

SEQUENTIAL MONTE CARLO SMOOTHING FOR GENERAL STATE SPACE HIDDEN MARKOV MODELS¹

BY RANDAL DOUC, AURÉLIEN GARIVIER, ERIC MOULINES
AND JIMMY OLSSON

*Institut Télécom/Télécom SudParis, CNRS UMR 5157, CNRS/Télécom ParisTech,
CNRS UMR 5141, Institut Télécom/Télécom ParisTech, CNRS UMR 5141
and Lund University*

Computing smoothing distributions, the distributions of one or more states conditional on past, present, and future observations is a recurring problem when operating on general hidden Markov models. The aim of this paper is to provide a foundation of particle-based approximation of such distributions and to analyze, in a common unifying framework, different schemes producing such approximations. In this setting, general convergence results, including exponential deviation inequalities and central limit theorems, are established. In particular, time uniform bounds on the marginal smoothing error are obtained under appropriate mixing conditions on the transition kernel of the latent chain. In addition, we propose an algorithm approximating the joint smoothing distribution at a cost that grows only linearly with the number of particles.

1. Introduction. Statistical inference in general state space hidden Markov models (HMM) involves computation of the *posterior distribution* of a set $X_{s:s'} \stackrel{\text{def}}{=} [X_s, \dots, X_{s'}]$ of state variables conditional on a record $Y_{0:T} = y_{0:t}$ of observations. This distribution will, in the following, be denoted by $\phi_{s:s'|T}$ where the dependence of this measure on the observed values $y_{0:T}$ is implicit. The posterior distribution can be expressed in closed-form only in very specific cases, principally, when the state space model is linear and Gaussian or when the state space of the hidden Markov chain is a finite set. In the vast majority of cases, nonlinearity or non-Gaussianity render analytic solutions intractable [3, 26, 33, 36].

This limitation has led to an increase of interest in alternative computational strategies handling more general state and measurement equations without constraining a priori the behavior of the posterior distributions. Among these, *sequential Monte Carlo (SMC) methods* play a central role. SMC methods—in which the *sequential importance sampling* and *sampling importance resampling* methods proposed by [23] and [35], respectively, are combined—refer to a class of

Received July 2009; revised March 2010.

¹Supported in part by the French National Research Agency, under the program ANR-07 ROBO 0002 CFLAM.

MSC2010 subject classifications. Primary 60G10, 60K35; secondary 60G18.

Key words and phrases. Sequential Monte Carlo methods, particle filter, smoothing, hidden Markov models.

algorithms approximating a *sequence of probability distributions*, defined on a *sequence of probability spaces*, by updating recursively a set of random *particles* with associated nonnegative *importance weights*. The SMC methodology has emerged as a key tool for approximating state posterior distribution flows in general state space models; see [9, 10, 12] for general introductions as well as theoretical results for SMC methods and [17, 31, 33] for applications of SMC within a variety of scientific fields.

The recursive formulas generating the *filter distributions* ϕ_T (short-hand notation for $\phi_{T:T|T}$) and the *joint smoothing distributions* $\phi_{0:T|T}$ are closely related; thus, executing the standard SMC scheme in the filtering mode provides, as a by-product, approximations of the joint smoothing distributions. More specifically, the branches of the genealogical tree associated with the historical evolution of the filtering particles up to time step T form, when combined with the corresponding importance weights of these filtering particles, a weighted sample approximating the joint smoothing distribution $\phi_{0:T|T}$; see [9], Section 3.4, for details. From these paths, one may readily obtain a weighted sample targeting the fixed lag or fixed interval smoothing distribution by extracting the required subsequence of states while retaining the weights. This appealingly simple scheme can be used successfully for estimating the joint smoothing distribution for small values of T or any marginal smoothing distribution $\phi_{s|T}$, with $s \leq T$, when s and T are close; however, when T is large and $s \ll T$, the associated particle approximations are inaccurate since the genealogical tree degenerates gradually as the interacting particle system evolves [20, 21].

In this article, we thus give attention to more sophisticated approaches and consider instead the *forward filtering backward smoothing (FFBSm) algorithm* and the *forward filtering backward simulation (FFBSi) sampler*. These algorithms share some similarities with the Baum–Welch algorithm for finite state space models and the Kalman filter-based smoother and simulation smoother for linear Gaussian state space models [8]. In the FFBSm algorithm, the particle weights obtained when approximating the filter distributions in a forward filtering pass are modified in a backward pass; see [18, 24, 27]. The FFBSi algorithm simulates, conditionally independently given the particles and particle weights produced in a similar forward filtering pass, state trajectories being approximately distributed according to the joint smoothing distribution; see [21].

The computational complexity of the FFBSm algorithm when used for estimating marginal fixed interval smoothing distributions or of the original formulation of the FFBSi sampler grows (in most situations) as the square of the number N of particles multiplied by the time horizon T . To alleviate this potentially very large computational cost, some methods using intricate data structures for storing the particles have been developed; see, for example, [28]. These algorithms have a complexity of order $O(N \log(N))$ and are thus amenable to practical applications; however, this reduction in complexity comes at the cost of introducing some level of approximation.

In this paper, a modification of the original FFBSi algorithm is presented. The proposed scheme has a complexity that grows only *linearly* in N and does not involve any numerical approximation techniques. This algorithm may be seen as an alternative to a recent proposal by [20] which is based on the so-called *two-filter algorithm* [2].

The smoothing weights computed in the backward pass of the FFBSm algorithm at a given time instant s (or the law of the FFBSi algorithm) are statistically dependent on all forward filtering pass particles and weights computed before and after this time instant. This intricate dependence structure makes the analysis of the resulting particle approximation challenging; up to our best knowledge, only a single consistency result is available in [21], but its proof is plagued by a (subtle) mistake that seems difficult to correct. Therefore, very little is known about the convergence of the schemes under consideration, and the second purpose of this paper is to fill this gap.² In this contribution, we focus first on finite time horizon approximations. Given a finite time horizon T , we derive *exponential deviation inequalities* stating that the probability of obtaining, when replacing $\phi_{s:T|T}$ by the corresponding FFBSm or FFBSi estimator, a Monte Carlo error exceeding a given $\varepsilon > 0$ is bounded by a quantity of order $O(\exp(-cN\varepsilon^2))$ where c is positive constant depending on T as well as the target function under consideration. The obtained inequalities, which are presented in Theorem 5 (FFBSm) and Corollary 6 (FFBSi), hold for any given number N of particles and are obtained by combining a novel backward error decomposition with an adaptation of the Hoeffding inequality to statistics expressed as ratios of random variables. We then consider the asymptotic (as the number N of particles tends to infinity) regime and establish a central limit theorem (CLT) with rate \sqrt{N} and with an explicit expression of the asymptotic variance; see Theorem 8. The proof of our CLT relies on a technique, developed gradually in [6, 15, 30], which is based on a CLT for triangular arrays of dependent random variables; however, since we are required to take the complex dependence structure of the smoothing weights into account, our proof is significantly more involved than in the standard filtering framework considered in the mentioned works.

The second part of the paper is devoted to time uniform results, and we here study the behavior of the particle-based marginal smoothing distribution approximations as the time horizon T tends to infinity. In this setting, we first establish, under the assumption that the Markov transition kernel M of the latent signal is strongly mixing (Assumption 4), time uniform deviation bounds of the type described above which hold for any particle population size N and where the constant c is *independent* of T ; see Theorem 11. This result may seem surprising,

²Since the first version of this paper has been released, an article [11] has been published. This work, developed completely independently from ours, complement the results presented in this manuscript. In particular, this paper presents a functional central limit theorems as well as nonasymptotic variance bounds. Additionally, this work shows how the forward filtering backward smoothing estimates of additive functionals can be computed using a forward only recursion.

and the nonobvious reason for its validity stems from the fact that the underlying Markov chain forgets, when evolving conditionally on the observations, its initial conditions in the forward *as well as* the backward directions. Finally, we prove (see Theorem 12), under the same uniform mixing assumption, that the asymptotic variance of the CLT for the particle-based marginal smoothing distribution approximations remains bounded as T tends to infinity. The uniform mixing assumption in Assumption 4 points typically to applications where the state space of the latent signal is compact; nevertheless, in the light of recent results on filtering stability [14, 29] one may expect the geometrical contraction of the backward kernel to hold for a significantly larger class of nonuniformly mixing models (see [14] for examples from, e.g., financial economics). But even though the geometrical mixing rate is supposed to be constant in this more general case, applying the mentioned results will yield a bound of contraction containing a multiplicative constant depending highly on the initial distributions as well as the observation record under consideration. Since there are currently no available results describing this dependence, applying such bounds to the instrumental decomposition used in the proof of Theorem 5 seems technically involved. Recently, [39] managed to derive *qualitative* time average convergence results for standard (bootstrap-type) particle filters under a mild tightness assumption being satisfied also in the noncompact case when the hidden chain is geometrically ergodic. Even though this technique does not (on the contrary to our approach) supply a rate of convergence, it could possibly be adopted to our framework in order to establish time average convergence of the particle-based marginal smoothing distribution approximations in a noncompact setting.

The paper is organized as follows. In Section 2, the FFBSm algorithm and the FFBSi sampler are introduced. An exponential deviation inequality for the fixed interval joint smoothing distribution is derived in Section 3.1, and a CLT is established in Section 3.2. In Section 4, time uniform exponential bounds on the error of the FFBSm marginal smoothing distribution estimator are computed under the mentioned mixing condition on the kernel M . Finally, under the same mixing condition, an explicit bound on the asymptotic variance of the marginal smoothing distribution estimator is derived in Section 4.2.

Notation and definitions. For any sequence $\{a_n\}_{n \geq 0}$ and any pair of integers $0 \leq m \leq n$, we denote $a_{m:n} \stackrel{\text{def}}{=} (a_m, \dots, a_n)$. We assume in the following that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sets \mathbb{X} and \mathbb{Y} are supposed to be Polish spaces and we denote by $\mathcal{B}(\mathbb{X})$ and $\mathcal{B}(\mathbb{Y})$ the associated Borel σ -algebras. $\mathcal{F}_b(\mathbb{X})$ denotes the set of all bounded $\mathcal{B}(\mathbb{X})/\mathcal{B}(\mathbb{R})$ -measurable functions from \mathbb{X} to \mathbb{R} . For any measure ζ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ and any ζ -integrable function f , we set $\zeta(f) \stackrel{\text{def}}{=} \int_{\mathbb{X}} f(x)\zeta(dx)$. Two measures ζ and ζ' are said to be *proportional* (written $\zeta \propto \zeta'$) if they differ only by a normalization constant.

A kernel V from $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ to $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ is a mapping from $\mathbb{X} \times \mathcal{B}(\mathbb{Y})$ into $[0, 1]$ such that, for each $A \in \mathcal{B}(\mathbb{Y})$, $x \mapsto V(x, A)$ is a nonnegative, bounded, and measurable function on \mathbb{X} , and, for each $x \in \mathbb{X}$, $A \mapsto V(x, A)$ is a measure on $\mathcal{B}(\mathbb{Y})$. For $f \in \mathcal{F}_b(\mathbb{X})$ and $x \in \mathbb{X}$, denote by $V(x, f) \stackrel{\text{def}}{=} \int V(x, dx') f(x')$; we will sometimes also use the abridged notation $Vf(x)$ instead of $V(x, f)$. For a measure ν on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, we denote by νV the measure on $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ defined by, for any $A \in \mathcal{B}(\mathbb{Y})$, $\nu V(A) \stackrel{\text{def}}{=} \int_{\mathbb{X}} V(x, A) \nu(dx)$.

Consider now a possibly nonlinear state space model, where the *state process* $\{X_t\}_{t \geq 0}$ is a Markov chain on the state space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. Even though t is not necessarily a temporal index, we will often refer to this index as “time.” We denote by χ and M the initial distribution and transition kernel, respectively, of this process. The state process is assumed to be hidden but partially observed through the *observations* $\{Y_t\}_{t \geq 0}$ which are \mathbb{Y} -valued random variables being conditionally independent given the latent state sequence $\{X_t\}_{t \geq 0}$; in addition, there exists a σ -finite measure λ on $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ and a nonnegative transition density function g on $\mathbb{X} \times \mathbb{Y}$ such that $\mathbb{P}[Y_t \in A | X_t] = \int_A g(X_t, y) \lambda(dy)$ for all $A \in \mathcal{B}(\mathbb{Y})$. The mapping $x \mapsto g(x, y)$ is referred to as the *likelihood function* of the state given an observed value $y \in \mathbb{Y}$. The kernel M as well as the transition density g are supposed to be known. In the setting of this paper, we assume that we have access to a record of arbitrary but fixed observations $y_{0:T} \stackrel{\text{def}}{=} [y_0, \dots, y_T]$, and our main task is to estimate the posterior distribution of (different subsets of) the state vector $X_{0:T}$ given these observations. For any $t \geq 0$, we denote by $g_t(x) \stackrel{\text{def}}{=} g(x, y_t)$ (where the dependence on y_t is implicit) the likelihood function of the state X_t given the observation y_t .

For simplicity, we consider a *fully dominated* state space model for which there exists a σ -finite measure ν on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ such that, for all $x \in \mathbb{X}$, $M(x, \cdot)$ has a transition probability density $m(x, \cdot)$ with respect to ν . For notational simplicity, $\nu(dx)$ will sometimes be replaced by dx .

For any initial distribution χ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ and any $0 \leq s \leq s' \leq T$, denote by $\phi_{s:s'|T}$ the posterior distribution of the state vector $X_{s:s'}$ given the observations $y_{0:T}$. For lucidity, the dependence of $\phi_{s:s'|T}$ on the initial distribution χ is omitted. Assuming that $\int \dots \int \chi(dx_0) \prod_{u=1}^T g_{u-1}(x_{u-1}) M(x_{u-1}, dx_u) g_T(x_T) > 0$, this distribution may be expressed as, for all $h \in \mathcal{F}_b(\mathbb{X}^{s'-s+1})$,

$$\phi_{s:s'|T}(h) = \frac{\int \dots \int \chi(dx_0) \prod_{u=1}^T g_{u-1}(x_{u-1}) M(x_{u-1}, dx_u) g_T(x_T) h(x_{s:s'})}{\int \dots \int \chi(dx_0) \prod_{v=1}^T g_{v-1}(x_{v-1}) M(x_{v-1}, dx_v) g_T(x_T)}$$

In the expression above, the dependence on the observation sequence is implicit. If $s = s'$, we use $\phi_{s|T}$ (the marginal smoothing distribution at time s) as shorthand for $\phi_{s:s|T}$. If $s = s' = T$, we denote by $\phi_s \stackrel{\text{def}}{=} \phi_{s|s}$ the filtering distribution at time s .

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. Denote by B_η the so-called *backward kernel* given by, for any probability measure η on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$,

$$(1) \quad B_\eta(x, h) \stackrel{\text{def}}{=} \frac{\int \eta(dx') m(x', x) h(x')}{\int \eta(dx') m(x', x)}, \quad h \in \mathcal{F}_b(\mathbb{X}).$$

The posterior distribution $\phi_{s:T|T}$ may be expressed as, for any integers $T > 0$, $s \in \{0, \dots, T - 1\}$ and any $h \in \mathcal{F}_b(\mathbb{X}^{T-s+1})$,

$$(2) \quad \phi_{s:T|T}(h) = \int \cdots \int \phi_T(dx_T) B_{\phi_{T-1}}(x_T, dx_{T-1}) \cdots B_{\phi_s}(x_{s+1}, dx_s) h(x_{s:T}).$$

Therefore, the joint smoothing distribution may be computed recursively, backward in time, according to

$$(3) \quad \phi_{s:T|T}(h) = \int \cdots \int B_{\phi_s}(x_{s+1}, dx_s) \phi_{s+1:T|T}(dx_{s+1:T}) h(x_{s:T}).$$

2.1. *The forward filtering backward smoothing algorithm.* As mentioned in the **Introduction**, the method proposed by [18, 24] for approximating the smoothing distribution is a two pass procedure. In the forward pass, particle approximations ϕ_s^N of the filter distributions ϕ_s are computed recursively for all time steps from $s = 0$ up to $s = T$. The filter distribution flow $\{\phi_s\}_{s \geq 0}$ satisfies the forward recursion

$$(4) \quad \phi_s(h) = \frac{\gamma_s(h)}{\gamma_s(\mathbf{1})} \quad \text{where } \gamma_0(h) = \chi(g_0 h), \gamma_s(h) \stackrel{\text{def}}{=} \gamma_{s-1} M(g_s h), s \geq 1,$$

for $h \in \mathcal{F}_b(\mathbb{X})$, with $\mathbf{1}$ being the unity function $x \mapsto 1$ on \mathbb{X} . In terms of SMC, each filter distribution ϕ_s is approximated by means of a set of particles $\{\xi_s^i\}_{i=1}^N$ and associated importance weights $\{\omega_s^i\}_{i=1}^N$ according to

$$(5) \quad \phi_s^N(h) \stackrel{\text{def}}{=} \frac{\gamma_s^N(h)}{\gamma_s^N(\mathbf{1})} \quad \text{where } \gamma_s^N(h) \stackrel{\text{def}}{=} N^{-1} \sum_{i=1}^N \omega_s^i h(\xi_s^i).$$

Having produced, using methods described in Section 2.4 below, a sequence of such weighted samples $\{(\xi_t^i, \omega_t^i)\}_{i=1}^N, 1 \leq t \leq T$, an approximation of the smoothing distribution is constructed in a backward pass by replacing, in (2), the filtering distribution by its particle approximation. This yields

$$(6) \quad \phi_{s:T|T}^N(h) \stackrel{\text{def}}{=} \int \cdots \int \phi_T^N(dx_T) B_{\phi_{T-1}^N}(x_T, dx_{T-1}) \cdots B_{\phi_s^N}(x_{s+1}, dx_s) h(x_{s:T})$$

for any $h \in \mathcal{F}_b(\mathbb{X}^{T-s+1})$. The approximation above can be computed recursively in the backward direction according to

$$(7) \quad \phi_{s:T|T}^N(h) = \int \cdots \int B_{\phi_s^N}(x_{s+1}, dx_s) \phi_{s+1:T|T}^N(dx_{s+1:T}) h(x_{s:T}).$$

Now, by definition,

$$B_{\phi_s^N}(x, h) = \sum_{i=1}^N \frac{\omega_s^i m(\xi_s^i, x)}{\sum_{\ell=1}^N \omega_s^\ell m(\xi_s^\ell, x)} h(\xi_s^i), \quad h \in \mathcal{F}_b(\mathbb{X}),$$

and inserting this expression into (6) gives

$$(8) \quad \phi_{s:T|T}^N(h) = \sum_{i_s=1}^N \cdots \sum_{i_T=1}^N \left(\prod_{u=s+1}^T \frac{\omega_{u-1}^{i_{u-1}} m(\xi_{u-1}^{i_{u-1}}, \xi_u^{i_u})}{\sum_{\ell=1}^N \omega_{u-1}^\ell m(\xi_{u-1}^\ell, \xi_u^{i_u})} \right) \frac{\omega_T^{i_T}}{\Omega_T} h(\xi_s^{i_s}, \dots, \xi_T^{i_T}),$$

of $\phi_{s:T|T}(h)$, where $h \in \mathcal{F}_b(\mathbb{X}^{T-s-1})$ and

$$(9) \quad \Omega_t \stackrel{\text{def}}{=} \sum_{i=1}^N \omega_t^i.$$

The estimator $\phi_{s:T|T}^N$ is impractical since the cardinality of its support grows exponentially with the number $T - s$ of time steps; nevertheless, it plays a key role in the theoretical developments that follow. A more practical approximation of this quantity will be defined in the next section. When the dimension of the input space is moderate, the computational cost of evaluating the estimator can be reduced to $O(N \log N)$ by using the *fast multipole method* as suggested in [28]; note, however, that this method involves approximations that introduce some bias. On the other hand, in certain specific scenarios, such as discrete Markov chains with sparse transition matrices over large state spaces, the complexity can even be reduced to $O(NT)$ without any truncation; see [1].

2.2. *The forward filtering backward simulation algorithm.* The estimator (8) may be understood alternatively by noting that the normalized smoothing weights define a probability distribution on the set $\{1, \dots, N\}^{T-s}$ of trajectories associated with an inhomogeneous Markov chain. Indeed, consider, for $t \in \{0, \dots, T - 1\}$, the Markov transition matrix $\{\Lambda_t^N(i, j)\}_{i,j=1}^N$ given by

$$(10) \quad \Lambda_t^N(i, j) = \frac{\omega_t^j m(\xi_t^j, \xi_{t+1}^i)}{\sum_{\ell=1}^N \omega_t^\ell m(\xi_t^\ell, \xi_{t+1}^i)}, \quad (i, j) \in \{1, \dots, N\}^2.$$

For $1 \leq t \leq T$, denote by

$$(11) \quad \mathcal{F}_t^N \stackrel{\text{def}}{=} \sigma\{Y_{0:T}, (\xi_s^i, \omega_s^i); 0 \leq s \leq t, 1 \leq i \leq N\}$$

the σ -algebra generated by the observations from time 0 to time T as well as the particles and importance weights produced in the forward pass up to time t . The transition probabilities defined in (10) induce an inhomogeneous Markov chain $\{J_u\}_{u=0}^T$ evolving backward in time as follows. At time T , the random index J_T is drawn from the set $\{1, \dots, N\}$ such that J_T takes the value i with a probability

proportional to ω_T^i . At time $t \leq T - 1$ and given that the index J_{t+1} was drawn at time step $t + 1$, the index J_t is drawn from the set $\{1, \dots, N\}$ such that J_t takes the value j with probability $\Lambda_t^N(J_t, j)$. The joint distribution of $J_{0:T}$ is therefore given by, for $j_{0:T} \in \{1, \dots, N\}^{T+1}$,

$$(12) \quad \mathbb{P}[J_{0:T} = j_{0:T} | \mathcal{F}_T^N] = \frac{\omega_T^{j_T}}{\Omega_T} \Lambda_T^N(J_T, j_{T-1}) \cdots \Lambda_0^N(j_1, j_0).$$

Thus, and this is a key observation, the FFBS estimator (8) of the joint smoothing distribution may be written as the conditional expectation

$$(13) \quad \phi_{0:T|T}^N(h) = \mathbb{E}[h(\xi_0^{J_0}, \dots, \xi_T^{J_T}) | \mathcal{F}_T^N], \quad h \in \mathcal{F}_b(\mathbb{X}^{T+1}).$$

We may therefore construct an unbiased estimator of the FFBS estimator by drawing, conditionally independently given \mathcal{F}_T^N , N paths of $\{J_{0:T}^\ell\}_{\ell=1}^N$ of the inhomogeneous Markov chain introduced above and then forming the (practical) estimator

$$(14) \quad \tilde{\phi}_{0:T|T}^N(h) = N^{-1} \sum_{\ell=1}^N h(\xi_0^{J_0^\ell}, \dots, \xi_T^{J_T^\ell}), \quad h \in \mathcal{F}_b(\mathbb{X}^{T+1}).$$

This practical estimator was introduced in [21] (Algorithm 1, page 158). For ease of notation, we have here simulated N replicates of the backward, index-valued Markov chain, but it would of course also be possible to sample a number of paths that is either larger or smaller than N . The estimator $\phi_{0:T|T}^N$ may be seen as a Rao–Blackwellized version of $\tilde{\phi}_{0:T|T}^N$. The variance of the latter is increased, but the gain in computational complexity is significant. The associated algorithm is referred in the sequel to as the forward filtering backward simulation (FFBSi) algorithm. In Section 4, forgetting properties of the inhomogeneous backward chain will play a key role when establishing time uniform stability properties of the proposed smoothing algorithm.

The computational complexity for sampling a single path of $J_{0:T}$ is $O(NT)$; therefore, the overall computational effort spent when estimating $\tilde{\phi}_{0:T|T}^N$ using the FFBSi sampler is $O(N^2T)$. Following [28], this complexity can be reduced further to $O(N \log(N)T)$ by means of the fast multipole method; however, here again computational work is gained at the cost of introducing additional approximations.

2.3. *A fast version of the forward filtering backward simulation algorithm.* We are now ready to describe one of the main contributions of this paper, namely a novel version of the FFBSi algorithm that can be proved to reach linear computational complexity under appropriate assumptions. At the end of the filtering phase of the FFBSi algorithm, all weighted particle samples $\{(\xi_s^i, \omega_s^i)\}_{i=1}^N, 0 \leq s \leq T$, are available, and it remains to sample efficiently index paths $\{J_{0:T}^\ell\}_{\ell=1}^N$ under the distribution (12). When the transition kernel m is bounded from above in the sense that $m(x, x') \leq \sigma_+$ for all $(x, x') \in \mathbb{X} \times \mathbb{X}$, the paths can be simulated

recursively backward in time using the following accept–reject procedure. As in the standard FFBSi algorithm, the recursion is initiated by sampling J_T^1, \dots, J_T^N multinomially with probabilities proportional to $\{\omega_T^i\}_{i=1}^N$. For $s \in \{0, \dots, T\}$, let \mathcal{G}_s^N the smallest σ -field containing \mathcal{F}_T^N and $\sigma(J_t^\ell : 1 \leq \ell \leq N, t \geq s)$; then in order to draw J_s^ℓ conditionally on \mathcal{G}_{s+1}^N , we draw, first, an index proposal I_s^ℓ taking the value $i \in \{1, \dots, N\}$ with a probability proportional to ω_i^ℓ and, second, an independent uniform random variable U_s^ℓ on $[0, 1]$. Then we set $J_s^\ell = I_s^\ell$ if $U_s^\ell \leq m(\xi_s^{I_s^\ell}, \xi_{s+1}^{J_s^\ell})/\sigma_+$; otherwise, we reject the proposed index and make another trial. To create samples of size $n \in \{1, \dots, N\}$ from a multinomial distribution on a set of N elements at lines 1 and 6, Algorithm 1 relies on an efficient procedure described in Appendix B.1 that requires $O(n(1 + \log(1 + N/n)))$ elementary operations; see Proposition 14. Using this technique, the computational complexity of Algorithm 1 can be upper-bounded as follows.

For the bootstrap particle filter as well as the fully adapted auxiliary particle filter (see Section 2.4 for precise descriptions of these SMC filters), it is possible to derive an asymptotic expression for the number of simulations required at line 8 of Algorithm 1 even if the kernel m is not bounded from below. The following result is obtained using theory derived in the coming section.

Algorithm 1 FFBSi-smoothing

```

1: sample  $J_T^1, \dots, J_T^N$  multinomially with probabilities proportional to  $\{\omega_T^i\}_{i=1}^N$ 
2: for  $s$  from  $T - 1$  down to  $0$  do
3:    $L \leftarrow (1, \dots, N)$ 
4:   while  $L$  is not empty do
5:      $n \leftarrow \text{size}(L)$ 
6:     sample  $I_1, \dots, I_n$  multinomially with probabilities proportional to
        $\{\omega_s^i\}_{i=1}^N$ 
7:     sample  $U_1, \dots, U_n$  independently and uniformly over  $[0, 1]$ 
8:      $nL \leftarrow \emptyset$ 
9:     for  $k$  from  $1$  to  $n$  do
10:      if  $U_k \leq m(\xi_s^{I(k)}, \xi_{s+1}^{J_s^{L(k)}})/\sigma_+$  then
11:         $J_s^{L(k)} \leftarrow I_k$ 
12:      else
13:         $nL \leftarrow nL \cup \{L(k)\}$ 
14:      end if
15:    end for
16:     $L \leftarrow nL$ 
17:  end while
18: end for

```

PROPOSITION 1. Assume that the transition kernel is bounded from above, $m(x, x') \leq \sigma_+$ for all $(x, x') \in \mathbb{X} \times \mathbb{X}$. At each iteration $s \in \{0, \dots, T - 1\}$, let Z_s^N be the number of simulations required in the accept–reject procedure of Algorithm 1.

- For the bootstrap auxiliary filter, Z_s^N / N converges in probability to

$$\alpha(s) \stackrel{\text{def}}{=} \sigma_+ \phi_{s|s-1}(g_s) \frac{\int \cdots \int dx_{s+1} \prod_{u=s+2}^T \int m(x_{u-1}, dx_u) g_u(x_u)}{\int \cdots \int \phi_{s|s-1}(dx_s) g_s(x_s) \prod_{u=s+1}^T \int m(x_{u-1}, dx_u) g_u(x_u)}$$

as N goes to infinity.

- In the fully adapted case, Z_s^N / N converges in probability to

$$\beta(s) \stackrel{\text{def}}{=} \sigma_+ \frac{\int \cdots \int dx_{s+1} \prod_{u=s+2}^T \int m(x_{u-1}, dx_u) g_u(x_u)}{\int \cdots \int \phi_s(dx_s) g_s(x_s) \prod_{u=s+1}^T \int m(x_{u-1}, dx_u) g_u(x_u)}$$

as N goes to infinity.

A sufficient condition for ensuring finiteness of $\alpha(s)$ and $\beta(s)$ is that $\int g_u(x_u) dx_u < \infty$ for all $u \geq 0$.

If the transition kernel satisfies stronger mixing conditions, it is possible to derive an upper-bound on the computational complexity of the FFBSi for any auxiliary particle filter, that is, the total number of computations (and not only the total number of simulations). Note that this result is not limited to the bootstrap and the fully adapted cases.

PROPOSITION 2. Assume that the transition kernel is bounded from below and above, that is, $\sigma_- \leq m(x, x') \leq \sigma_+$ for all $(x, x') \in \mathbb{X} \times \mathbb{X}$. Let $C(N, T)$ denote the number of elementary operations required in Algorithm 1. Then, there exists a constant K such that $\mathbb{E}[C(N, T)] \leq KNT\sigma_+/\sigma_-$.

The proofs of Propositions 1 and 2 involve theory developed in the coming section and are postponed to Section 5.

Before concluding this section on reduced complexity, let us mention that efficient smoothing strategies have been considered by [19] using quasi-Monte Carlo methods. The smoother (restricted to be one-dimensional) presented in this work has a complexity that grows quadratically in the number of particles N ; nevertheless, since the variance of the same decays as $O(N^{-2})$ (or faster) thanks to the use of quasi-random numbers, the method is equivalent to methods with complexity growing linearly in N [since the standard Monte Carlo variance is $O(N^{-1})$]. This solution is of course attractive; we are however not aware of extensions of this approach to multiple dimensions.

2.4. *Auxiliary particle filters.* It remains to describe in detail how to produce sequentially the weighted samples $\{(\xi_s^i, \omega_s^i)\}_{i=1}^N, 0 \leq s \leq T$, which can be done in several different ways (see [3, 17, 31] and the references therein). Still, most algorithms may be formulated within the unifying framework of the *auxiliary particle filter* described in the following. Let $\{\xi_0^i\}_{i=1}^N$ be i.i.d. random variables such that $\xi_0^i \sim \rho_0$ and set $\omega_0^i \stackrel{\text{def}}{=} d\chi/d\rho_0(\xi_0^i)g_0(\xi_0^i)$. The weighted sample $\{(\xi_0^i, \omega_0^i)\}_{i=1}^N$ then targets the initial filter ϕ_0 in the sense that $\phi_0^N(h)$ estimates $\phi_0(h)$ for $h \in \mathcal{F}_b(\mathbb{X})$. In order to describe the sequential structure of the auxiliary particle filter, we proceed inductively and assume that we have at hand a weighted sample $\{(\xi_{s-1}^i, \omega_{s-1}^i)\}_{i=1}^N$ targeting ϕ_{s-1} in the same sense. Next, we aim at simulating new particles from the target $\phi_s^{N,t}$ defined as

$$(15) \quad \phi_s^{N,t}(h) = \frac{\gamma_{s-1}^N M(g_s h)}{\gamma_{s-1}^N M(g_s)}, \quad h \in \mathcal{F}_b(\mathbb{X}),$$

in order to produce an updated particle sample approximating the subsequent filter ϕ_s . Following [32], this may be done by considering the *auxiliary target distribution*

$$(16) \quad \phi_s^{N,a}(i, h) \stackrel{\text{def}}{=} \frac{\omega_{s-1}^i M(\xi_{s-1}^i, g_s h)}{\sum_{\ell=1}^N \omega_{s-1}^\ell M(\xi_{s-1}^\ell, g_s h)}, \quad h \in \mathcal{F}_b(\mathbb{X}),$$

on the product space $\{1, \dots, N\} \times \mathbb{X}$ equipped with the product σ -algebra $\mathcal{P}(\{1, \dots, N\}) \otimes \mathcal{B}(\mathbb{X})$. By construction, $\phi_s^{N,t}$ is the marginal distribution of $\phi_s^{N,a}$ with respect to the particle index. Therefore, we may approximate the target distribution $\phi_s^{N,t}$ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ by simulating from the auxiliary distribution and then discarding the indices. More specifically, we first simulate pairs $\{(I_s^i, \xi_s^i)\}_{i=1}^N$ of indices and particles from the instrumental distribution

$$(17) \quad \pi_{s|s}(i, h) \propto \omega_{s-1}^i \vartheta_s(\xi_{s-1}^i) P_s(\xi_{s-1}^i, h), \quad h \in \mathcal{F}_b(\mathbb{X}),$$

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

$$(18) \quad \omega_s^i \stackrel{\text{def}}{=} \frac{m(\xi_{s-1}^{I_s^i}, \xi_s^i) g_s(\xi_s^i)}{\vartheta_s(\xi_{s-1}^{I_s^i}) p_s(\xi_{s-1}^{I_s^i}, \xi_s^i)},$$

such that $\omega_s^i \propto d\phi_s^{N,a}/d\pi_{s|s}(I_s^i, \xi_s^i)$, and associate it to the corresponding particle position ξ_s^i . Finally, the indices $\{I_s^i\}_{i=1}^N$ are discarded whereupon $\{(\xi_s^i, \omega_s^i)\}_{i=1}^N$ is taken as an approximation of ϕ_s . The simplest choice, yielding to the so-called *bootstrap particle filter algorithm* proposed by [22], consists of setting, for all

$x \in \mathbb{X}$, $\vartheta_s(x) \equiv 1$ and $p_s(x, \cdot) \equiv m(x, \cdot)$. A more appealing—but often computationally costly—choice consists of using the adjustment weights $\vartheta_s(x) \equiv \vartheta_s^*(x) \stackrel{\text{def}}{=} \int m(x, x')g_s(x') dx'$, $x \in \mathbb{X}$, and the proposal transition density

$$p_s^*(x, x') \stackrel{\text{def}}{=} \frac{m(x, x')g_s(x')}{\vartheta_s^*(x)}, \quad (x, x') \in \mathbb{X} \times \mathbb{X}.$$

In this case, the auxiliary particle filter is referred to as *fully adapted*. Other choices are discussed in [16] and [7].

3. Convergence of the FFBS and FFBSi algorithms. In this section, the convergence of the FFBS and FFBSi algorithms are studied. For these two algorithms, nonasymptotic Hoeffding-type deviation inequalities and CLTs are obtained. We also introduce a decomposition, serving as a basis for most results obtained in this paper, of the error $\phi_{0:T|T}^N - \phi_{0:T|T}$ and some technical conditions under which the results are derived.

For any function $f: \mathbb{X}^d \rightarrow \mathbb{R}$, we define by $|f|_\infty \stackrel{\text{def}}{=} \sup_{x \in \mathbb{X}^d} |f(x)|$ and $\text{osc}(f) \stackrel{\text{def}}{=} \sup_{(x, x') \in \mathbb{X}^d \times \mathbb{X}^d} |f(x) - f(x')|$ the supremum and oscillator norms, respectively. Denote $\bar{\mathbb{N}} \stackrel{\text{def}}{=} \mathbb{N} \cup \{\infty\}$ and consider the following assumptions where T is the time horizon which can be either a finite integer or infinity.

ASSUMPTION 1. For all $0 \leq t \leq T$, $g_t(\cdot) > 0$ and $\sup_{0 \leq t \leq T} |g_t|_\infty < \infty$.

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

$$(19) \quad \omega_0(x) \stackrel{\text{def}}{=} \frac{d\chi}{d\rho_0}(x)g_0(x) \quad \text{and} \quad \omega_t(x, x') \stackrel{\text{def}}{=} \frac{m(x, x')g_t(x')}{\vartheta_t(x)p_t(x, x')}, \quad t \geq 1.$$

ASSUMPTION 2. $\sup_{1 \leq t \leq T} |\vartheta_t|_\infty < \infty$ and $\sup_{0 \leq t \leq T} |\omega_t|_\infty < \infty$.

The latter assumption is rather mild; it holds in particular under Assumption 1 for the bootstrap filter ($p_t = m$ and $\vartheta_t \equiv 1$) and is automatically fulfilled in the fully adapted case ($\omega_t \equiv 1$).

The coming proofs are based on a decomposition of the joint smoothing distribution that we introduce below. For $0 \leq t < T$ and $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$, define the kernel $L_{t,T}: \mathbb{X}^{t+1} \times \mathcal{B}(\mathbb{X})^{\otimes T+1} \rightarrow [0, 1]$ by

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

and set $L_{T,T}(x_{0:T}, h) \stackrel{\text{def}}{=} h(x_{0:T})$. By construction, for every $t \in \{0, \dots, T\}$, the joint smoothing distribution may be expressed as

$$(21) \quad \phi_{0:T|T}(h) = \frac{\phi_{0:t|t}[L_{t,T}(\cdot, h)]}{\phi_{0:t|t}[L_{t,T}(\cdot, \mathbf{1})]}.$$

This expression extends the classical forward–backward decomposition to the joint smoothing distribution; here $L_{t,T}(\cdot, h)$ plays the role of the so-called backward variable. This suggests to decompose the error $\phi_{0:T|T}^N(h) - \phi_{0:T|T}(h)$ as the following telescoping sum:

$$(22) \quad \begin{aligned} \phi_{0:T|T}^N(h) - \phi_{0:T|T}(h) &= \frac{\phi_0^N[L_{0,T}(\cdot, h)]}{\phi_0^N[L_{0,T}(\cdot, \mathbf{1})]} - \frac{\phi_0[L_{0,T}(\cdot, h)]}{\phi_0[L_{0,T}(\cdot, \mathbf{1})]} \\ &+ \sum_{t=1}^T \left\{ \frac{\phi_{0:t|t}^N[L_{t,T}(\cdot, h)]}{\phi_{0:t|t}^N[L_{t,T}(\cdot, \mathbf{1})]} - \frac{\phi_{0:t-1|t-1}[L_{t-1,T}(\cdot, h)]}{\phi_{0:t-1|t-1}[L_{t-1,T}(\cdot, \mathbf{1})]} \right\}. \end{aligned}$$

The first term on RHS of the decomposition above can be easily dealt with since ϕ_0^N is a weighted empirical distribution associated to i.i.d. random variables.

To cope with the terms in the sum of the RHS in (22), we introduce some kernels (depending on the *past* particles) that stress the dependence with respect to the *current* particles. More precisely, $\phi_{0:t|t}^N[L_{t,T}(\cdot, h)]$ is expressed as

$$(23) \quad \phi_{0:t|t}^N[L_{t,T}(\cdot, h)] = \phi_t^N[\mathcal{L}_{t,T}^N(\cdot, h)] = \frac{\gamma_t^N[\mathcal{L}_{t,T}^N(\cdot, h)]}{\gamma_t^N(\mathbf{1})},$$

where the random kernels $\mathcal{L}_{t,T}^N : \mathbb{X} \times \mathcal{B}(\mathbb{X})^{\otimes(T+1)} \rightarrow [0, 1]$ are defined by: for all $0 < t \leq T$, and $x_t \in \mathbb{X}$,

$$(24) \quad \mathcal{L}_{t,T}^N(x_t, h) \stackrel{\text{def}}{=} \int \cdots \int \mathbf{B}_{\phi_{t-1}^N}(x_t, dx_{t-1}) \cdots \mathbf{B}_{\phi_0^N}(x_1, dx_0) L_{t,T}(x_{0:t}, h),$$

and

$$(25) \quad \mathcal{L}_{0,T}^N(x, h) \stackrel{\text{def}}{=} L_{0,T}(x, h).$$

We stress that the kernels $\mathcal{L}_{t,T}^N$ depend on the particles and weights $(\xi_s^i, \omega_s^i)_{i=1}^N$, $0 \leq s \leq t - 1$, through the particle approximations $\phi_{t-1}^N, \dots, \phi_0^N$ of the filter distributions. When proving the CLT for the FFBS algorithm, it will be crucial to establish that for any $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$, $\mathcal{L}_{t,T}^N(\cdot, h)$ converges (see Lemma 7 below), as the number N of particles tends to infinity, to a deterministic function $\mathcal{L}_{t,T}(\cdot, h)$ given by

$$(26) \quad \mathcal{L}_{t,T}(x_t, h) \stackrel{\text{def}}{=} \int \cdots \int \mathbf{B}_{\phi_{t-1}}(x_t, dx_{t-1}) \cdots \mathbf{B}_{\phi_0}(x_1, dx_0) L_{t,T}(x_{0:t}, h).$$

In the sequel, the case $h = \mathbf{1}$ will be of particular importance; in that case, $L_{t,T}(x_{0:t}, \mathbf{1})$ does not depend on $x_{0:t-1}$, yielding

$$(27) \quad \mathcal{L}_{t,T}^N(x_t, \mathbf{1}) = \mathcal{L}_{t,T}(x_t, \mathbf{1}) = L_{t,T}(x_{0:t}, \mathbf{1})$$

for all $x_{0:t} \in \mathbb{X}^{t+1}$. Using these functions, the difference appearing in the sum in (22) may then be rewritten as

$$\begin{aligned}
 & \frac{\phi_{0:t|t}^N[L_{t,T}(\cdot, h)]}{\phi_{0:t|t}^N[L_{t,T}(\cdot, \mathbf{1})]} - \frac{\phi_{0:t-1|t-1}^N[L_{t-1,T}(\cdot, h)]}{\phi_{0:t-1|t-1}^N[L_{t-1,T}(\cdot, \mathbf{1})]} \\
 (28) \quad &= \frac{1}{\gamma_i^N[\mathcal{L}_{i,T}^N(\cdot, \mathbf{1})]} \left(\gamma_i^N[\mathcal{L}_{i,T}^N(\cdot, h)] - \frac{\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, h)]}{\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, \mathbf{1})]} \gamma_i^N[\mathcal{L}_{i,T}^N(\cdot, \mathbf{1})] \right) \\
 &= \frac{N^{-1} \sum_{\ell=1}^N \omega_i^\ell G_{i,T}^N(\xi_i^\ell, h)}{N^{-1} \sum_{\ell=1}^N \omega_i^\ell \mathcal{L}_{i,T}(\xi_i^\ell, \mathbf{1})},
 \end{aligned}$$

where the kernel $G_{i,T}^N : \mathbb{X} \times \mathcal{B}(\mathbb{X})^{T+1} \rightarrow [0, 1]$ is defined by, for $x \in \mathbb{X}$,

$$(29) \quad G_{i,T}^N(x, h) \stackrel{\text{def}}{=} \mathcal{L}_{i,T}^N(x, h) - \frac{\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, h)]}{\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, \mathbf{1})]} \mathcal{L}_{i,T}^N(x, \mathbf{1}).$$

Similarly to $\mathcal{L}_{i,T}^N(\cdot, h)$, the functions $G_{i,T}^N(\cdot, h)$ depend on the past particles; it will however be shown (see Lemma 7 below) that $G_{i,T}^N(\cdot, h)$ converges to the deterministic function given by, for $x \in \mathbb{X}$,

$$(30) \quad G_{i,T}(x, h) \stackrel{\text{def}}{=} \mathcal{L}_{i,T}(x, h - \phi_{0:T|T}(h)).$$

The key property of this decomposition is stated in the following lemma.

LEMMA 3. *Assume that Assumptions 1–2 hold for some $T < \infty$. Then, for any $0 \leq t \leq T$, the variables $\{\omega_i^\ell G_{i,T}^N(\xi_i^\ell, h)\}_{\ell=1}^N$ are, conditionally on the σ -field \mathcal{F}_{t-1}^N , i.i.d. with zero mean. Moreover, there exists a constant C (that may depend on t and T) such that, for all $N \geq 1$, $\ell \in \{1, \dots, N\}$, and $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$,*

$$|\omega_i^\ell G_{i,T}^N(\xi_i^\ell, h)| \leq |\omega_i|_\infty |G_{i,T}^N(\xi_i^\ell, h)| \leq C \text{osc}(h).$$

PROOF. By construction, all pairs of particles and weights of the weighted sample $\{(\xi_i^\ell, \omega_i^\ell)\}_{\ell=1}^N$ are i.i.d. conditionally on the σ -field \mathcal{F}_{t-1}^N . This implies immediately that the variables $\{\omega_i^\ell G_{i,T}^N(\xi_i^\ell, h)\}_{\ell=1}^N$ are also i.i.d. conditionally on the same σ -field \mathcal{F}_{t-1}^N . We now show that $\mathbb{E}[\omega_i^1 G_{i,T}^N(\xi_i^1, h) | \mathcal{F}_{t-1}^N] = 0$. Using the definition of $G_{i,T}^N$ and the fact that $\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, h)]$ and $\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, \mathbf{1})]$ are \mathcal{F}_{t-1}^N -measurable, we have

$$\begin{aligned}
 & \mathbb{E}[\omega_i^1 G_{i,T}^N(\xi_i^1, h) | \mathcal{F}_{t-1}^N] \\
 &= \mathbb{E}[\omega_i^1 \mathcal{L}_{i,T}^N(x, h) | \mathcal{F}_{t-1}^N] - \frac{\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, h)]}{\phi_{i-1}^N[\mathcal{L}_{i-1,T}^N(\cdot, \mathbf{1})]} \mathbb{E}[\omega_i^1 \mathcal{L}_{i,T}^N(x, \mathbf{1}) | \mathcal{F}_{t-1}^N],
 \end{aligned}$$

which is equal to zero provided that the relation

$$(31) \quad \mathbb{E}[\omega_t^1 \mathcal{L}_{t,T}^N(\xi_t^1, h) | \mathcal{F}_{t-1}^N] = \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N(\cdot, h)]}{\phi_{t-1}^N(\vartheta_t)}$$

holds for any $h \in \mathcal{F}_b(\mathbb{X})$. We now turn to the proof of (31). Note that for any $f \in \mathcal{F}_b(\mathbb{X})$,

$$(32) \quad \begin{aligned} \mathbb{E}[\omega_t^1 f(\xi_t^1) | \mathcal{F}_{t-1}^N] &= \frac{\sum_{\ell=1}^N \omega_{t-1}^\ell \int M(\xi_{t-1}^\ell, dx) g_t(x) f(x)}{\sum_{\ell=1}^N \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell)} \\ &= \frac{\phi_{t-1}^N[M(\cdot, g_t f)]}{\phi_{t-1}^N(\vartheta_t)}. \end{aligned}$$

It turns out that (31) is a consequence of (32) with $f(\cdot) = \mathcal{L}_{t,T}^N(\cdot, h)$, but since $\mathcal{L}_{t-1,T}^N(\cdot, h)$ is in general different from $M(\cdot, g_t \mathcal{L}_{t,T}^N(\cdot, h))$, we have to prove directly that

$$(33) \quad \phi_{t-1}^N[\mathcal{L}_{t-1,T}^N(\cdot, h)] = \phi_{t-1}^N[M(\cdot, g_t \mathcal{L}_{t,T}^N(\cdot, h))].$$

Write

$$(34) \quad \begin{aligned} &\phi_{t-1}^N[M(\cdot, g_t \mathcal{L}_{t,T}^N(\cdot, h))] \\ &= \Omega_t^{-1} \sum_{\ell=1}^N \omega_{t-1}^\ell \int \cdots \int m(\xi_{t-1}^\ell, x_t) g_t(x_t) \left(\prod_{u=1}^t \mathbf{B}_{\phi_{u-1}^N}(x_u, dx_{u-1}) \right) \\ &\quad \times L_{t,T}(x_{0:t}, h) dx_t. \end{aligned}$$

To simplify the expression in the RHS, we will use the two following equalities:

$$(35) \quad \left(\sum_{\ell=1}^N \omega_{t-1}^\ell m(\xi_{t-1}^\ell, x_t) \right) \mathbf{B}_{\phi_{t-1}^N}(x_t, dx_{t-1}) = \sum_{\ell=1}^N \omega_{t-1}^\ell m(x_{t-1}, x_t) \delta_{\xi_{t-1}^\ell}(dx_{t-1}),$$

$$(36) \quad \int M(x_{t-1}, dx_t) g_t(x_t) L_{t,T}(x_{0:t}, h) = L_{t-1,T}(x_{0:t-1}, h).$$

The first relation is derived directly from the definition (1) of the backward kernel, the second is a recursive expression of $L_{t,T}$ which is straightforward from the definition (20). Now, (35) and (36) allow for writing

$$\begin{aligned} &\sum_{\ell=1}^N \omega_{t-1}^\ell \int \cdots \int m(\xi_{t-1}^\ell, x_t) g_t(x_t) \prod_{u=1}^t \mathbf{B}_{\phi_{u-1}^N}(x_u, dx_{u-1}) L_{t,T}(x_{0:t}, h) dx_t \\ &= \sum_{\ell=1}^N \omega_{t-1}^\ell \int \cdots \int M(x_{t-1}, dx_t) g_t(x_t) \delta_{\xi_{t-1}^\ell}(dx_{t-1}) \end{aligned}$$

$$\begin{aligned}
 & \times \prod_{u=1}^{t-1} \mathbf{B}_{\phi_{u-1}^N}(x_u, dx_{u-1}) L_{t,T}(x_{0:t}, h) \\
 &= \sum_{\ell=1}^N \omega_{t-1}^\ell \int \cdots \int \delta_{\xi_{t-1}^\ell}(dx_{t-1}) \prod_{u=1}^{t-1} \mathbf{B}_{\phi_{u-1}^N}(x_u, dx_{u-1}) L_{t-1,T}(x_{0:t-1}, h) \\
 &= \sum_{\ell=1}^N \omega_{t-1}^\ell \mathcal{L}_{t-1}^N(\xi_{t-1}^\ell, h).
 \end{aligned}$$

By plugging this expression into (34), we obtain (33) from which (31) follows via (32). Finally, $\mathbb{E}[\omega_t^1 G_{t,T}^N(\xi_t^1, h) | \mathcal{F}_{t-1}^N] = 0$. It remains to check that the random variable $\omega_t^1 G_{t,T}^N(\xi_t^1, h)$ is bounded. But this is immediate since

$$\begin{aligned}
 (37) \quad |\omega_t^1 G_{t,T}^N(\xi_t^1, h)| &= |\omega_t|_\infty \left| \mathcal{L}_{t,T}^N(\cdot, h) - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N(\cdot, h)]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N(\cdot, \mathbf{1})]} \mathcal{L}_{t,T}(\cdot, \mathbf{1}) \right|_\infty \\
 &\leq 2|\omega_t|_\infty |\mathcal{L}_{t,T}^N(\cdot, \mathbf{1})|_\infty \text{osc}(h) \leq 2|\omega_t|_\infty |L_{t,T}(\cdot, \mathbf{1})|_\infty \text{osc}(h).
 \end{aligned}$$

□

3.1. Exponential deviation inequality. We first establish a nonasymptotic deviation inequality. Considering (28), we are led to prove a Hoeffding inequality for ratios. For this purpose, we use the following elementary lemma which will play a key role in the sequel. The proof is postponed to Appendix A.

LEMMA 4. *Assume that a_N, b_N and b are random variables defined on the same probability space such that there exist positive constants 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 $\varepsilon > 0$ and all $N \geq 1, \mathbb{P}[|b_N - b| > \varepsilon] \leq B e^{-CN\varepsilon^2}$,
- (III) for all $\varepsilon > 0$ and all $N \geq 1, \mathbb{P}[|a_N| > \varepsilon] \leq B e^{-CN(\varepsilon/M)^2}$.

Then

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

THEOREM 5. *Assume that Assumptions 1–2 hold for some $T < \infty$. Then, there exist constants $0 < B$ and $C < \infty$ (depending on T) such that for all $N, \varepsilon > 0$, and all measurable functions $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$,*

$$(38) \quad \mathbb{P}[|\phi_{0:T|T}^N(h) - \phi_{0:T|T}(h)| \geq \varepsilon] \leq B e^{-CN\varepsilon^2/\text{osc}^2(h)}.$$

In addition,

$$(39) \quad N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1}) \xrightarrow{P}_{N \rightarrow \infty} \frac{\phi_{t-1}[\mathcal{L}_{t-1,T}(\cdot, \mathbf{1})]}{\phi_{t-1}(\vartheta_t)}.$$

REMARK 1. As a by-product, Theorem 5 provides an exponential inequality for the particle approximation of the filter. For any $h \in \mathcal{F}_b(\mathbb{X})$, define the function $h_{0:T} : \mathbb{X}^{T+1} \rightarrow \mathbb{R}$ by $h_{0:T}(x_{0:T}) = h(x_T)$. By construction, $\phi_{0:T|T}(h_{0:T}) = \phi_T(h)$ and $\phi_{0:T|T}^N(h_{0:T}) = \phi_T^N(h)$. With this notation, equation (38) may be rewritten as

$$\mathbb{P}[|\phi_T^N(h) - \phi_T(h)| \geq \varepsilon] \leq B e^{-CN\varepsilon^2 / \text{osc}^2(h)}.$$

An inequality of this form was first obtained by [12] (see also [9], Chapter 7).

PROOF. We prove (38) by induction on T using the decomposition (22). Assume that (38) holds at time $T - 1$, for $\phi_{0:T-1|T-1}^N(h)$. Let $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$ and assume without loss of generality that $\phi_{0:T|T}(h) = 0$. Then (21) implies that $\phi_0[L_{0,T}(\cdot, h)] = 0$ and the first term of the decomposition (22) thus becomes

$$(40) \quad \frac{\phi_0^N[L_{0,T}(\cdot, h)]}{\phi_0^N[L_{0,T}(\cdot, \mathbf{1})]} = \frac{N^{-1} \sum_{i=0}^N \frac{d\chi}{d\rho_0}(\xi_0^i) g_0(\xi_0^i) L_{0,T}(\xi_0^i, h)}{N^{-1} \sum_{\ell=0}^N \frac{d\chi}{d\rho_0}(\xi_0^\ell) g_0(\xi_0^\ell) L_{0,T}(\xi_0^\ell, \mathbf{1})},$$

where $\{\xi_0^i\}_{i=1}^N$ are i.i.d. random variables with distribution ρ_0 . We obtain an exponential inequality for (40) by applying Lemma 4 with

$$\begin{cases} a_N = N^{-1} \sum_{i=0}^N \frac{d\chi}{d\rho_0}(\xi_0^i) g_0(\xi_0^i) L_{0,T}(\xi_0^i, h), \\ b_N = N^{-1} \sum_{i=0}^N \frac{d\chi}{d\rho_0}(\xi_0^i) g_0(\xi_0^i) L_{0,T}(\xi_0^i, \mathbf{1}), \\ b = \beta = \chi[g_0(\cdot) L_{0,T}(\cdot, \mathbf{1})]. \end{cases}$$

Condition (I) is trivially satisfied and conditions (II) and (III) follow from the Hoeffding inequality for i.i.d. variables.

By (22) and (28), it is now enough to establish an exponential inequality for

$$(41) \quad \frac{\phi_{0:t|t}^N[L_{t,T}(\cdot, h)]}{\phi_{0:t|t}^N[L_{t,T}(\cdot, \mathbf{1})]} - \frac{\phi_{0:t-1|t-1}^N[L_{t-1,T}(\cdot, h)]}{\phi_{0:t-1|t-1}^N[L_{t-1,T}(\cdot, \mathbf{1})]} = \frac{N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, h)}{N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1})},$$

where $0 < t \leq T$. For that purpose, we use again Lemma 4 with

$$(42) \quad \begin{cases} a_N = N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, h), \\ b_N = N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1}), \\ b = \beta = \frac{\phi_{t-1}[L_{t-1,T}(\cdot, \mathbf{1})]}{\phi_{t-1}(\vartheta_t)}. \end{cases}$$

By considering the LHS of (41), $|a_N/b_N| \leq 2|h|_\infty$, verifying condition (I) in Lemma 4. By Lemma 3, Hoeffding’s inequality implies that there exist constants B and C such that for all $N, \varepsilon > 0$, and all measurable function $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$,

$$\begin{aligned} & \mathbb{P} \left[\left| N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, h) \right| \geq \varepsilon \right] \\ &= \mathbb{E} \left[\mathbb{P} \left[\left| N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, h) \right| \geq \varepsilon \mid \mathcal{F}_{t-1}^N \right] \right] \leq B e^{-CN\varepsilon^2 / \text{osc}^2(h)}, \end{aligned}$$

verifying condition (III) in Lemma 4. It remains to verify condition (II). Since the pairs of particles and weights of the weighted sample $\{(\xi_t^\ell, \omega_t^\ell)\}_{\ell=1}^N$ are i.i.d. conditionally on \mathcal{F}_{t-1}^N , Hoeffding’s inequality implies that

$$(43) \quad \mathbb{P} \left[\left| b_N - \mathbb{E} \left[N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1}) \mid \mathcal{F}_{t-1}^N \right] \right| \geq \varepsilon \right] \leq B e^{-CN\varepsilon^2}.$$

Moreover, by (32), (27), and the definition (20), we have

$$(44) \quad \begin{aligned} & \mathbb{E} \left[N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1}) \mid \mathcal{F}_{t-1}^N \right] - b \\ &= \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}(\cdot, \mathbf{1})]}{\phi_{t-1}^N(\vartheta_t)} - \frac{\phi_{t-1}[\mathcal{L}_{t-1,T}(\cdot, \mathbf{1})]}{\phi_{t-1}(\vartheta_t)} = \frac{\phi_{t-1}^N(H)}{\phi_{t-1}^N(\vartheta_t)}, \end{aligned}$$

with $H(\cdot) \stackrel{\text{def}}{=} \mathcal{L}_{t-1,T}(\cdot, \mathbf{1}) - \phi_{t-1}[\mathcal{L}_{t-1,T}(\cdot, \mathbf{1})]\vartheta_t(\cdot)/\phi_{t-1}(\vartheta_t)$. To obtain an exponential deviation inequality for (44), we apply again Lemma 4 with

$$\begin{cases} a'_N = \phi_{t-1}^N(H), \\ b'_N = \phi_{t-1}^N(\vartheta_t), \\ b' = \beta' = \phi_{t-1}(\vartheta_t). \end{cases}$$

By using the inequality

$$\begin{aligned} & \mathcal{L}_{t-1,T}(x_{t-1}, \mathbf{1}) \\ &= \vartheta_t(x_{t-1}) \int \frac{m(x_{t-1}, x_t) g_t(x_t)}{\vartheta_t(x_{t-1}) p_t(x_{t-1}, x_t)} p_t(x_{t-1}, x_t) \mathcal{L}_{t,T}(x_t, \mathbf{1}) dx_t \\ &\leq \vartheta_t(x_{t-1}) |\omega_t|_\infty |\mathcal{L}_{t,T}(\cdot, \mathbf{1})|_\infty, \end{aligned}$$

we obtain the bound $|\phi_{t-1}^N(H)/\phi_{t-1}^N(\vartheta_t)| \leq 2|\omega_t|_\infty |\mathcal{L}_{t,T}(\cdot, \mathbf{1})|_\infty$ which verifies condition (I). Now, since $t - 1 \leq T - 1$ and $\phi_{t-1}(H) = 0$, the induction assumption implies that conditions (II) and (III) are satisfied for $|b'_N - b'|$ and $|a'_N|$. Hence, Lemma 4 shows that

$$(45) \quad \mathbb{P} \left[\left| \mathbb{E} \left[N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1}) \mid \mathcal{F}_{t-1}^N \right] - b \right| > \varepsilon \right] \leq B e^{-CN\varepsilon^2}.$$

Finally, (43) and (45) ensure that condition (II) in Lemma 4 is satisfied and an exponential deviation inequality for (41) follows. The proof of (38) is complete. The last statement (39) of the theorem is a consequence of (43) and (45). \square

The exponential inequality of Theorem 5 may be more or less immediately extended to the FFBSi estimator.

COROLLARY 6. *Under the assumptions of Theorem 5 there exist constants $0 < B$ and $C < \infty$ (depending on T) such that for all $N, \varepsilon > 0$, and all measurable functions h ,*

$$(46) \quad \mathbb{P}[|\tilde{\phi}_{0:T|T}^N(h) - \phi_{0:T|T}(h)| \geq \varepsilon] \leq B e^{-CN\varepsilon^2/\text{osc}^2(h)},$$

where $\tilde{\phi}_{0:T|T}^N(h)$ is defined in (14).

PROOF. Using (13) and the definition of $\tilde{\phi}_{s:T|T}^N(h)$, we may write

$$\begin{aligned} &\tilde{\phi}_{0:T|T}^N(h) - \phi_{0:T|T}^N(h) \\ &= N^{-1} \sum_{\ell=1}^N [h(\xi_0^{J_\ell}, \dots, \xi_T^{J_\ell}) - \mathbb{E}[h(\xi_0^{J_0}, \dots, \xi_T^{J_T}) | \mathcal{F}_T^N]], \end{aligned}$$

which implies (46) by the Hoeffding inequality and (38). \square

3.2. Asymptotic normality. We now extend the theoretical analysis of the forward-filtering backward-smoothing estimator (6) to a CLT. Consider the following mild assumption on the proposal distribution.

ASSUMPTION 3. $|m|_\infty < \infty$ and $\sup_{0 \leq t \leq T} |p_t|_\infty < \infty$.

CLTs for interacting particle models have been established in [9, 12, 15]; the application to these results to auxiliary particle filters is presented in [25] and [16], Theorem 3.2. Here, we base our proof on techniques developed in [15] (extending [6] and [30]). As noted in the previous section, it turns out crucial that $G_{t,T}^N(\cdot, h)$ converges to a deterministic function as $N \rightarrow \infty$. This convergence is stated in the following lemma.

LEMMA 7. *Assume Assumptions 1–3. Then, for any $h \in \mathcal{F}_b(\mathbb{X})$ and $x \in \mathbb{X}$,*

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathcal{L}_{t,T}^N(x, h) &= \mathcal{L}_{t,T}(x, h), & \mathbb{P}\text{-a.s.}, \\ \lim_{N \rightarrow \infty} G_{t,T}^N(x, h) &= G_{t,T}(x, h), & \mathbb{P}\text{-a.s.}, \end{aligned}$$

where $\mathcal{L}_{t,T}^N, \mathcal{L}_{t,T}, G_{t,T}^N$ and $G_{t,T}$ are defined in (24), (26), (29) and (30). Moreover, there exists a constant C (that may depend on t and T) such that for all $N \geq 1, \ell \in \{1, \dots, N\}$, and $h \in \mathcal{F}_b(\mathbb{X})$,

$$|\omega_t^\ell G_{t,T}(\xi_t^\ell, h)| \leq |\omega_t|_\infty |G_{t,T}(\xi_t^\ell, h)| \leq C \text{osc}(h), \quad \mathbb{P}\text{-a.s.}$$

PROOF OF LEMMA 7. Let $h \in \mathcal{F}_b(\mathbb{X})$ and $x_t \in \mathbb{X}$. By plugging (1) with $\eta = \phi_{t-1}^N$ into the definition (24) of $\mathcal{L}_{t,T}^N(x_t, h)$, we obtain immediately

$$\begin{aligned} \mathcal{L}_{t,T}^N(x_t, h) &= \frac{\int \cdots \int \phi_{t-1}^N(dx_{t-1}) \prod_{u=0}^{t-2} \mathbf{B}_{\phi_u^N}(x_{u+1}, dx_u) m(x_{t-1}, x_t) L_{t,T}(x_{0:t}, h)}{\int \phi_{t-1}^N(dx_{t-1}) m(x_{t-1}, x_t)} \\ &= \frac{\phi_{0:t-1|t-1}^N[H([\cdot, x_t])]}{\phi_{t-1}^N[m(\cdot, x_t)]} \quad \text{with } H(x_{0:t}) \stackrel{\text{def}}{=} m(x_{t-1}, x_t) L_{t,T}(x_{0:t}, h). \end{aligned}$$

The convergence of $\mathcal{L}_{t,T}^N(\cdot, h)$ follows from Theorem 5. The proof of the convergence of $G_{t,T}^N(\cdot, h)$ follows the same lines. Finally, the final statement of the lemma is derived from Lemma 3 and the almost sure convergence of $G_{t,T}^N(\cdot, h)$ to $G_{t,T}(\cdot, h)$. \square

Now, we may state the CLT with an asymptotic variance given by a finite sum of terms involving the limiting kernel $G_{t,T}$.

THEOREM 8. Assume Assumptions 1–3. Then, for any $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$,

$$(47) \quad \sqrt{N}(\phi_{0:T|T}^N(h) - \phi_{0:T|T}(h)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma_{0:T|T}[h])$$

with

$$(48) \quad \Gamma_{0:T|T}[h] \stackrel{\text{def}}{=} \frac{\rho_0[\omega_0^2(\cdot) G_{0,T}^2(\cdot, h)]}{\rho_0^2[\omega_0(\cdot) \mathcal{L}_{0,T}(\cdot, \mathbf{1})]} + \sum_{t=1}^T \frac{\phi_{t-1}[\nu_{t,T}(\cdot, h)] \phi_{t-1}(\vartheta_t)}{\phi_{t-1}^2[\mathcal{L}_{t-1,T}(\cdot, \mathbf{1})]},$$

$$(49) \quad \nu_{t,T}(\cdot, h) \stackrel{\text{def}}{=} \vartheta_t(\cdot) \int P_t(\cdot, dx) \omega_t^2(\cdot, x) G_{t,T}^2(x, h).$$

PROOF. Without loss of generality, we assume that $\phi_{0:T|T}(h) = 0$. We show that $\sqrt{N}\phi_{0:T|T}^N(h)$ may be expressed as

$$(50) \quad \sqrt{N}\phi_{0:T|T}^N(h) = \sum_{t=0}^T \frac{V_{t,T}^N(h)}{W_{t,T}^N},$$

where the sequence of random vectors $[V_{0,T}^N(h), \dots, V_{T,T}^N(h)]$ is asymptotically normal and $[W_{0,T}^N, \dots, W_{T,T}^N]$ converge in probability to a deterministic vector.

The proof of (47) then follows from Slutsky’s lemma. Actually, the decomposition (50) follows immediately from the backward decomposition (22) by setting, for $t \in \{1, \dots, T\}$,

$$\begin{aligned}
 V_{0,T}^N(h) &\stackrel{\text{def}}{=} N^{-1/2} \sum_{\ell=1}^N \frac{d\chi}{d\rho_0}(\xi_0^\ell) g_0(\xi_0^\ell) G_{0,T}(\xi_0^\ell, h), \\
 V_{t,T}^N(h) &\stackrel{\text{def}}{=} N^{-1/2} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, h), \\
 W_{0,T}^N &\stackrel{\text{def}}{=} N^{-1} \sum_{\ell=1}^N \frac{d\chi}{d\rho_0}(\xi_0^\ell) g_0(\xi_0^\ell) \mathcal{L}_{0,T}(\xi_0^\ell, \mathbf{1}), \\
 W_{t,T}^N &\stackrel{\text{def}}{=} N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1}).
 \end{aligned}$$

The convergence

$$\begin{aligned}
 W_{0,T}^N &\xrightarrow{P}_{N \rightarrow \infty} \chi[g_0(\cdot) \mathcal{L}_{0,T}(\cdot, \mathbf{1})], \\
 W_{t,T}^N &\xrightarrow{P}_{N \rightarrow \infty} \frac{\phi_{t-1}[\mathcal{L}_{t-1,T}(\cdot, \mathbf{1})]}{\phi_{t-1}(\vartheta_t)}
 \end{aligned}$$

of $[W_{0,T}^N, \dots, W_{T,T}^N]$ to a deterministic vector is established immediately using (39) and noting that the initial particles $(\xi_0^i)_{i=1}^N$ are i.i.d. We devote the rest of the proof to showing that the sequence of random vectors $[V_{0,T}^N(h), \dots, V_{T,T}^N(h)]$ is asymptotically normal. Proceeding recursively in time, we prove by induction over $t \in \{0, \dots, T\}$ (starting with $t = 0$) that $[V_{0,T}^N(h), \dots, V_{t,T}^N(h)]$ is asymptotically normal. More precisely, using the Cramér–Wold device, it is enough to show that for all scalars $(\alpha_0, \dots, \alpha_t) \in \mathbb{R}^{t+1}$,

$$(51) \quad \sum_{r=0}^t \alpha_r V_{r,T}^N(h) \xrightarrow{\mathcal{D}}_{N \rightarrow \infty} \mathcal{N}\left(0, \sum_{r=0}^t \alpha_r^2 \sigma_{r,T}^2[h]\right),$$

where, for $r \geq 1$,

$$\sigma_{0,T}^2[h] \stackrel{\text{def}}{=} \rho_0[\omega_0^2 G_{0,T}^2(\cdot, h)], \quad \sigma_{t,T}^2[h] \stackrel{\text{def}}{=} \frac{\phi_{t-1}[\nu_{t,T}(\cdot, h)]}{\phi_{t-1}(\vartheta_t)}.$$

The case $t = 0$ is elementary since the initial particles $\{\xi_0^i\}_{i=1}^N$ are i.i.d. Assume now that (51) holds for some $t - 1 \leq T$; for all scalars $(\alpha_1, \dots, \alpha_{t-1}) \in \mathbb{R}^{t-1}$,

$$(52) \quad \sum_{r=s}^{t-1} \alpha_r V_{r,T}^N(h) \xrightarrow{\mathcal{D}}_{N \rightarrow \infty} \mathcal{N}\left(0, \sum_{r=s}^{t-1} \alpha_r^2 \sigma_{r,T}^2[h]\right).$$

The sequence of random variable $V_{t,T}^N(h)$ may be expressed as an additive function of a triangular array of random variables,

$$V_{t,T}^N(h) = \sum_{\ell=1}^N U_{N,\ell}, \quad U_{N,\ell} \stackrel{\text{def}}{=} \omega_t^\ell G_{t,T}^N(\xi_t^\ell, h) / \sqrt{N},$$

where $G_{t,T}^N(x, h)$ is defined in (29). Lemma 3 implies that $\mathbb{E}[V_{t,T}^N(h) | \mathcal{F}_{t-1}^N] = 0$, yielding

$$\mathbb{E} \left[\sum_{r=0}^t \alpha_r V_{r,T}^N(h) \middle| \mathcal{F}_{t-1}^N \right] = \sum_{r=0}^{t-1} \alpha_r V_{r,T}^N(h) \xrightarrow{\mathcal{D}}_{N \rightarrow \infty} \mathcal{N} \left(0, \sum_{r=1}^{t-1} \alpha_r^2 \sigma_{r,T}^2[h] \right),$$

where the last limit follows by the induction assumption hypothesis (52). By [15], Theorem A.3, page 2360, as the random variables $\{U_{N,\ell}\}_{\ell=1}^N$ are centered and conditionally independent given \mathcal{F}_{t-1}^N , (51) holds provided that the asymptotic smallness condition

$$(53) \quad \sum_{\ell=1}^N \mathbb{E}[U_{N,\ell}^2 \mathbb{1}_{\{|U_{N,\ell}| \geq \varepsilon\}} | \mathcal{F}_{t-1}^N] \xrightarrow{P}_{N \rightarrow \infty} 0$$

holds for any $\varepsilon > 0$ and that the conditional variance converges:

$$(54) \quad \sum_{\ell=1}^N \mathbb{E}[U_{N,\ell}^2 | \mathcal{F}_{t-1}^N] \xrightarrow{P}_{N \rightarrow \infty} \sigma_{t,T}^2[h].$$

Lemma 3 implies that $|U_{N,\ell}| \leq C \text{osc}(h) / \sqrt{N}$, verifying immediately the asymptotic smallness condition (53). To conclude the proof, we thus only need to establish the convergence (54) of the asymptotic variance. Via Lemma 3 and straightforward computations, we conclude that

$$\begin{aligned} \sum_{\ell=1}^N \mathbb{E}[U_{N,\ell}^2 | \mathcal{F}_{t-1}^N] &= \mathbb{E}[(\omega_t^1 G_{t,T}^N(\xi_t^1, h))^2 | \mathcal{F}_{t-1}^N] \\ &= \int \sum_{\ell=1}^N \frac{\omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell) P_t(\xi_{t-1}^\ell, dx)}{\sum_{j=1}^N \omega_{t-1}^j \vartheta_t(\xi_{t-1}^j)} (\omega_t(\xi_{t-1}^\ell, x) G_{t,T}^N(x, h))^2 \\ (55) \quad &= \left(\frac{\Omega_{t-1}}{\sum_{j=1}^N \omega_{t-1}^j \vartheta_t(\xi_{t-1}^j)} \right) \left(\frac{1}{\Omega_{t-1}} \sum_{\ell=1}^N \omega_{t-1}^\ell v_{t,T}^N(\xi_{t-1}^\ell, h) \right) \\ &= \frac{\phi_{t-1}^N[v_{t,T}^N(\cdot, h)]}{\phi_{t-1}^N(\vartheta_t)}, \end{aligned}$$

where Ω_t is defined in (9) and

$$v_{t,T}^N(\cdot, h) \stackrel{\text{def}}{=} \vartheta_t(\cdot) \int P_t(\cdot, dx) \omega_t^2(\cdot, x) [G_{t,T}^N(x, h)]^2.$$

The denominator in on RHS of (55) converges evidently in probability to $\phi_{t-1}(\vartheta_t)$ by Theorem 5. The numerator is more complex since $v_{t,T}^N$ depends on $G_{t,T}^N$ whose definition involves all the approximations $\phi_{t-1}^N, \dots, \phi_0^N$ of the past filters. To obtain its convergence, note that, by Theorem 5, $\phi_{t-1}^N(v_{t,T}(\cdot, h)) \xrightarrow{\mathbb{P}} \phi_{t-1}(v_{t,T}(\cdot, h))$ as N tends to infinity; hence, it only remains to prove that

$$(56) \quad \phi_{t-1}^N[v_{t,T}^N(\cdot, h) - v_{t,T}(\cdot, h)] \xrightarrow{\mathbb{P}}_{N \rightarrow \infty} 0.$$

For that purpose, introduce the following notation: for all $x \in \mathbb{X}$,

$$A_N(x) \stackrel{\text{def}}{=} \phi_{t-1}^N[\vartheta_t(\cdot) p_t(\cdot, x) \omega_t^2(\cdot, x) |(G_{t,T}^N(x, h))^2 - G_{t,T}^2(x, h)|],$$

$$B_N(x) \stackrel{\text{def}}{=} \phi_{t-1}^N[\vartheta_t(\cdot) p_t(\cdot, x)].$$

Applying Fubini’s theorem,

$$(57) \quad \lim_{N \rightarrow \infty} \mathbb{E} \left[\int A_N(x) dx \right] = \lim_{N \rightarrow \infty} \int \mathbb{E}[A_N(x)] dx = 0,$$

where the last equality is due to the generalized Lebesgue convergence theorem [34], Proposition 18, page 270, with $f_N(x) = \mathbb{E}[A_N(x)]$ and $g_N(x) = 2C \text{osc}(h) \mathbb{E}[B_N(x)]$ provided that the following conditions hold:

- (i) for any $x \in \mathbb{X}$, $\mathbb{E}[A_N(x)] \leq 2C^2 \text{osc}^2(h) \mathbb{E}[B_N(x)]$,
- (ii) for any $x \in \mathbb{X}$, $\lim_{N \rightarrow \infty} \mathbb{E}[A_N(x)] = 0$, \mathbb{P} -a.s.,
- (iii) $\lim_{N \rightarrow \infty} \int \mathbb{E}[B_N(x)] dx = \int \lim_{N \rightarrow \infty} \mathbb{E}[B_N(x)] dx$.

Proof of (i). The bound follows directly from Lemmas 7 and 3.

Proof of (ii). Using again Lemmas 7 and 3, for any $x \in \mathbb{X}$,

$$A_N(x) \leq 2C^2 |\vartheta_t|_\infty |p_t|_\infty \text{osc}^2(h),$$

$$\limsup_{N \rightarrow \infty} A_N(x) \leq |\vartheta_t p_t \omega_t^2|_\infty \limsup_{N \rightarrow \infty} |(G_{t,T}^N(x, h))^2 - G_{t,T}^2(x, h)| = 0, \quad \mathbb{P}\text{-a.s.}$$

These two inequalities combined with $A_N(x) \geq 0$ allow for applying the Lebesgue dominated convergence theorem, verifying condition (ii).

Proof of (iii). We have

$$\begin{aligned} \lim_{N \rightarrow \infty} \int \mathbb{E}[B_N(x)] dx &\stackrel{(a)}{=} \lim_{N \rightarrow \infty} \mathbb{E} \left[\phi_{t-1}^N \left(\vartheta_t(\cdot) \int p_t(\cdot, x) dx \right) \right] \\ &\stackrel{(b)}{=} \phi_{t-1}(\vartheta_t) \stackrel{(c)}{=} \int \phi_{t-1}(\vartheta_t(\cdot) p_t(\cdot, x)) dx \\ &\stackrel{(d)}{=} \int \lim_{N \rightarrow \infty} \mathbb{E}[\phi_{t-1}^N(\vartheta_t(\cdot) p_t(\cdot, x))] dx \\ &= \int \lim_{N \rightarrow \infty} \mathbb{E}[B_N(x)] dx, \end{aligned}$$

where (a) and (c) are consequences of Fubini’s theorem and (b) and (d) follows from the L^1 -convergence of $\phi_t^N(h)$ to $\phi_t(h)$ (see Theorem 5) with $h(\cdot) = \vartheta_t(\cdot)$ and $h(\cdot) = \vartheta_t(\cdot) p_t(\cdot, x)$.

Thus, (57) holds, yielding that $\int A_N(x) dx \xrightarrow{P} 0$ as N tends to infinity. This in turn implies (56) via the inequality

$$|\phi_{t-1}^N[v_{t,T}^N(\cdot, h) - v_{t,T}(\cdot, h)]| \leq \int A_N(x) dx.$$

This establishes (51) and therefore completes the proof. \square

The weak convergence of $\sqrt{N}(\phi_{0:T|T}^N(h) - \phi_{0:T|T}(h))$ for the FFBS algorithm implies more or less immediately the one of $\sqrt{N}(\tilde{\phi}_{0:T|T}^N(h) - \phi_{0:T|T}(h))$ for the FFBSi algorithm.

COROLLARY 9. *Under the assumptions of Theorem 8,*

$$(58) \quad \begin{aligned} &\sqrt{N}(\tilde{\phi}_{0:T|T}^N(h) - \phi_{0:T|T}(h)) \\ &\xrightarrow{D} \mathcal{N}(0, \phi_{0:T|T}^2[h - \phi_{0:T|T}(h)] + \Gamma_{0:T|T}[h - \phi_{0:T|T}(h)]). \end{aligned}$$

PROOF. Using (13) and the definition of $\tilde{\phi}_{0:T|T}^N(h)$, we may write

$$\begin{aligned} &\sqrt{N}(\tilde{\phi}_{0:T|T}^N(h) - \phi_{0:T|T}(h)) \\ &= N^{-1/2} \sum_{\ell=1}^N [h(\xi_0^{J_\ell^0}, \dots, \xi_T^{J_\ell^T}) - \mathbb{E}[h(\xi_0^{J_0}, \dots, \xi_T^{J_T}) | \mathcal{F}_T^N]] \\ &\quad + \sqrt{N}(\phi_{0:T|T}^N(h) - \phi_{0:T|T}(h)). \end{aligned}$$

Note that since $\{J_{0:T}^\ell\}_{\ell=1}^N$ are i.i.d. conditional on \mathcal{F}_T^N , (58) follows from (47) and direct application of [15], Theorem A.3, page 2360, by noting that

$$\begin{aligned} &N^{-1} \sum_{\ell=1}^N \mathbb{E}[\{h(\xi_0^{J_\ell^0}, \dots, \xi_T^{J_\ell^T}) - \mathbb{E}[h(\xi_0^{J_0}, \dots, \xi_T^{J_T}) | \mathcal{F}_T^N]\}^2 | \mathcal{F}_T^N] \\ &= (\phi_{0:T|T}^N[h - \phi_{0:T|T}^N(h)])^2 \xrightarrow{P} (\phi_{0:T|T}[h - \phi_{0:T|T}(h)])^2. \quad \square \end{aligned}$$

4. Time uniform bounds. Most often, it is not required to compute the joint smoothing distribution but rather the marginal smoothing distributions $\phi_{s|T}$. Considering (8) for a function h that depends on the component x_s only, we obtain particle approximations of the marginal smoothing distributions by associating the

set $\{\xi_s^j\}_{j=1}^N$ of particles with weights obtained by marginalizing the joint smoothing weights according to

$$\omega_{s|T}^{i_s} = \sum_{i_{s+1}=1}^N \cdots \sum_{i_T=1}^N \prod_{u=s+1}^t \frac{\omega_{u-1}^{i_{u-1}} m(\xi_{u-1}^{i_{u-1}}, \xi_u^{i_u})}{\sum_{\ell=1}^N \omega_{u-1}^\ell m(\xi_{u-1}^\ell, \xi_u^{i_u})} \frac{\omega_T^{i_T}}{\Omega_T}.$$

It is easily seen that these marginal weights may be recursively updated backward in time as

$$(59) \quad \omega_{s|T}^i = \sum_{j=1}^N \frac{\omega_s^j m(\xi_s^i, \xi_{s+1}^j)}{\sum_{\ell=1}^N \omega_s^\ell m(\xi_s^\ell, \xi_{s+1}^j)} \omega_{s+1|T}^j.$$

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 type of mixing condition on the Markov transition kernel; see [5] and the references therein. For simplicity, we consider elementary but strong conditions which are similar to the ones used in [9], Chapter 7.4, or [3], Chapter 4; these conditions, which points to applications where the state space \mathbb{X} is compact, can be relaxed, but at the expense of many technical difficulties [4, 37, 38, 40].

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

$$(60) \quad \sigma_- \leq m(x, x') \leq \sigma_+.$$

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

$$(61) \quad \inf_{x \in \mathbb{X}} \int M(x, dx')g_t(x') \geq c_- > 0.$$

Assumption 4 implies that $\nu(\mathbb{X}) < \infty$; in the sequel, we will consider without loss of generality that $\nu(\mathbb{X}) = 1$. Note also that, under Assumption 4, the average number of simulations required in the accept–reject mechanism per sample of the FFBSi algorithm is bounded by σ_+/σ_- .

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 and, under the same assumptions, that the variance of the CLT is uniformly bounded in time.

For obtaining these results, we will need upper-bounds on $G_{t,T}^N$ and $G_{t,T}$ that are more precise than the ones stated in Lemmas 3 and 7. For any function $h \in \mathcal{F}_b(\mathbb{X})$ and $s \leq T$, define the extension $\Pi_{s,T}h \in \mathcal{F}_b(\mathbb{X}^{T+1})$ of h to \mathbb{X}^{T+1} by

$$(62) \quad \Pi_{s,T}h(x_{0:T}) \stackrel{\text{def}}{=} h(x_s), \quad x_{0:T} \in \mathbb{X}^{T+1}.$$

LEMMA 10. Assume that Assumptions 1–4 hold with $T = \infty$. Let $s \leq T$. Then, for all $t, T, N \geq 1$, and $h \in \mathcal{F}_b(\mathbb{X})$,

$$(63) \quad |G_{t,T}^N(\cdot, \Pi_{s,Th})|_\infty \leq \rho^{|t-s|} \text{osc}(h) |\mathcal{L}_{t,T}(\cdot, \mathbf{1})|_\infty,$$

where $\mathcal{L}_{t,T}$ is defined in (26) and

$$(64) \quad \rho = 1 - \frac{\sigma_-}{\sigma_+}.$$

Moreover, for all $t, T \geq 1$, and $h \in \mathcal{F}_b(\mathbb{X})$,

$$(65) \quad |G_{t,T}(\cdot, \Pi_{s,Th})|_\infty \leq \rho^{|t-s|} \text{osc}(h) |\mathcal{L}_{t,T}(\cdot, \mathbf{1})|_\infty.$$

PROOF. Using (27) and (29),

$$(66) \quad \frac{G_{t,T}^N(x, \Pi_{s,Th})}{\mathcal{L}_{t,T}(x, \mathbf{1})} = \frac{\mathcal{L}_{t,T}^N(x, \Pi_{s,Th})}{\mathcal{L}_{t,T}^N(x, \mathbf{1})} - \frac{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N(\cdot, \Pi_{s,Th})]}{\phi_{t-1}^N[\mathcal{L}_{t-1,T}^N(\cdot, \mathbf{1})]}.$$

To prove (63), we will rewrite (66) and obtain an exponential bound by either using ergodicity properties of the ‘‘a posteriori’’ chain (when $t \leq s$), or by using ergodicity properties of the backward kernel (when $t > s$).

Assume first that $t \leq s$. The quantity $L_{t,T}(x_{0:t}, \Pi_{s,Th})$ does not depend on $x_{0:t-1}$ so that by (24) and definition (20) of $L_{t,T}$,

$$(67) \quad \begin{aligned} \mathcal{L}_{t,T}^N(x_t, \Pi_{s,Th}) &= L_{t,T}(x_{0:t}, \Pi_{s,Th}) \\ &= \int \cdots \int \left(\prod_{u=t+1}^T M(x_{u-1}, dx_u) g_u(x_u) \right) h(x_s) \\ &= \mathcal{L}_{t,T}(x_t, \Pi_{s,Th}). \end{aligned}$$

Now, by construction, for any $t \leq s$,

$$(68) \quad \mathcal{L}_{t-1,T}(x_{t-1}, \Pi_{s,Th}) = \int M(x_{t-1}, dx_t) g_t(x_t) \mathcal{L}_{t,T}(x_t, \Pi_{s,Th}).$$

The relations (66), (67) and (68) imply that

$$(69) \quad \frac{G_{t,T}^N(x, \Pi_{s,Th})}{\mathcal{L}_{t,T}(x, \mathbf{1})} = \frac{\mu[\mathcal{L}_{t,T}(\cdot, \Pi_{s,Th})]}{\mu[\mathcal{L}_{t,T}(\cdot, \mathbf{1})]} - \frac{\mu'[\mathcal{L}_{t,T}(\cdot, \Pi_{s,Th})]}{\mu'[\mathcal{L}_{t,T}(\cdot, \mathbf{1})]},$$

where $\mu \stackrel{\text{def}}{=} \delta_x$ and μ' is the nonnegative finite measure defined by

$$\mu'(A) \stackrel{\text{def}}{=} \iint \phi_{t-1}^N(dx_{t-1}) M(x_{t-1}, dx_t) g_t(x_t) \mathbf{1}_A(x_t), \quad A \in \mathcal{B}(\mathbb{X}).$$

Now, for any finite measure μ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, the quantity

$$\begin{aligned} & \frac{\mu[\mathcal{L}_{t,T}(\cdot, \Pi_{s,T}h)]}{\mu[\mathcal{L}_{t,T}(\cdot, \mathbf{1})]} \\ &= \frac{\int \cdots \int \mu(dx_t) \prod_{u=t+1}^T M(x_{u-1}, dx_u) g_u(x_u) h(x_s)}{\int \cdots \int \mu(dx_t) \prod_{u=t+1}^T M(x_{u-1}, dx_u) g_u(x_u)} \\ &= \frac{\int \cdots \int \mu(dx_t) \prod_{u=t+1}^s M(x_{u-1}, dx_u) g_u(x_u) h(x_s) \mathcal{L}_{s,T}(x_s, \mathbf{1})}{\int \cdots \int \mu(dx_t) \prod_{u=t+1}^s M(x_{u-1}, dx_u) g_u(x_u) \mathcal{L}_{s,T}(x_s, \mathbf{1})} \end{aligned}$$

may be seen as the expectation of $h(X_s)$ conditionally on $Y_{t:T}$, where X_t is distributed according to $A \mapsto \mu(A)/\mu(\mathbb{X})$. Under the strong mixing condition (Assumption 4), it is shown in [12] (see also [9]) that, for any $t \leq s \leq T$, any finite measure μ and μ' on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, any function $h \in \mathcal{F}_b(\mathbb{X})$, that

$$\begin{aligned} & \left| \frac{\int \cdots \int \mu(dx_t) \prod_{u=t+1}^s M(x_{u-1}, dx_u) g_u(x_u) h(x_s) \mathcal{L}_{s,T}(x_s, \mathbf{1})}{\int \cdots \int \mu(dx_t) \prod_{u=t+1}^s M(x_{u-1}, dx_u) g_u(x_u) \mathcal{L}_{s,T}(x_s, \mathbf{1})} \right. \\ & \quad \left. - \frac{\int \cdots \int \mu'(dx_t) \prod_{u=t+1}^s M(x_{u-1}, dx_u) g_u(x_u) h(x_s) \mathcal{L}_{s,T}(x_s, \mathbf{1})}{\int \cdots \int \mu'(dx_t) \prod_{u=t+1}^s M(x_{u-1}, dx_u) g_u(x_u) \mathcal{L}_{s,T}(x_s, \mathbf{1})} \right| \\ & \leq \rho^{s-t} \text{osc}(h), \end{aligned}$$

where ρ is defined in (64). This shows (63) when t is smaller than s .

Consider now the case $s < t \leq T$. By definition,

$$\begin{aligned} (70) \quad \mathcal{L}_{t,T}^N(x_t, \Pi_{s,T}h) &= \int \cdots \int L_{t,T}(x_{0:t}, \Pi_{s,T}h) \prod_{u=s+1}^t B_{\phi_{u-1}^N}(x_u, dx_{u-1}) \\ &= \int \cdots \int \mathcal{L}_{t,T}(x_t, \mathbf{1}) \prod_{u=s+1}^t B_{\phi_{u-1}^N}(x_u, dx_{u-1}) h(x_s), \end{aligned}$$

where the last expression is obtained from the following equality, valid for $s < t$:

$$\begin{aligned} L_{t,T}(x_{0:t}, \Pi_{s,T}h) &= h(x_s) \int \cdots \int \prod_{u=t+1}^T M(x_{u-1}, dx_u) g_u(x_u) \\ &= h(x_s) \mathcal{L}_{t,T}(x_t, \mathbf{1}). \end{aligned}$$

Moreover, combining (33) and (70),

$$\begin{aligned} & \phi_{t-1}^N[\mathcal{L}_{t-1,T}^N(\cdot, \Pi_{s,T}h)] \\ &= \int \cdots \int \phi_{t-1}^N(du_{t-1}) M(u_{t-1}, dx_t) g_t(x_t) \mathcal{L}_{t,T}^N(x_t, \Pi_{s,T}h) \end{aligned}$$

$$= \int \cdots \int \phi_{t-1}^N(du_{t-1})M(u_{t-1}, dx_t)g_t(x_t)\mathcal{L}_{t,T}(x_t, \mathbf{1}) \\ \times \prod_{u=s+1}^t B_{\phi_{u-1}^N}(x_u, dx_{u-1})h(x_s).$$

By plugging this expression and (70) into (66), we obtain

$$\frac{G_{t,T}^N(x, \Pi_{s,T}h)}{\mathcal{L}_{t,T}(x, \mathbf{1})} = \int \cdots \int \left\{ \frac{\mu(dx_t)}{\mu(\mathbb{X})} - \frac{\mu'(dx_t)}{\mu'(\mathbb{X})} \right\} \prod_{u=s+1}^t B_{\phi_{u-1}^N}(x_u, dx_{u-1})h(x_s),$$

with $\mu(dx_t) = \delta_x(dx_t)\mathcal{L}_{t,T}(x_t, \mathbf{1})$ and μ' being the nonnegative measure defined by

$$\mu'(A) = \int \phi_{t-1}^N[m(\cdot, x_t)]g_t(x_t)\mathcal{L}_{t,T}(x_t, \mathbf{1})\mathbf{1}_A(x_t) dx_t.$$

Under the uniform ergodicity condition (Assumption 4) it holds, for any probability measure η on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, and any $A \in \mathcal{B}(\mathbb{X})$,

$$B_\eta(x, A) = \frac{\int_A \eta(dx')m(x', x)}{\int \eta(dx')m(x', x)} \geq \frac{\sigma_-}{\sigma_+} \eta(A);$$

thus, the transition kernel B_η is uniformly Doeblin with minorizing constant σ_-/σ_+ and the proof of (63) for $s < t \leq T$ follows. The last statement of the Lemma follows from (63) and the almost-sure convergence

$$\lim_{N \rightarrow \infty} G_{t,T}^N(x, h) = G_{t,T}(x, h), \quad \mathbb{P}\text{-a.s.},$$

for all $x \in \mathbb{X}$, which was established in Lemma 7. \square

4.1. *A time uniform exponential deviation inequality.* Under the strong mixing Assumption 4, a time uniform deviation inequality for the *marginal smoothing* approximation can be derived using the exponentially decreasing bound on the quantity $G_{t,T}^N$ obtained in Lemma 10.

THEOREM 11. *Assume Assumptions 1–4 hold with $T = \infty$. Then, there exist constants $0 \leq B, C < \infty$ such that for all integers N, s, T , with $s \leq T$, and for all $\varepsilon > 0$,*

$$(71) \quad \mathbb{P}[|\phi_{s|T}^N(h) - \phi_{s|T}(h)| \geq \varepsilon] \leq B e^{-CN\varepsilon^2/\text{osc}^2(h)},$$

$$(72) \quad \mathbb{P}[|\tilde{\phi}_{s|T}^N(h) - \phi_{s|T}(h)| \geq \varepsilon] \leq B e^{-CN\varepsilon^2/\text{osc}^2(h)},$$

where $\phi_{s|T}^N(h)$ and $\tilde{\phi}_{s|T}^N(h)$ are defined in (6) and (14).

Letting $s = T$ in Theorem 11 provides, as a special case, the (already known) time uniform deviation inequality for the *filter* approximation; however, the novelty of the bounds obtained here is that these confirm the stability of the FFBSm and FFBSi marginal smoothing approximations also when s is fixed and T tends to infinity (see [9] for further discussion).

PROOF OF THEOREM 11. Combining (27) with the definition (20) and Assumption 4 yields, for all $x \in \mathbb{X}$,

$$(73) \quad \frac{\sigma_-}{\sigma_+} \leq \frac{\mathcal{L}_{t,T}(x, \mathbf{1})}{|\mathcal{L}_{t,T}(\cdot, \mathbf{1})|_\infty} \leq 1.$$

Let $h \in \mathcal{F}_b(\mathbb{X}^{T+1})$ and assume without loss of generality that $\phi_{0:T|T}(h) = 0$. Then, (21) implies that $\phi_0[L_{0,T}(\cdot, h)] = 0$ and the first term of the decomposition (22) thus becomes

$$(74) \quad \frac{\phi_0^N[L_{0,T}(\cdot, h)]}{\phi_0^N[L_{0,T}(\cdot, \mathbf{1})]} = \frac{N^{-1} \sum_{i=0}^N \frac{d\chi}{d\rho_0}(\xi_0^i) g_0(\xi_0^i) L_{0,T}(\xi_0^i, h)}{N^{-1} \sum_{\ell=0}^N \frac{d\chi}{d\rho_0}(\xi_0^\ell) g_0(\xi_0^\ell) L_{0,T}(\xi_0^\ell, \mathbf{1})},$$

where $(\xi_0^\ell)_{\ell=1}^N$ are i.i.d. random variables with distribution ρ_0 . Noting that $L_{0,T} = \mathcal{L}_{0,T}$ we obtain an exponential deviation inequality for (74) by applying Lemma 4 with

$$\begin{cases} a_N = N^{-1} \sum_{i=0}^N \frac{d\chi}{d\rho_0}(\xi_0^i) g_0(\xi_0^i) \mathcal{L}_{0,T}(\xi_0^i, h) / |\mathcal{L}_{0,T}(\cdot, h)|_\infty, \\ b_N = N^{-1} \sum_{i=0}^N \frac{d\chi}{d\rho_0}(\xi_0^i) g_0(\xi_0^i) \mathcal{L}_{0,T}(\xi_0^i, \mathbf{1}) / |\mathcal{L}_{0,T}(\cdot, h)|_\infty, \\ b = \chi[g_0(\cdot) \mathcal{L}_{0,T}(\cdot, \mathbf{1})] / |\mathcal{L}_{0,T}(\cdot, h)|_\infty, \\ \beta = \chi(g_0) \sigma_- / \sigma_+. \end{cases}$$

Here, condition (I) is trivially satisfied and conditions (II) and (III) follow from the Hoeffding inequality for i.i.d. variables.

According to (22) and (28), it is now required, for any $1 \leq t \leq T$, to derive an exponential inequality for

$$A_{t,T}^N \stackrel{\text{def}}{=} \frac{N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, \Pi_{s,Th})}{N^{-1} \sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,T}(\xi_t^\ell, \mathbf{1})}.$$

Note first that, using (73), we have

$$|A_{t,T}^N| \leq \left(\frac{\sigma_+}{\sigma_-} \right) \frac{N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, \Pi_{s,Th}) / |\mathcal{L}_{t,T}(\cdot, \mathbf{1})|_\infty}{N^{-1} \sum_{\ell=1}^N \omega_t^\ell}.$$

We use again Lemma 4 with

$$\begin{cases} a_N = N^{-1} \sum_{\ell=1}^N \omega_t^\ell G_{t,T}^N(\xi_t^\ell, \Pi_{s,T}h) / |\mathcal{L}_{t,T}(\cdot, \mathbf{1})|_\infty, \\ b_N = N^{-1} \sum_{\ell=1}^N \omega_t^\ell, \\ b = \mathbb{E}[\omega_t^1 | \mathcal{F}_{t-1}^N] = \phi_{t-1}^N[M(\cdot, g_t)] / \phi_{t-1}^N(\vartheta_t), \\ \beta = c_- / |\vartheta_t|_\infty. \end{cases}$$

Assumption 4 shows that $b \geq \beta$ and Lemma 10 shows that $|a_N/b_N| \leq M \stackrel{\text{def}}{=} \rho^{|t-s|} \text{osc}(h)$, where ρ is defined in (64). Therefore, condition (I) of Lemma 4 is satisfied and the Hoeffding inequality gives

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

establishing condition (II) in Lemma 4. Finally, Lemma 10 and the Hoeffding inequality imply that

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

Lemma 4 therefore yields

$$\mathbb{P}\left(\left|\frac{a_N}{b_N}\right| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{N\varepsilon^2 c_-^2}{2 \text{osc}^2(h) \rho^{2|t-s|} |\omega_t|_\infty^2 |\vartheta_t|_\infty^2}\right),$$

so that

$$\mathbb{P}(|A_{t,T}^N| \geq \varepsilon) \leq 2 \exp\left(-\frac{N\varepsilon^2 c_-^2 \sigma_-^2}{2 \text{osc}^2(h) \rho^{2|t-s|} |\omega_t|_\infty^2 |\vartheta_t|_\infty^2 \sigma_+^2}\right).$$

A time uniform exponential deviation inequality for $\sum_{t=1}^T A_{t,T}$ then follows from Lemma 13 and the proof is complete. \square

4.2. *A time uniform bound on the variance of the marginal smoothing distribution.* Analogous to the result obtained in the previous section, a time uniform bound on the asymptotic variance in the CLT for the *marginal smoothing* approximations can, again under the strong mixing Assumption 4, be easily obtained from the exponentially decreasing bound on $G_{t,T}$ stated and proved in Lemma 10 for the quantity.

THEOREM 12. *Assume Assumptions 1–4 hold with $T = \infty$. Then, for all $s \leq T$,*

$$\Gamma_{0:T|T}[\Pi_{s,T}h] \leq \left(\frac{\sigma_+}{\sigma_-} \left(1 \vee \sup_{t \geq 1} |\vartheta_t|_\infty \right) \sup_{t \geq 0} |\omega_t|_\infty \text{osc}(h) \right)^2 \frac{1 + \rho^2}{1 - \rho^2},$$

where $\Gamma_{0:T|T}$ is defined in (48).

In accordance with the results of the previous section, letting $s = T$ in the previous theorem provides a time uniform bound on the asymptotic variance for the filter approximation; nevertheless, as mentioned previously, the situation of interest for us is when s is fixed and T goes to infinity.

PROOF OF THEOREM 12. Combining (73) and (65) with $\rho_0(\omega_0) = 1$ yields

$$\frac{\rho_0(\omega_0^2(\cdot)G_{0,T}^2(\cdot, \Pi_{s,T}h))}{\rho_0^2[\omega_0(\cdot)\mathcal{L}_{0,T}(\cdot, \mathbf{1})]} \leq \left(\frac{\sigma_+}{\sigma_-} |\omega_0|_\infty \text{osc}(h) \rho^s \right)^2.$$

Moreover, by inserting, for any $0 < t \leq T$, the bound obtained in (65) into the expression (49) of $\nu_{t,T}$ we obtain

$$\frac{\phi_{t-1}(\nu_{t,T}(\cdot, \Pi_{s,T}h))\phi_{t-1}(\vartheta_t)}{\phi_{t-1}^2[\mathcal{L}_{t-1,T}(\cdot, \mathbf{1})]} \leq \left(\frac{\sigma_+}{\sigma_-} |\vartheta_t|_\infty |\omega_t|_\infty \text{osc}(h) \rho^{|t-s|} \right)^2.$$

Finally, plugging the two bounds above into (48) gives

$$\Gamma_{0:T|T}[\Pi_{s,T}h] \leq \left(\frac{\sigma_+}{\sigma_-} \left(1 \vee \sup_{t \geq 1} |\vartheta_t|_\infty \right) \sup_{t \geq 0} |\omega_t|_\infty \text{osc}(h) \right)^2 \left(\sum_{t=0}^{\infty} \rho^{2|t-s|} \right),$$

which completes the proof. \square

5. Proofs of Propositions 1 and 2. Having at hand the theory established in the previous sections, we are now ready to present the proofs of Propositions 1 and 2.

PROOF OF PROPOSITION 1. The average number of simulations required to sample J_s^ℓ conditionally on \mathcal{G}_{s+1}^N is $\sigma_+ \Omega_s / \sum_{u=1}^N \omega_s^u m(\xi_s^u, \xi_{s+1}^{J_s^\ell})$. Hence, the number of simulations Z_s^N required to sample $\{J_s^\ell\}_{\ell=1}^N$ has conditional expectation

$$\mathbb{E}[Z_s^N | \mathcal{G}_{s+1}^N] = \sum_{\ell=1}^N \frac{\sigma_+ \Omega_s}{\sum_{i=1}^N \omega_s^i m(\xi_s^i, \xi_{s+1}^{J_s^\ell})}.$$

We denote $\omega_{s|T}^i \stackrel{\text{def}}{=} \mathbb{P}[J_s^1 = i | \mathcal{F}_T^N]$ and $\omega_{s:s+1|T}^{\ell i} \stackrel{\text{def}}{=} \mathbb{P}[J_s^1 = \ell, J_{s+1}^1 = i | \mathcal{F}_T^N]$ and write

$$\begin{aligned} \mathbb{E}[Z_s^N | \mathcal{F}_T^N] &= \sum_{i=1}^N \omega_{s+1|T}^i \frac{\sigma_+ \Omega_s}{\sum_{j=1}^N \omega_s^j m(\xi_s^j, \xi_{s+1}^i)} \\ &= \sigma_+ \Omega_s \sum_{i=1}^N \sum_{\ell=1}^N \frac{\omega_{s+1|T}^i \omega_s^\ell m(\xi_s^\ell, \xi_{s+1}^i)}{\sum_{j=1}^N \omega_s^j m(\xi_s^j, \xi_{s+1}^i)} \times \frac{1}{\omega_s^\ell m(\xi_s^\ell, \xi_{s+1}^i)} \\ &= \sigma_+ \Omega_s \sum_{i=1}^N \sum_{\ell=1}^N \omega_{s:s+1|T}^{\ell i} \frac{1}{\omega_s^\ell m(\xi_s^\ell, \xi_{s+1}^i)}. \end{aligned}$$

For the bootstrap particle filter, $\omega_s^\ell \equiv g_s(\xi_s^\ell)$; Theorem 5 then implies that $\Omega_s / N \xrightarrow{P} N \rightarrow \infty \phi_{s|s-1}(g_s)$ and

$$\begin{aligned} \sum_{i=1}^N \sum_{\ell=1}^N \omega_{s:s+1|T}^{\ell i} \frac{1}{\omega_s^\ell m(\xi_s^\ell, \xi_{s+1}^i)} \\ \xrightarrow{P} N \rightarrow \infty \iint \phi_{s:s+1|T}(dx_{s:s+1}) \frac{1}{g_s(x_s) m(x_s, x_{s+1})}. \end{aligned}$$

Besides,

$$\begin{aligned} &\iint \phi_{s:s+1|T}(dx_{s:s+1}) \frac{1}{g_s(x_s) m(x_s, x_{s+1})} \\ &= \int \cdots \int \phi_{s|s-1}(dx_s) \frac{g_s(x_s) M(x_s, dx_{s+1})}{g_s(x_s) m(x_s, x_{s+1})} g_{s+1}(x_{s+1}) \\ &\quad \times \prod_{u=s+2}^T M(x_{u-1}, dx_u) g_u(x_u) \\ &= \frac{\int \cdots \int \phi_{s|s-1}(dx_s) g_s(x_s) \prod_{u=s+1}^T M(x_{u-1}, dx_u) g_u(x_u)}{\int \cdots \int \phi_{s|s-1}(dx_s) g_s(x_s) \prod_{u=s+1}^T M(x_{u-1}, dx_u) g_u(x_u)}. \end{aligned}$$

Similarly, in the fully adapted case we have $\omega_s^i \equiv 1$ for all $i \in \{1, \dots, N\}$; thus, $\Omega_s = N$ and

$$\begin{aligned} \sum_{i=1}^N \sum_{\ell=1}^N \omega_{s:s+1|T}^{\ell i} \frac{1}{\omega_s^\ell m(\xi_s^\ell, \xi_{s+1}^i)} \\ \xrightarrow{P} N \rightarrow \infty \iint \phi_{s:s+1|T}(dx_{s:s+1}) \frac{1}{m(x_s, x_{s+1})} \end{aligned}$$

$$= \frac{\int \cdots \int g_{s+1}(x_{s+1}) dx_{s+1} \prod_{u=s+2}^T \int M(x_{u-1}, dx_u) g_u(x_u)}{\int \cdots \int \phi_s(dx_s) \prod_{u=s+1}^T M(x_{u-1}, dx_u) g_u(x_u)}.$$

In both cases, the numerator can be bounded from above by

$$\frac{\sigma_+^{T-s-1} \prod_{u=s+1}^T \int g_u(x_u) dx_u}{\int \cdots \int \phi_{|s-1}(dx_s) g_s(x_s) \prod_{u=s+1}^T M(x_{u-1}, dx_u) g_u(x_u)}$$

if $\int g_u(x_u) dx_u < \infty$ for all $u \geq 0$. \square

PROOF OF PROPOSITION 2. Fix a time step s of the algorithm and denote by C_s the number of elementary operations required for this step. For $k \in \{1, \dots, n\}$, let T_s^k be the number of times that k appears in list L at time s in the ‘while’ loop. Let also $N_s^u \stackrel{\text{def}}{=} \sum_{k=1}^N \mathbb{1}_{\{T_s^k \geq u\}}$ be the size of L (i.e., the value of n at line 6) after u iterations of the ‘while’ loop, with $N_s^0 \stackrel{\text{def}}{=} N$. Then, using Proposition 14 there exists a constant C such that

$$C_s \leq C \sum_{u=0}^{\infty} N_s^u \left(1 + \log \left(1 + \frac{N}{N_s^u} \right) \right).$$

As $n \rightarrow n(1 + \log(1 + N/n))$ is a concave, increasing function, it holds by Jensen’s inequality that

$$\mathbb{E}[C_s] \leq C \sum_{u=0}^{\infty} \mathbb{E}[N_s^u] \left(1 + \log \left(1 + \frac{N}{\mathbb{E}[N_s^u]} \right) \right).$$

Besides,

$$\mathbb{E}[N_s^u] = \sum_{k=1}^N \mathbb{P}(T_s^k \geq u) \leq N \left(1 - \frac{\sigma_-}{\sigma_+} \right)^u$$

as σ_-/σ_+ is a lower bound on the acceptance probability. Thus,

$$\mathbb{E}[C_s] \leq CN \sum_{u=0}^{\infty} \left(1 - \frac{\sigma_-}{\sigma_+} \right)^u \left(1 + \log \left(1 + \frac{1}{(1 - \sigma_-/\sigma_+)^u} \right) \right) \leq \frac{KN\sigma_+}{\sigma_-}. \quad \square$$

APPENDIX A: PROOF OF LEMMA 4

Write

$$\left| \frac{a_N}{b_N} \right| \leq b^{-1} \left| \frac{a_N}{b_N} \right| |b - b_N| + b^{-1} |a_N| \leq \beta^{-1} M |b - b_N| + \beta^{-1} |a_N|, \quad \mathbb{P}\text{-a.s.}$$

Thus,

$$\left\{ \left| \frac{a_N}{b_N} \right| \geq \varepsilon \right\} \subseteq \left\{ |b - b_N| \geq \frac{\varepsilon\beta}{2M} \right\} \cup \left\{ |a_N| \geq \frac{\varepsilon\beta}{2} \right\},$$

from which the proof follows.

APPENDIX B: TECHNICAL RESULTS

LEMMA 13. *Let $\{Y_{n,i}\}_{i=1}^n$ be a triangular array of random variables such that there exist constants $B > 0$, $C > 0$, and ρ with $0 < \rho < 1$ satisfying, for all n , $i \in \{1, \dots, n\}$, and $\varepsilon > 0$,*

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

Then, there exist constants $\bar{B} > 0$ and $\bar{C} > 0$ such that, for any n and $\varepsilon > 0$,

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

PROOF. Set $S \stackrel{\text{def}}{=} \sum_{\ell=1}^{\infty} \sqrt{\ell}\rho^\ell$; one easily concludes that

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

Set $\varepsilon_0 > 0$. The proof follows by noting that, for any $\varepsilon \geq \varepsilon_0$,

$$\sum_{i=1}^n \exp(-CS^{-1}\varepsilon^2i) \leq \exp(CS^{-1}\varepsilon_0^2) \exp(-CS^{-1}\varepsilon^2)/(1 - \exp(CS^{-1}\varepsilon_0^2)). \quad \square$$

B.1. Description of the sampling procedure. In this section, we describe and analyze an efficient multinomial sampling procedure, detailed in Algorithm 2. Given a probability distribution (p_1, \dots, p_N) on the set $\{1, \dots, N\}$, it returns a sample of size n of that distribution. Compared to the procedure described in Section 7.4.1 in [3], its main virtue is to be efficient for both large and small samples sizes: if $n = 1$, the complexity is $O(\log(N))$, while if $n = N$, the complexity is $O(N)$.

PROPOSITION 14. *The number of elementary operations required by Algorithm 2 is $O(n + n \log(1 + N/n))$.*

PROOF. The order statistics at line 5 and the permutation at line 6 can be sampled using $O(n)$ operations; see [13], Chapter V and XIII. For each value of k between 1 and n , denote by G_k the number of times lines 11–13 are executed. Observe that line 18 is executed the same number of times, and thus the number of elementary operations required by call to Algorithm 2 is $O(n + \sum_{k=1}^n G_k)$. But the value of l is increased during iteration k by at least $2^{G_k} - 1$, and as the final value of l is at most equal to N , it holds that

$$\sum_{k=1}^n 2^{G_k} \leq N + n.$$

Algorithm 2 Multinomial sampling

```

1:  $q_1 \leftarrow p_1$ 
2: for  $k$  from 1 to  $N$  do
3:    $q_k \leftarrow q_{k-1} + p_k$ 
4: end for
5: sample an order statistics  $U_{(1)}, \dots, U_{(n)}$  of an i.i.d. uniform distribution
6: uniformly sample a permutation  $\sigma$  on  $\{1, \dots, n\}$ 
7:  $l \leftarrow 0, r \leftarrow 1$ 
8: for  $k$  from 1 to  $n$  do
9:    $d \leftarrow 1$ 
10:  while  $U_{(k)} \geq q_r$  do
11:     $l \leftarrow r$ 
12:     $r \leftarrow \min(r + 2^d, N)$ 
13:     $d \leftarrow d + 1$ 
14:  end while
15:  while  $r - l > 1$  do
16:     $m \leftarrow \lfloor (l + r)/2 \rfloor$ 
17:    if  $U_{(k)} \geq q_m$  then
18:       $l \leftarrow m$ 
19:    else
20:       $r \leftarrow m$ 
21:    end if
22:  end while
23:   $I_{\sigma(k)} \leftarrow r$ 
24: end for

```

By convexity,

$$\exp\left(\frac{\log(2)}{n} \sum_{k=1}^n G_k\right) \leq \frac{1}{n} \sum_{k=1}^n 2^{G_k} \leq 1 + \frac{N}{n},$$

which implies that

$$\sum_{k=1}^n G_k \leq n \log\left(1 + \frac{N}{n}\right) / \log(2). \quad \square$$

REFERENCES

- [1] BAREMBRUCH, S., GARIVIER, A. and MOULINES, E. (2009). On approximate maximum-likelihood methods for blind identification: How to cope with the curse of dimensionality. *IEEE Trans. Signal Process.* **57** 4247–4259. [MR2683174](#)
- [2] BRIERS, M., DOUCET, A. and MASKELL, S. (2010). Smoothing algorithms for state-space models. *Ann. Inst. Statist. Math.* **62** 61–89. [MR2577439](#)
- [3] CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. [MR2159833](#)

- [4] CHIGANSKY, P. and LIPTSER, R. (2004). Stability of nonlinear filters in nonmixing case. *Ann. Appl. Probab.* **14** 2038–2056. [MR2099662](#)
- [5] CHIGANSKY, P., LIPTSER, R. and VAN HANDEL, R. (2011). Intrinsic methods in filter stability. In *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, Oxford.
- [6] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. [MR2153989](#)
- [7] CORNEBISE, J., MOULINES, É. and OLSSON, J. (2008). Adaptive methods for sequential importance sampling with application to state space models. *Statist. Comput.* **18** 461–480. [MR2461889](#)
- [8] DE JONG, P. and SHEPHARD, N. (1995). The simulation smoother for time series models. *Biometrika* **82** 339–350. [MR1354233](#)
- [9] DEL MORAL, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York. [MR2044973](#)
- [10] DEL MORAL, P. and DOUCET, A. (2009). Particle methods: An introduction with applications. Available at <http://hal.inria.fr/docs/00/40/39/17/PDF/RR-6991.pdf>.
- [11] DEL MORAL, P., DOUCET, A. and SINGH, S. (2010). A backward particle interpretation of Feynman–Kac formulae. *ESAIM M2AN* **44** 947–975.
- [12] DEL MORAL, P. and MICLO, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV. Lecture Notes in Math.* **1729** 1–145. Springer, Berlin. [MR1768060](#)
- [13] DEVROYE, L. (1986). *Nonuniform Random Variate Generation*. Springer, New York. [MR0836973](#)
- [14] DOUC, R., FORT, G., MOULINES, E. and PRIOURET, P. (2009). Forgetting the initial distribution for hidden Markov models. *Stochastic Process. Appl.* **119** 1235–1256. [MR2508572](#)
- [15] DOUC, R. and MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. [MR2458190](#)
- [16] DOUC, R., MOULINES, É. and OLSSON, J. (2009). Optimality of the auxiliary particle filter. *Probab. Math. Statist.* **29** 1–28. [MR2552996](#)
- [17] DOUCET, A., DE FREITAS, N. and GORDON, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York. [MR1847783](#)
- [18] DOUCET, A., GODSILL, S. and ANDRIEU, C. (2000). On sequential Monte-Carlo sampling methods for Bayesian filtering. *Statist. Comput.* **10** 197–208.
- [19] FEARNHEAD, P. (2005). Using random quasi-Monte-Carlo within particle filters, with application to financial time series. *J. Comput. Graph. Statist.* **14** 751–769. [MR2211365](#)
- [20] FEARNHEAD, P., WYNOLL, D. and TAWN, J. (2010). A sequential smoothing algorithm with linear computational cost. *Biometrika* **97** 447–464. [MR2650750](#)
- [21] GODSILL, S. J., DOUCET, A. and WEST, M. (2004). Monte Carlo smoothing for non-linear time series. *J. Amer. Statist. Assoc.* **50** 438–449.
- [22] 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.
- [23] HANDSCHIN, J. E. and MAYNE, D. Q. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *Internat. J. Control* (1) **9** 547–559. [MR0246490](#)
- [24] HÜRZELER, M. and KÜNSCH, H. R. (1998). Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.* **7** 175–193. [MR1649366](#)
- [25] JOHANSEN, A. M. and DOUCET, A. (2008). A note on auxiliary particle filters. *Statist. Probab. Lett.* **78** 1498–1504. [MR2528342](#)
- [26] KAILATH, T., SAYED, A. and HASSIBI, B. (2000). *Linear Estimation*. Prentice Hall, New York.

- [27] KITAGAWA, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5** 1–25. [MR1380850](#)
- [28] 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.
- [29] KLEPTSZYNA, M. L. and VERETENNIKOV, A. Y. (2008). On discrete time ergodic filters with wrong initial data. *Probab. Theory Related Fields* **141** 411–444. [MR2391160](#)
- [30] KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. [MR2211077](#)
- [31] LIU, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York. [MR1842342](#)
- [32] PITT, M. K. and SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. [MR1702328](#)
- [33] RISTIC, B., ARULAMPALAM, M. and GORDON, A. (2004). *Beyond Kalman Filters: Particle Filters for Target Tracking*. Artech House, Norwood, MA.
- [34] ROYDEN, H. L. (1988). *Real Analysis*, 3rd ed. Macmillan Co., New York. [MR1013117](#)
- [35] 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. Amer. Statist. Assoc.* **82** 543–546.
- [36] VAN HANDEL, R. (2007). Filtering, stability, and robustness. Ph.D. thesis, CalTech.
- [37] VAN HANDEL, R. (2008). Discrete time nonlinear filters with informative observations are stable. *Electron. Commun. Probab.* **13** 562–575. [MR2461532](#)
- [38] VAN HANDEL, R. (2009). Uniform observability of hidden Markov models and filter stability for unstable signals. *Ann. Appl. Probab.* **19** 1172–1199.
- [39] VAN HANDEL, R. (2009). The stability of conditional Markov processes and Markov chains in random environments. *Ann. Probab.* **37** 1876–1925. [MR2561436](#)
- [40] VAN HANDEL, R. (2009). Uniform time average consistency of Monte Carlo particle filters. *Stochastic Process. Appl.* **119** 3835–3861. [MR2552307](#)

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

A. GARIVIER
E. MOULINES
TÉLÉCOM PARISTECH
46 RUE BARRAULT
75634 PARIS, CEDEX 13
FRANCE
E-MAIL: aurelien.garivier@telecom-paristech.fr
eric.moulines@telecom-paristech.fr

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