_{s|T}\odot\gamma^{-1}_{s+1}\right]}\;.$$

The proof therefore amounts to establish a CLT for $\sum_{\ell=1}^N U_{N,\ell}$ and then to use Slutsky Lemma. The limit distribution of $\sum_{\ell=1}^N U_{N,\ell}$ is again obtained using the invariance principle for triangular array of dependent random variables derived in [7]. As

$$\sum_{i=1}^{N}\mathbb{E}\left[U_{N,i}\,|\,\mathcal{G}_{N,i-1}\right]$$
$$= \sqrt{N}\frac{\sum_{i,j=1}^{N}\omega^i_{s-1}\check{\omega}^j_{s+1|T}\int q^{[2]}(\cdot;x_s)g_s(x_s)\tilde{h}_{s,T}(x_s)\mathrm{d}x_s\odot\gamma^{-1}_{s+1}(\xi^i_{s-1},\check{\xi}^j_{s+1|T})}{\sum_{i,j=1}^{N}\omega^i_{s-1}\check{\omega}^j_{s+1|T}\gamma^{-1}_{s+1}(\check{\xi}^j_{s+1|T})\tilde{\vartheta}_{s|T}(\xi^i_{s-1},\check{\xi}^j_{s+1|T})}\;,$$

it follows from Theorems 1 and 4 that

$$\sum_{i=1}^{N}\mathbb{E}\left[U_{N,i}\,|\,\mathcal{G}_{N,i-1}\right] \xrightarrow{\mathrm{P}}_{N\to\infty} \mathcal{N}\left(0,\frac{\Sigma_s[\tilde{h}_{s,T}]}{\left(\phi_{\chi,s-1}\otimes\psi_{s+1|T}[\tilde{\vartheta}_{s|T}\odot\gamma^{-1}_{s+1}]\right)^2}\right)\;.$$

Using that

$$\phi_{\chi,s|T}[\tilde{h}_{s,T}] = \frac{\phi_{\chi,s-1}\otimes\psi_{s+1|T}\left[\int q^{[2]}(\cdot;x)g_s(x)\tilde{h}_{s,T}(x)\mathrm{d}x\right]\odot\gamma^{-1}_{s+1}}{\phi_{\chi,s-1}\otimes\psi_{s+1|T}\left[\int q^{[2]}(\cdot;x)g_s(x)\mathrm{d}x\right]\odot\gamma^{-1}_{s+1}} = 0\;,$$

Theorem 1 yields

$$\sum_{i=1}^{N}\mathbb{E}\left[U_{N,i}\,|\,\mathcal{G}_{N,i-1}\right]^2 =$$
$$\left(\frac{\sum_{i,j=1}^{N}\omega^i_{s-1}\check{\omega}^j_{s+1|T}\gamma^{-1}_{s+1}(\check{\xi}^j_{s+1|T})\int q^{[2]}(\xi^i_{s-1},\check{\xi}^j_{s+1|T};x)g_s(x)\tilde{h}_{s,T}(x)}{\sum_{i,j=1}^{N}\omega^i_{s-1}\check{\omega}^j_{s+1|T}\gamma^{-1}_{s+1}(\check{\xi}^j_{s+1|T})\tilde{\vartheta}_{s|T}(\xi^i_{s-1},\check{\xi}^j_{s+1|T})}\right)^2$$
$$\xrightarrow{\mathrm{P}}_{N\to\infty}\left(\frac{\phi_{\chi,s-1}\otimes\psi_{s+1|T}\left[\int q^{[2]}(\cdot;x)g_s(x)\tilde{h}_{s,T}(x)\mathrm{d}x\odot\gamma^{-1}_{s+1}\right]}{\phi_{\chi,s-1}\otimes\psi_{s+1|T}[\tilde{\vartheta}_{s|T}\odot\gamma^{-1}_{s+1}]}\right)^2 = 0\;.$$

Similarly, using again Theorem 1,

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i}^2 \,\middle|\, \mathcal{G}_{N,i-1} \right] \xrightarrow{\mathrm{P}}_{N\to\infty} \frac{\phi_{\chi,s-1} \otimes \psi_{s+1|T}\left[\int \tilde{\omega}_{s|T}(\cdot\,;x) q^{[2]}(\cdot\,;x) g_s(x) \tilde{h}_{s,T}^2(x)\mathrm{d}x \odot \gamma_{s+1}^{-1} \right]}{\phi_{\chi,s-1} \otimes \psi_{s+1|T}[\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}]} \;.$$

Since under A2, $|U_{N,i}| \le N^{-1/2} \left|\tilde{\omega}_{s|T}\right|_\infty \operatorname{osc}(h)$, for any $\epsilon > 0$,

$$\sum_{i=1}^{N} \mathbb{E}\left[U_{N,i}^2 \mathbf{1}\{|U_{N,i}| \ge \epsilon\} \,\middle|\, \mathcal{G}_{N,i-1} \right] \xrightarrow{\mathrm{P}}_{N\to\infty} 0 \;,$$

which concludes the proof.

A Exponential deviation inequalities for the forward filter and the backward information filter

The following result is proved in [6].

Lemma 7. *Assume that a_N, b_N, and b are random variables defined on the same probability space such that there exist positive constants β, B, C, and M satisfying*

(i) $|a_N/b_N| \leq M$, *$\mathbb{P}$-a.s. and* $b \geq \beta$, *$\mathbb{P}$-a.s.,*

(ii) For all $\epsilon > 0$ *and all* $N \geq 1$, $\mathbb{P}\left[|a_N| > \epsilon\right] \leq B\mathrm{e}^{-CN(\epsilon/M)^2}$,

(iii) For all $\epsilon > 0$ *and all* $N \geq 1$, $\mathbb{P}\left[|b_N - b| > \epsilon\right] \leq B\mathrm{e}^{-CN\epsilon^2}$.

Then, for all $\varepsilon > 0$,

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

Proposition 8 provides an exponential deviation inequality for the forward filter and is proved in [6].

Proposition 8. *Assume that A1 and A2 hold for some* $T > 0$. *Then, for all* $s \geq 1$, *there exist* $0 < B_s, C_s < \infty$ *such that for all* $N \geq 1$, $\epsilon > 0$, *and all* $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,

$$\mathbb{P}\left(\left|\Omega_s^{-1}\sum_{i=1}^N \omega_s^i h(\xi_s^i) - \phi_{\chi,s}[h]\right| \geq \epsilon\right) \leq B_s\mathrm{e}^{-C_s N\epsilon^2/\operatorname{osc}(h)^2} .$$

Proposition 9 provides an exponential inequality for the backward information filter $\psi_{\gamma,t|T}$ and its unnormalized approximation. Its proof is similar to the proof [6, Theorem 5] and is omitted.

Proposition 9. *Assume that A1 and A3 hold for some* $T > 0$. *Then, for all* $0 \leq t \leq T$, *there exist* $0 < B_{t|T}, C_{t|T} < \infty$ *such that for all* $N \geq 1$, $\epsilon > 0$, *and all* $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,

$$\mathbb{P}\left[\left|\check{\Omega}_{t|T}^{-1}\sum_{i=1}^N \check{\omega}_{t|T}^i h(\check{\xi}_{t|T}^i) - \psi_{\gamma,t|T}[h]\right| \geq \epsilon\right] \leq B_{t|T}\mathrm{e}^{-C_{t|T}N\epsilon^2/\operatorname{osc}(h)^2} .$$

B Asymptotic normality of the forward filter and the backward information filter

Proposition 10 provides a CLT for the weighted particles $\{(\omega_s^i, \xi_s^i)\}_{i=1}^N$ approximating the filtering distribution $\phi_{\chi,s}$ and is proved for instance in [4].

Proposition 10. *Assume that A1 and A2 hold for some* $T > 0$. *Then, for all* $0 \leq s \leq T$ *and all* $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,

$$N^{1/2}\left(\sum_{i=1}^N \frac{\omega_s^i}{\Omega_s} h(\xi_s^i) - \phi_{\chi,s}[h]\right) \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \Gamma_{\chi,s}\left[h - \phi_{\chi,s}[h]\right]\right) ,$$

where

$$\Gamma_{\chi,0}\left[h\right] := \frac{\int \rho_0(\mathrm{d}x_0)\omega_0^2(x_0)h^2(x_0)}{\left(\int \rho_0(\mathrm{d}x_0)\omega_0(x_0)\right)^2} \quad \textit{and for all } s \geq 1 \;,$$

$$\Gamma_{\chi,s}\left[h\right] := \frac{\Gamma_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)h(x_s)\mathrm{d}x_s\right]}{\phi^2_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]} + \frac{\phi_{\chi,s-1}\left[\int \omega_s(\cdot,x_s)q(\cdot,x_s)g_s(x_s)h^2(x_s)\mathrm{d}x_s\right]\phi_{\chi,s-1}[\vartheta_s]}{\phi^2_{\chi,s-1}\left[\int q(\cdot,x_s)g_s(x_s)\mathrm{d}x_s\right]} \;.$$

Proposition 11 provides a CLT for the weighted particles $\{(\check{\omega}^j_{t|T}, \check{\xi}^j_{t|T})\}_{j=1}^N$ approximating the backward information filter. Its proof follows the same lines as the proof of Proposition 10 and is omitted for brevity.

Proposition 11. *Assume that A1 and A3 hold. Then, for all $0 \leq t \leq T$ and all $h \in \mathbb{F}_b(\mathsf{X}, \mathcal{X})$,*

$$N^{1/2}\left(\sum_{j=1}^N \frac{\check{\omega}^j_{t|T}}{\check{\Omega}_{t|T}} h(\check{\xi}^j_{t|T}) - \psi_{\gamma,t|T}[h]\right) \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \check{\Gamma}_{\gamma,t|T}\left[h - \psi_{\gamma,t|T}[h]\right]\right) \;,$$

where

$$\check{\Gamma}_{\gamma,T|T}\left[h\right] := \frac{\int \check{\rho}_T(\mathrm{d}x_T)\check{\omega}^2_{T|T}(x_T)h^2(x_T)}{\left(\int \check{\rho}_T(\mathrm{d}x_T)\check{\omega}_{T|T}(x_T)\right)^2} \quad \textit{and for all } t \leq T-1 \;,$$

$$\check{\Gamma}_{\gamma,t|T}\left[h\right] := \frac{\check{\Gamma}_{\gamma,t+1|T}\left[\int \gamma_t(x_t)g_t(x_t)q(x_t,\cdot)\gamma^{-1}_{t+1}(\cdot)h(x_t)\mathrm{d}x_t\right]}{\psi^2_{\gamma,t+1|T}\left[\int \gamma_t(x_t)g_t(x_t)q(x_t,\cdot)\gamma^{-1}_{t+1}(\cdot)\mathrm{d}x_t\right]} + \frac{\psi_{\gamma,t+1|T}\left[\int \check{\omega}_t(x_t,\cdot)q(x_t,\cdot)g_t(x_t)\gamma_t(x_t)\gamma^{-1}_{t+1}(\cdot)h^2(x_t)\mathrm{d}x_t\right]\psi_{\gamma,t+1|T}\left[\vartheta_{t|T}\gamma^{-1}_{t+1}\right]}{\psi^2_{\gamma,t+1|T}\left[\int \gamma_t(x_t)g_t(x_t)q(x_t,\cdot)\gamma^{-1}_{t+1}(\cdot)\mathrm{d}x_t\right]} \;.$$

References

[1] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state-space models. *Annals Institute Statistical Mathematics*, 62(1):61–89, 2010.

[2] O. Cappe. Online EM algorithm for hidden Markov models. *Journal of Computational and Graphical Statistics*, 20(73):728–749, 2011.

[3] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

[4] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[5] P. Del Moral, A. Doucet, and S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5):947–975, 2010.

[6] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.*, 21(6):2109–2145, 2011.

[7] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, 36(5):2344–2376, 2008.

[8] R. Douc, E. Moulines, and D.S. Stoffer. *Nonlinear time series: theory, methods and applications with R examples.* CRC Press, 2014.

[9] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice.* Springer, 2001.

[10] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10:197–208, 2000.

[11] C. Dubarry and S. Le Corff. Non-asymptotic deviation inequalities for smoothed additive functionals in nonlinear state-space models. *Bernoulli*, 19(5B):2222–2249, 2013.

[12] P. Fearnhead, D. Wyncoll, and J. Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447–464, 2010.

[13] S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.*, 50:438–449, 2004.

[14] N. Gordon, D. Salmond, and A.F. Smith. Novel approach to nonlinear/non-Gaussian bayesian state estimation. *IEE Proc. F, Radar Signal Process*, 140:107–113, 1993.

[15] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.*, 7:175–193, 1998.

[16] P.E. Jacob, L.M. Murray, and S. Rubenthaler. Path storage in the particle filter. *Statistics and Computing*, pages 1–10, 2013.

[17] N. Kantas, A. Doucet, S.S. Singh, J. Maciejowski, and N. Chopin. On particle methods for parameter estimation in state-space models. *Statist. Sci.*, 30(3):328–351, 2015.

[18] G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, 1:1–25, 1996.

[19] H. R. Künsch. Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.*, 33(5):1983–2021, 2005.

[20] S. Le Corff and G. Fort. Convergence of a particle-based approximation of the block online Expectation Maximization algorithm. *ACM Transactions on Modeling and Computer Simulation*, 23(1):2, 2013.

[21] S. Le Corff and G. Fort. Online Expectation Maximization based algorithms for inference in hidden Markov models. *Electronic Journal of Statistics*, 7:763–792, 2013.

[22] G. Mongillo and S. Deneve. Online learning with hidden Markov models. *Neural Computation*, 20(7):1706–1716, 2008.

[23] J. Olsson and J. Westerborn. Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm. *ArXiv:1412.7550*, 2015.

[24] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, 94(446):590–599, 1999.

[25] G. Poyiadjis, A. Doucet, and S.S Singh. Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98:65–80, 2011.