



HAL
open science

Convergence of a Particle-based Approximation of the Block Online Expectation Maximization Algorithm

Sylvain Le Corff, Gersende Fort

► **To cite this version:**

Sylvain Le Corff, Gersende Fort. Convergence of a Particle-based Approximation of the Block Online Expectation Maximization Algorithm. 2012. hal-00638388v2

HAL Id: hal-00638388

<https://hal.science/hal-00638388v2>

Preprint submitted on 24 Feb 2012 (v2), last revised 29 May 2012 (v3)

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

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

Convergence of a Particle-based Approximation of the Block Online Expectation Maximization Algorithm

Sylvain Le Corff^{*†} and Gersende Fort[†]

February 24, 2012

Abstract

Online variants of the Expectation Maximization algorithm have recently been proposed to perform parameter inference with large data sets or data streams, in independent latent models and in hidden Markov models. Nevertheless, the convergence properties of these algorithms remain an open problem at least in the Hidden Markov case. [25] introduced a new online EM algorithm which updates the parameter at some deterministic times, and derived the convergence properties even in general latent models such as the Hidden Markov one. These properties rely on the assumption that some intermediate quantities are available analytically. Unfortunately, this is not the case in hidden Markov models with general state-spaces. In this paper, we propose an algorithm which approximates these quantities using Sequential Monte Carlo methods. The convergence of this algorithm and of an averaged version is established and their performance are illustrated through Monte Carlo experiments.

This extended version of the paper “Convergence of a Particle-based Approximation of the Block Online Expectation Maximization Algorithm“, by S. Le Corff and G. Fort, provides detailed proofs which have been omitted in the submitted paper since they are very close to existing results. These additional proofs are postponed to Appendix B.

1 Introduction

The Expectation Maximization (EM) algorithm is a well-known iterative algorithm to solve maximum likelihood estimation in incomplete data models [16].

^{*}This work is partially supported by the French National Research Agency, under the programs ANR-07-ROBO-0002 and ANR-08-BLAN-0218.

[†]LTCI, CNRS and TELECOM ParisTech, 46 rue Barrault 75634 Paris Cedex 13, France

Each iteration is decomposed into two steps: in the E-step the conditional expectation of the complete log-likelihood (log of the joint distribution of the hidden states and the observations) given the observations is computed; and the M-step updates the parameter estimate. The EM algorithm is mostly practicable if the model belongs to the exponential family, see [31, Section 1.5] and [7, Section 10.1], so that we assume below that our model belongs to this family. Under mild regularity conditions, this algorithm is known to converge to the stationary points of the log-likelihood of the observations [38]. However, the original EM algorithm cannot be used to perform online estimation or when the inference task relies on large data sets. Each iteration requires the whole data set so that each piece of data needs to be stored and is scanned to produce a new parameter estimate. *Online* variants of the EM algorithm were first proposed for independent and identically distributed (i.i.d.) observations. In [37], the parameter estimate is updated each time a new observation is available using a stochastic gradient approach. [6] proposed to replace the original E-step by a stochastic approximation using the new observation. Other solutions have been proposed to perform online EM based inference in hidden Markov models (HMM). [5] provides an algorithm for finite state-space HMM. It relies on recursive computations of the filtering distributions combined with a stochastic approximation step. In the case of finite state-spaces, deterministic approximations of these distributions are available. This algorithm has been extended to the case of general state-space models, the approximations of the filtering distributions being handled with Sequential Monte Carlo (SMC) algorithms, see [4, 12, 27]. Unfortunately, it is quite challenging to address the asymptotic behavior of these algorithms (in the HMM case) since the recursive computation of the filtering distributions relies on approximations which are really difficult to control.

In [25], another online variant of the EM algorithm in HMM is proposed, called the Block Online EM (BOEM) algorithm. In this case, the data stream is decomposed into blocks of increasing sizes. Within each block, the parameter estimate is kept fixed and the update occurs at the end of the block. This update is based on a single scan of the observations, so that it is not required to store any block of observations. [25] provides results on the convergence and on the convergence rates of both the BOEM and an averaged version. However, these analyses rely on the assumption that the E-step (computed on each block) is available analytically. This is not the case in many general state-spaces.

In this paper, we introduce an algorithm which replaces the E-step of BOEM by SMC approximations: the filtering distributions are approximated using a set of random particles with their importance weights, see [7, 10]. The Monte Carlo approximation is based on an online variant of the Forward Filtering Backward Smoothing algorithm (FFBS) proposed in [5, 12]. This method is appealing for two reasons: first, it can be implemented forwardly in time i.e. within a block, each observation is scanned once and never stored and the approximation computed on each block does not require a backward step - this is crucial in our online estimation framework. Secondly, recent work on SMC approximations provides L_p -mean control of the Monte Carlo error, see e.g. [21, 11]. This control, combined with the results in [25], sparks off the convergence results

and the convergence rates provided in this contribution.

The paper is organized as follows: our new algorithm called *Particle Block Online EM* algorithm (P-BOEM) is derived in Section 2 together with an *averaged* version. Section 3 is devoted to practical applications: P-BOEM is used to perform parameter inference in stochastic volatility models and in the more challenging framework of the Simultaneous Localization And Mapping problem (SLAM). The convergence properties of P-BOEM are addressed in Section 4. Finally, in Section 5, we derive results on the rate of convergence of P-BOEM and its averaged version.

2 The Particle Block Online EM algorithms

In Section 2.1, we fix notations that will be used throughout this paper. We then derive our online algorithms in Sections 2.2 and 2.3. We finally detail in Section 2.4, the SMC procedure that makes our algorithm a true online algorithm.

2.1 Notations and Model assumptions

A hidden Markov model on $\mathbb{X} \times \mathbb{Y}$ is defined by an initial distribution χ on $(\mathbb{X}, \mathcal{X})$ and two families of transition kernels. In this paper, the transition kernels are parametrized by $\theta \in \Theta$, $\Theta \subseteq \mathbb{R}^{d_\theta}$. In the sequel, the initial distribution χ on $(\mathbb{X}, \mathcal{X})$ is assumed to be known and fixed. The parameter is estimated online in the maximum likelihood sense using a sequence of observations \mathbf{Y} . Online maximum likelihood parameter inference algorithms were proposed either with a gradient approach or an EM approach. In the case of finite state-spaces HMM, [28] proposed a recursive maximum likelihood procedure. The asymptotic properties of this algorithm have recently been addressed in [36]. This algorithm has been adapted to general state-spaces HMM with SMC methods (see [20]). The main drawback of gradient methods is the necessity to scale the gradient components. As an alternative to perform online inference in HMM, online EM based algorithms have been proposed for finite state-spaces (see [5]) or general state-spaces HMM (see [4, 12, 25]). [12] proposed a SMC method giving encouraging experimental results. Nevertheless, it relies on a combination of stochastic approximations and SMC computations so that its analysis is quite challenging. In [25], the convergence of an online EM based algorithm is established. This algorithm requires the exact computation of intermediate quantities available explicitly in finite-state spaces HMM or in linear Gaussian models. We propose to extend this algorithm to more general models where these quantities are replaced by SMC approximations.

We now detail the model assumptions. Consider a family of transition kernels $\{m_\theta(x, x')d\lambda(x')\}_{\theta \in \Theta}$ on $\mathbb{X} \times \mathcal{X}$, where \mathbb{X} is a general state-space equipped with a countably generated σ -field \mathcal{X} , and λ is a bounded non-negative measure on $(\mathbb{X}, \mathcal{X})$. Let $\{g_\theta(x, y)d\nu(y)\}_{\theta \in \Theta}$ be a family of transition kernels on $(\mathbb{X} \times \mathbb{Y})$, where \mathbb{Y} is a general space endowed with a countably generated σ -field \mathcal{Y} and

ν is a non-negative measure on $(\mathbb{Y}, \mathcal{Y})$. Let $\mathbf{Y} = \{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$ be the observation process defined on $(\Omega, \mathbb{P}_*, \mathcal{F})$ and taking values in $\mathbb{Y}^{\mathbb{Z}}$. Hereafter, we will use $x_{r:t}$ as a shorthand notation for the sequence (x_r, \dots, x_t) , $r \leq t$.

The batch EM algorithm is an offline maximum likelihood procedure which iteratively produces parameter estimates using the complete data log-likelihood (log of the joint distribution of the observations and the states) and a fixed set of observations, see [16]. In the HMM context presented above, given T observations $\mathbf{Y}_{1:T}$, the missing data $x_{0:T}$ and a parameter θ , the complete data log-likelihood may be written as (up to the initial distribution χ which is assumed to be known)

$$\ell_\theta(x_{0:T}, \mathbf{Y}_{1:T}) \stackrel{\text{def}}{=} \sum_{t=1}^T \{\log m_\theta(x_{t-1}, x_t) + \log g_\theta(x_t, \mathbf{Y}_t)\} . \quad (1)$$

Each iteration of the batch EM is decomposed into two steps. The E-step computes, for all $\theta \in \Theta$, an expectation of the complete data log-likelihood under the conditional probability of the hidden states given the observations and the current parameter estimate $\hat{\theta}$. In the HMM context, due to the additive form of the complete data log-likelihood (1), the E-step is decomposed into T expectations under the conditional probabilities $\Phi_{\hat{\theta}, t, T}^{\chi, 0}(\cdot, \mathbf{y})$ where

$$\Phi_{\hat{\theta}, s, t}^{\chi, r}(h, \mathbf{y}) \stackrel{\text{def}}{=} \frac{\int \chi(dx_r) \{\prod_{i=r}^{t-1} m_\theta(x_i, x_{i+1}) g_\theta(x_{i+1}, \mathbf{y}_{i+1})\} h(x_{s-1}, x_s, \mathbf{y}_s) d\lambda(x_{r+1:t})}{\int \chi(dx_r) \{\prod_{i=r}^{t-1} m_\theta(x_i, x_{i+1}) g_\theta(x_{i+1}, \mathbf{y}_{i+1})\} d\lambda(x_{r+1:t})} , \quad (2)$$

for any bounded function h , any $\theta \in \Theta$, any $r < s \leq t$ and any sequence $\mathbf{y} \in \mathbb{Y}^{\mathbb{Z}}$. Then, given the current value of the parameter $\hat{\theta}$, the E-step amounts to computing the quantity

$$Q_T(\theta, \hat{\theta}) \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=1}^T \Phi_{\hat{\theta}, t, T}^{\chi, 0}(\log m_\theta + \log g_\theta, \mathbf{Y}) , \quad (3)$$

for any $\theta \in \Theta$. The M-step sets the new parameter estimate as a maximum of this expectation over θ .

The computation of $\theta \mapsto Q_T(\theta, \hat{\theta})$ for any $\theta \in \Theta$ is usually intractable except in the case of exponential complete data models see [31, Section 1.5] and [7, Section 10.1]. Therefore, in the sequel, the following assumption is assumed to hold:

- A1** (a) There exist continuous functions $\phi : \Theta \rightarrow \mathbb{R}$, $\psi : \Theta \rightarrow \mathbb{R}^d$ and $S : \mathbb{X} \times \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}^d$ s.t.

$$\log m_\theta(x, x') + \log g_\theta(x', y) = \phi(\theta) + \langle S(x, x', y), \psi(\theta) \rangle ,$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product on \mathbb{R}^d .

- (b) There exists an open subset \mathcal{S} of \mathbb{R}^d that contains the convex hull of $S(\mathbb{X} \times \mathbb{X} \times \mathbb{Y})$.

(c) There exists a continuous function $\bar{\theta} : \mathcal{S} \rightarrow \Theta$ s.t. for any $s \in \mathcal{S}$,

$$\bar{\theta}(s) = \operatorname{argmax}_{\theta \in \Theta} \{ \phi(\theta) + \langle s, \psi(\theta) \rangle \} .$$

Under A1, the quantity $Q_T(\theta, \hat{\theta})$ defined by (3) becomes

$$Q_T(\theta, \hat{\theta}) = \phi(\theta) + \left\langle \frac{1}{T} \sum_{t=1}^T \Phi_{\hat{\theta}, t, T}^{\chi, 0}(S, \mathbf{Y}), \psi(\theta) \right\rangle , \quad (4)$$

so that the definition of the function $\theta \mapsto Q_T(\theta, \hat{\theta})$ requires the computation of an expectation $\frac{1}{T} \sum_{t=1}^T \Phi_{\hat{\theta}, t, T}^{\chi, 0}(S, \mathbf{Y})$ independently of θ .

The M-step of the batch EM iteration amounts to computing

$$\bar{\theta} \left(\frac{1}{T} \sum_{t=1}^T \Phi_{\hat{\theta}, t, T}^{\chi, 0}(S, \mathbf{Y}) \right) .$$

This batch EM algorithm is designed for a fixed set of observations. A natural extension of this algorithm to the online context is to define a sequence of parameter estimates by

$$\theta_{t+1} = \operatorname{argmax}_{\theta} Q_{t+1}(\theta, \theta_t) .$$

Unfortunately, the computation of $Q_{t+1}(\theta, \theta_t)$ requires the whole set of observations to be stored and scanned for each estimation. For large data sets the computation cost of the E-step makes it intractable in this case. To overcome this difficulty, several online variants of the batch EM algorithm have been proposed, based on a recursive approximation of the function $\theta \mapsto Q_{t+1}(\cdot, \theta_t)$ (see [4, 12, 25]). In this paper, we focus on the Block Online EM (BOEM) algorithm, see [25].

2.2 Particle Block Online EM (P-BOEM)

The BOEM, introduced in [25], is an online variant of the EM algorithm. The observations are processed sequentially per block and the parameter estimate is updated at the end of each block. More precisely, let $\{\tau_k\}_{k \geq 1}$ be a sequence of positive integers denoting the length of the blocks and set

$$T_n \stackrel{\text{def}}{=} \sum_{k=1}^n \tau_k \quad \text{and} \quad T_0 \stackrel{\text{def}}{=} 0 ; \quad (5)$$

$\{T_k\}_{k \geq 1}$ are the deterministic times at which the parameter updates occur. Define, for all integers $\tau > 0$ and $T \geq 0$ and all $\theta \in \Theta$,

$$\bar{S}_{\tau}^{\chi, T}(\theta, \mathbf{Y}) \stackrel{\text{def}}{=} \frac{1}{\tau} \sum_{t=T+1}^{T+\tau} \Phi_{\theta, t, T+\tau}^{\chi, T}(S, \mathbf{Y}) . \quad (6)$$

The quantity $\bar{S}_\tau^{\chi, T}(\theta, \mathbf{Y})$ corresponds to the intermediate quantity in (4) with the observations $\mathbf{Y}_{T+1:T+\tau}$.

BOEM iteratively defines a sequence of parameter estimates $\{\theta_n\}_{n \geq 0}$ as follows: given the current parameter estimate θ_n ,

- (i) compute the quantity $\bar{S}_{\tau_{n+1}}^{\chi, T_n}(\theta_n, \mathbf{Y})$,
- (ii) compute a candidate for the new value of the parameter: $\theta_{n+1/2} = \bar{\theta} \left(\bar{S}_{\tau_{n+1}}^{\chi, T_n}(\theta_n, \mathbf{Y}) \right)$,
- (iii) stabilization step: define θ_{n+1} .

To make the exposition easier, we assume that the initial distribution χ is the same on each block. The dependence of $\bar{S}_\tau^T(\theta, \mathbf{Y})$ on χ is thus dropped from the notation for a better clarity.

The first step when proving the convergence of stochastic algorithms is to prove that the sequence of parameters $\{\theta_n\}_{n \geq 0}$ remains in a compact set - which may depend upon the path - . To ensure this stability, we introduce a stabilization step in our algorithm: we propose to use a reprojection scheme adapted from [8]. Given a sequence of compact sets $\{\Theta_n\}_{n \geq 0}$ satisfying

$$\forall n \geq 0, \Theta_n \subset \Theta_{n+1} \quad \text{and} \quad \Theta = \bigcup_{n \geq 0} \Theta_n, \quad (7)$$

an initial value $\theta_0 \in \Theta_0$ and starting with $p_0 = 0$, the stabilization step is

$$\theta_{n+1} = \begin{cases} \theta_{n+1/2} & \text{if } \theta_{n+1/2} \in \Theta_{p_n}, \text{ and set } p_{n+1} = p_n, \\ \theta_0 & \text{otherwise and set } p_{n+1} = p_n + 1. \end{cases} \quad (8)$$

p_n counts the number of “truncations”.

BOEM relies on the assumption that $\bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y})$ is available analytically, which is true e.g. in the case of linear Gaussian models and HMM with finite state-spaces. In HMM with general state-spaces $\bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y})$ cannot be computed explicitly and we propose to compute an approximation of $\bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y})$ using SMC algorithms thus yielding to the so-called *Particle-BOEM* (P-BOEM) algorithm. Different methods can be used to compute these approximations (see e.g. [11, 12, 18]). We will discuss in Section 2.4 below some SMC approximations that use the data sequentially.

Let N_n be the number of particles used on the block n ; and denote by $\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta, \mathbf{Y})$ the SMC approximation of $\bar{S}_{\tau_{n+1}}^{T_n}(\theta, \mathbf{Y})$. The P-BOEM iteratively defines a sequence of parameter estimate $\{\theta_n\}_{n \geq 0}$ as follows: given the current parameter estimate θ_n ,

- (i) compute the quantity $\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})$,
- (ii) compute a candidate for the new value of the parameter: $\theta_{n+1/2} = \bar{\theta} \left(\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) \right)$,
- (iii) stabilization step: define θ_{n+1} as in (8).

We give in Algorithm 1 lines 1 to 9 an algorithmic description of P-BOEM.

2.3 Averaged Particle Block Online EM

Following the same lines as in [25], we propose to replace the P-BOEM sequence $\{\theta_n\}_{n \geq 0}$ by an *averaged* sequence. This new sequence can be computed recursively, simultaneously with the P-BOEM sequence, and does not require additional storage of the data. The proposed averaged P-BOEM is defined as follows (see also lines 10 and 11 of Algorithm 1): the step (iii) of P-BOEM presented above is followed by

(iv) compute the quantity

$$\tilde{\Sigma}_{n+1} = \frac{T_n}{T_{n+1}} \tilde{\Sigma}_n + \frac{\tau_{n+1}}{T_{n+1}} \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}), \quad (9)$$

(v) define

$$\tilde{\theta}_{n+1} \stackrel{\text{def}}{=} \bar{\theta}(\tilde{\Sigma}_{n+1}). \quad (10)$$

We set $\tilde{\Sigma}_0 = 0$ so that

$$\tilde{\Sigma}_n = \frac{1}{T_n} \sum_{j=1}^n \tau_j \tilde{S}_{\tau_j}^{N_j, T_{j-1}}(\theta_{j-1}, \mathbf{Y}); \quad (11)$$

we will prove in Section 5 that the rate of convergence of the averaged sequence $\{\tilde{\theta}_n\}_{n \geq 0}$, computed from the averaged statistics $\{\tilde{\Sigma}_n\}_{n \geq 0}$, is better than the non-averaged one. We will also observe this property in Section 3 by comparing the variability of the P-BOEM and the averaged P-BOEM sequences in numerical applications. Note that there is no need to apply the stabilization step to the averaged sequence $\{\tilde{\theta}_n\}_{n \geq 0}$ - this will be established in Section 4.2.

2.4 The SMC approximation step

As P-BOEM is an online algorithm, the SMC algorithm should use the data sequentially: no backward pass is allowed to browse all the data at the end of the block. Hence, the approximation is computed recursively within each block, each observation being used once and never stored. These SMC algorithms will be referred to as *forward only SMC*. We detail below a forward only SMC algorithm for the computation of $\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})$ which has been proposed by [5] (see also [12]).

For notational convenience, we detail the computation of $\tilde{S}_{\tau}^{N, 0}(\theta, \mathbf{Y})$ and omit the dependence upon the block number n ; for the block n , the algorithm below has to be applied with $(\tau, N) \leftarrow (\tau_{n+1}, N_{n+1})$, $Y_{1:\tau} \leftarrow Y_{T_n+1, T_n+\tau_{n+1}}$ and $\theta \leftarrow \theta_n$.

$\tilde{S}_{\tau}^{N, 0}(\theta, \mathbf{Y})$ is a particle approximation of $\bar{S}_{\tau}^0(\theta, \mathbf{Y})$. The key property is to observe that

$$\bar{S}_{\tau}^0(\theta, \mathbf{Y}) = \phi_{\tau}^{\theta}(R_{\theta, \tau}) \quad (12)$$

Algorithm 1 P-BOEM and averaged P-BOEM

Require: $\theta_0, \{\tau_n\}_{n \geq 1}, \{N_n\}_{n \geq 1}, \{\mathbf{Y}_t\}_{t \geq 0}$.

Ensure: $\{\theta_n\}_{n \geq 0}$ and $\{p_n\}_{n \geq 0}$.

Set $\tilde{\Sigma}_0 = 0$ and $p_0 = 0$.

for all $i \geq 0$ **do**

 Compute sequentially $\tilde{S}_{\tau_{i+1}}^{N_{i+1}, T_i}(\theta_i, \mathbf{Y})$.

 Set $\theta_{i+1/2} = \bar{\theta}(\tilde{S}_{\tau_{i+1}}^{N_{i+1}, T_i}(\theta_i, \mathbf{Y}))$.

if $\theta_{i+1/2} \in \Theta_{p_i}$ **then**

 Set $\theta_{i+1} = \theta_{i+1/2}$ and $p_{i+1} = p_i$

else

 Set $\theta_{i+1} = \theta_0$ and $p_{i+1} = p_i + 1$.

end if

 Set

$$\tilde{\Sigma}_{i+1} = \frac{T_i}{T_{i+1}} \tilde{\Sigma}_i + \frac{\tau_{i+1}}{T_{i+1}} \tilde{S}_{\tau_{i+1}}^{N_{i+1}, T_i}(\theta_i, \mathbf{Y}).$$

 Set $\tilde{\theta}_{i+1} = \bar{\theta}(\tilde{\Sigma}_{i+1})$.

end for

where ϕ_t^θ is the filtering distribution at time t , and the functions $R_{t,\theta} : \mathbb{X} \rightarrow \mathcal{S}$, $1 \leq t \leq \tau$, satisfy the following equations

$$R_{t,\theta}(x) = \frac{1}{t} \mathbf{B}_t^\theta(x, S(\cdot, x, Y_t)) + \frac{t-1}{t} \mathbf{B}_t^\theta(x, R_{t-1,\theta}), \quad (13)$$

where \mathbf{B}_t^θ denotes the backward smoothing kernel at time t

$$\mathbf{B}_t^\theta(x, dx') = \frac{m_\theta(x', x)}{\int m_\theta(u, x) \phi_{t-1}^\theta(du)} \phi_{t-1}^\theta(dx'). \quad (14)$$

By convention, $R_{0,\theta}(x) = 0$ and $\phi_0^\theta = \chi$. A proof of the equalities (12) to (14) can be found in [5, 12]. Therefore, a careful reading of Eqs (12) to (14) shows that, for an iterative particle approximation of $\tilde{S}_\tau^0(\theta, \mathbf{Y})$, it is sufficient to update from time $t-1$ to t

- (i) N weighted samples $\{(\xi_t^\ell, \omega_t^\ell); \ell \in \{1, \dots, N\}\}$, approximating the filtering distribution $\phi_t^\theta(dx)$.
- (ii) the intermediate quantities $\{R_{t,\theta}^\ell\}_{\ell=1}^N$, approximating the function $R_{t,\theta}$ at point $x = \xi_t^\ell$, $\ell \in \{1, \dots, N\}$.

We describe below such an algorithm. An algorithmic description is also provided in Appendix A, Algorithm 2.

Given *instrumental* Markov transition kernels $\{q_t(x, x'), t \leq \tau\}$ on $\mathbb{X} \times \mathcal{X}$ and adjustment multipliers $\{v_t, t \leq \tau\}$, the procedure goes as follows:

- (i) *line 1 in Algorithm 2*: sample independently N particles $\{\xi_0^\ell\}_{\ell=1}^N$ with the same distribution χ .

- (ii) *line 6 in Algorithm 2*: at each time step $t \in \{1, \dots, \tau\}$, pairs $\{(J_t^\ell, \xi_t^\ell)\}_{\ell=1}^N$ of indices and particles are sampled independently (conditionally to $Y_{1:t}$, θ and $\{(J_{t-1}^\ell, \xi_{t-1}^\ell)\}_{\ell=1}^N$) from the instrumental distribution:

$$\pi_t(i, dx) \propto \omega_{t-1}^i v_t(\xi_{t-1}^i) q_t(\xi_{t-1}^i, x) \lambda(dx), \quad (15)$$

on the product space $\{1, \dots, N\} \times \mathbb{X}$. For any $t \in \{1, \dots, \tau\}$ and any $\ell \in \{1, \dots, N\}$, J_t^ℓ denotes the index of the selected particle at time $t - 1$ used to produce ξ_t^ℓ .

- (iii) *line 7 in Algorithm 2*: once the new particles $\{\xi_t^\ell\}_{\ell=1}^N$ have been sampled, their importance weights $\{\omega_t^\ell\}_{\ell=1}^N$ are computed.

- (iv) *lines 8 in Algorithm 2*: update the intermediate quantities $\{R_{t,\theta}^\ell\}_{\ell=1}^N$.

If, for all $x \in \mathbb{X}$, $v_t(x) = 1$ and if the kernels q_t are chosen such that $q_t = m_\theta$, lines 6-7 in Algorithm 2 are known as the *Bootstrap filter*. Other choices of q_t and v_t can be made, see e.g. [7].

3 Applications to Bayesian inverse problems in Hidden Markov Models

3.1 Stochastic volatility model

Consider the following stochastic volatility model (SVM):

$$X_{t+1} = \phi X_t + \sigma U_t, \quad Y_t = \beta e^{\frac{X_t}{2}} V_t,$$

where $X_0 \sim \mathcal{N}(0, (1 - \phi^2)^{-1} \sigma^2)$ and $\{U_t\}_{t \geq 0}$ and $\{V_t\}_{t \geq 0}$ are two sequences of i.i.d. standard Gaussian r.v., independent from X_0 .

We illustrate the convergence of the P-BOEM algorithms and discuss the choice of some design parameters such as the pair (τ_n, N_n) . Data are sampled using $\phi = 0.95$, $\sigma^2 = 0.1$ and $\beta^2 = 0.6$; we estimate $\theta = (\phi, \sigma^2, \beta^2)$ by applying the P-BOEM algorithm and its averaged version. All runs are started from $\phi = 0.1$, $\sigma^2 = 0.6$ and $\beta^2 = 2$.

Figure 1 displays the estimation of the three parameters as a function of the number of observations, over 50 independent Monte Carlo runs. The block-size sequence is of the form $\tau_n \propto n^{1.2}$. For the SMC step, we choose $N_n = 0.25 \cdot \tau_n$; particles are sampled as described in Algorithm 2 (see Appendix A) with the bootstrap filter. For each parameter, Figure 1 displays the empirical median (bold line) and first and last quartiles (dotted line). The averaging procedure is started after 1500 observations. Both algorithms converge to the true values of the parameters and, once the averaging procedure is started, the variance of the estimation decreases (estimation of ϕ and β^2). The estimation of σ^2 shows that if the averaging procedure is started with too few observations, the estimation can be slowed down.

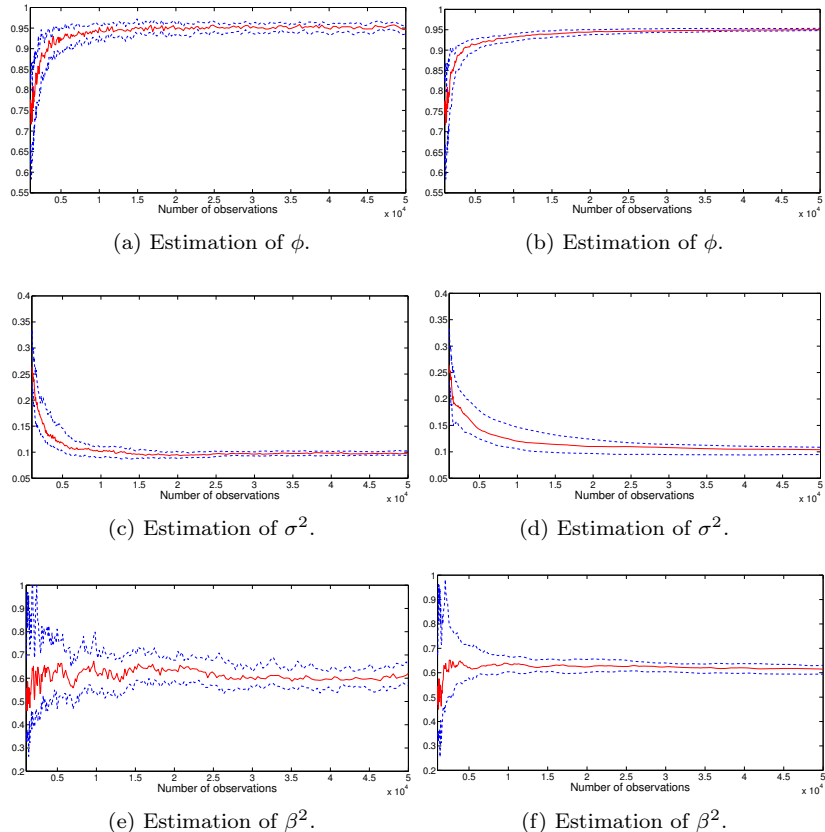
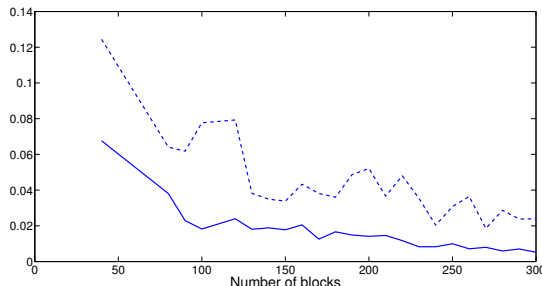


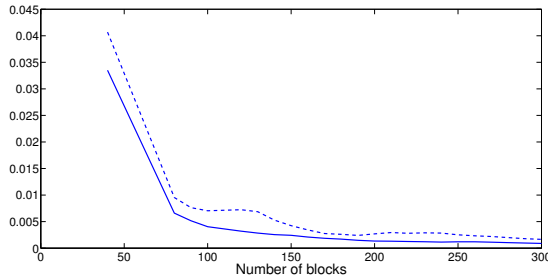
Figure 1: Estimation of ϕ , σ^2 and β^2 without (left) and with (right) averaging. Each graph represents the empirical median (bold line) and first and last quartiles (dotted line) over 50 independent Monte Carlo runs. The averaging procedure is started after 1500 observations. The first 1000 observations are not displayed for a better clarity.

We now discuss the role of the pairs (τ_n, N_n) . Roughly speaking (see section 4 for a rigorous decomposition), τ controls the rate of convergence of $\bar{S}_\tau^T(\theta, \mathbf{Y})$ to $\lim_{\tau \rightarrow \infty} \bar{S}_\tau^T(\theta, \mathbf{Y})$; and N controls the error between $\bar{S}_\tau^T(\theta, \mathbf{Y})$ and the SMC approximation $\tilde{S}_\tau^{N,T}(\theta, \mathbf{Y})$. We will show in Section 4 that $\lim_n \tau_n = \lim_n N_n = +\infty$ are part of some sufficient conditions for P-BOEM algorithms to converge. We thus choose increasing sequences $\{\tau_n, N_n\}_{n \geq 1}$. The role of τ_n has been illustrated in [25, Section 3]. Hence, in this illustration, we fix τ_n and discuss the role of N_n . Figure 2 compares the algorithms when applied with $\tau_n \propto n^{1.1}$ and $N_n = \sqrt{\tau_n}$ or $N_n = \tau_n$. The empirical variance (over 50 independent Monte Carlo runs) of the estimation of β^2 is displayed, as a function of the number of blocks. First, Figure 2 illustrates the variance decrease provided by the averaged

procedure, whatever the block size sequence. Moreover, increasing the number of particles per block improves the variance of the estimation given by P-BOEM while the impact on the variance of the averaged estimation is less important. On average, the variance is reduced by a factor of 3.0 for P-BOEM and by a factor of 1.8 for its averaged version when the number of particles goes from $N_n = \sqrt{\tau_n}$ to $N_n = \tau_n$. These practical considerations illustrate the theoretical results derived in Section 5.



(a) P-BOEM: empirical variance of the estimation of β^2 with $N_n = \sqrt{\tau_n}$ (dashed line) and $N_n = \tau_n$ (bold line).



(b) Averaged P-BOEM: empirical variance of the estimation of β^2 with $N_n = \sqrt{\tau_n}$ (dashed line) and $N_n = \tau_n$ (bold line).

Figure 2: Empirical variance of the estimation of β^2 with P-BOEM (top) and its averaged version (bottom). The averaging procedure is started after the 25-th block and the variance is displayed after a burn-in time of 35 blocks.

Finally, we discuss the role of the initial distribution χ . In all the applications above, we have the same distribution $\chi \equiv \mathcal{N}(0, (1 - \phi^2)^{-1} \sigma^2)$ at the beginning of each block. We could choose a different distribution χ_n for each block such as, e.g., the filtering distribution at the end of the previous block. We have observed that this particular choice of χ_n leads to the same behavior for both algorithms.

To end this section, the P-BOEM is compared to the Online EM algorithm outlined in [5] and [12]. These algorithms rely on a combination of stochastic approximation and SMC methods. According to classical results on stochastic

approximation, it is expected that the rate of convergence of the Online EM behaves like $\gamma_n^{1/2}$, where $\{\gamma_n\}_{n \geq 0}$ is the so called step-size sequence. Hence, γ_n in the Online EM is chosen such that $\gamma_n \propto n^{-0.55}$ and the block size sequence in the P-BOEM such that $\tau_n \propto n^{1.2}$. The number of particles used in the Online EM is fixed and chosen so that the computational costs of both algorithms are similar. Provided that $N_n \propto \tau_n$ in the P-BOEM, this leads to a choice of 70 particles for the Online EM. We report in Figure 3, the estimation of ϕ and σ^2 for a Polyak-Ruppert averaged Online EM (see [35]) and the averaged P-BOEM as a function of the number of observations. The averaging procedure is started after about 1500 observations. As noted in [25, Section 3] for a constant sequence $\{N_n\}_{n \geq 0}$ this figure shows that both algorithms behave similarly. For the estimation of ϕ and β^2 , the variance is smaller for P-BOEM and the convergence is faster for P-BOEM in the case of β^2 . Conclusions are different for the estimation of σ^2 : the variance is smaller for P-BOEM but the Online EM converges a bit faster. The main advantage of P-BOEM is that it relies on approximations which can be controlled in such a way that we are able to show that the limiting points of P-BOEM algorithms are the stationary points of the limiting normalized log-likelihood of the observations.

3.2 P-BOEM applied to Simultaneous Localization And Mapping

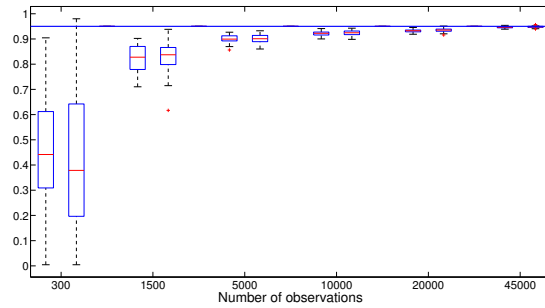
The Simultaneous Localization And Mapping (SLAM) problem arises when a mobile device wants to build a map of an unknown environment and, at the same time, has to estimate its position in this map. The common statistical approach for SLAM is to introduce a state space model. Many solutions have been proposed depending on the assumptions made on the transition and observation models, and on the map (see e.g. [3, 30, 33]). In [30, 27], it is proposed to see the SLAM as an inference problem in HMM: the localisation of the robot is the hidden state with Markovian dynamic, and the map is seen as an unknown parameter. Therefore, the mapping problem is answered by solving the inference task, and the localization problem is answered by approximating the conditional distribution of the hidden states given the observations.

In this application, we consider a statistical model for a landmark-based SLAM problem for a bicycle manoeuvring on a plane surface.

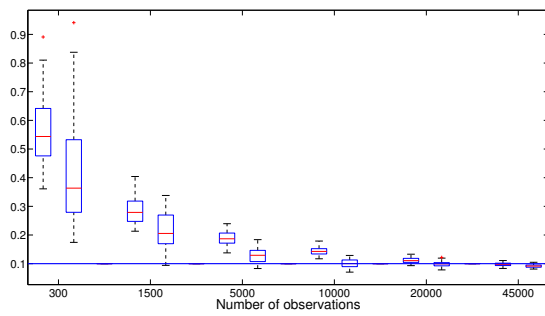
Let $x_t \stackrel{\text{def}}{=} \{x_{t,i}\}_{i=1}^3$ be the robot pose, where $x_{t,1}$ and $x_{t,2}$ are the robot's cartesian coordinates and $x_{t,3}$ its orientation. At each time step, deterministic controls are sent to the robot so that it explores a given part of the environment. Controls are denoted by (v_t, ψ_t) where ψ_t stands for the robot's heading direction and v_t its velocity. The robot pose at time t , given its previous pose at time $t-1$ and the noisy controls $(\hat{v}_t, \hat{\psi}_t)$, can be written as:

$$x_t = f(x_{t-1}, \hat{v}_t, \hat{\psi}_t), \quad (16)$$

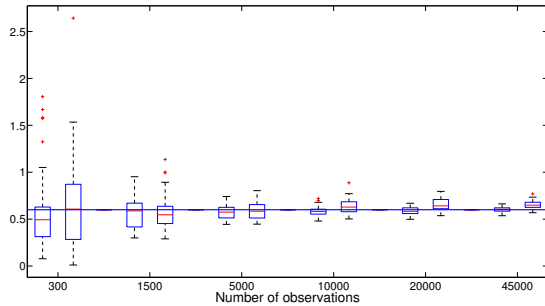
where $(\hat{v}_t, \hat{\psi}_t)$ is a 2-dimensional Gaussian distribution with mean (v_t, ψ_t) and known covariance matrix Q . In this contribution, we use the kinematic model



(a) Estimation of ϕ .



(b) Estimation of σ^2 .



(c) Estimation of β^2 .

Figure 3: Estimation of ϕ , σ^2 and β^2 with the averaged P-BOEM (left) and a Polyak-Ruppert averaged version of the Online EM (right) after 300, 1500, 5000, 10000, 20000 and 45000 observations. The averaging procedure is started after about 1000 observations (which corresponds to the 25-th block for P-BOEM).

of the front wheel of a bicycle (see e.g. [1]) where the function f in (16) is given

by

$$f(x_{t-1}, \hat{v}_t, \hat{\psi}_t) = x_{t-1} + \begin{pmatrix} \hat{v}_t d_t \cos(x_{t-1,3} + \hat{\psi}_t) \\ \hat{v}_t d_t \sin(x_{t-1,3} + \hat{\psi}_t) \\ \hat{v}_t d_t B^{-1} \sin(\hat{\psi}_t) \end{pmatrix},$$

where d_t is the time period between two successive poses and B is the robot wheelbase.

The 2-dimensional environment is represented by a set of landmarks $\theta \stackrel{\text{def}}{=} \{\theta_j\}_{1 \leq j \leq q}$, $\theta_j \in \mathbb{C}$ being the position of the j -th landmark. The total number of landmarks q and the association between observations and landmarks are assumed to be known.

At time t , the robot observes the distance and the angular position of all landmarks in its neighborhood; let $c_t \subseteq \{1, \dots, q\}$ be the set of observed landmarks at time t . It is assumed that the observations $\{y_{t,i}\}_{i \in c_t}$ are independent and satisfy

$$y_{t,i} = h(x_t, \theta_i) + \delta_{t,i},$$

where h is defined by

$$h(x, \tau) \stackrel{\text{def}}{=} \begin{pmatrix} \sqrt{(\tau_1 - x_1)^2 + (\tau_2 - x_2)^2} \\ \arctan \frac{\tau_2 - x_2}{\tau_1 - x_1} - x_3 \end{pmatrix},$$

and the noise vectors $\{\delta_{t,i}\}_{t,i}$ are i.i.d Gaussian $\mathcal{N}(0, R)$. R is assumed to be known.

The model presented in this Section does not take into account all the issues arising in the SLAM problem (such as the association process which is assumed to be known and the known covariance matrices). The aim is to prove that BOEM and its averaged version have a satisfying behavior even in the challenging framework described above. The observation and motion models are highly non linear and we show that BOEM remains stable in this experiment. Several solutions have been proposed to solve the association problem (see e.g. [3] for a solution based on the likelihood of the observations) and could be adapted to our case.

We want to estimate $\theta = \{\theta_j\}_{j=1}^q$ by applying the P-BOEM algorithms. In this paper, we use simulated data. $q = 15$ landmarks are drawn in a square of size 45x45. The robot path is sampled with a given set of controls. Using the *true* positions of all landmarks in the map and the true path of the robot (see the dots and the bold line on Figure 4), observations are sampled by setting:

$$R = \begin{pmatrix} \sigma_r^2 & \rho \\ \rho & \sigma_b^2 \end{pmatrix}, \text{ where } \sigma_r = 0.5\text{m}, \sigma_b = \frac{\pi}{60}\text{rad} \text{ and } \rho = 0.01. \text{ We choose } Q = \text{diag}(\sigma_v^2, \sigma_\phi^2) \text{ where } \sigma_v = 0.5\text{m.s}^{-1}, \sigma_\phi = \frac{\pi}{60}\text{rad} \text{ and } B = 1.5\text{m}.$$

In this model, the transition denoted by m_θ does not depend on the map θ (see (16)) and the marginal likelihood g_θ is such that the model does not belong to the exponential family:

$$\sum_{i \in c_t} \ln g_\theta(x_t, y_{t,i}) \propto \sum_{i \in c_t} [y_{t,i} - h(x_t, \theta_i)]^* R^{-1} [y_{t,i} - h(x_t, \theta_i)]. \quad (17)$$

Hence, in order to apply Algorithm 1, at the beginning of each block, g_θ is approximated by a function depending on the current parameter estimate so that the resulting approximated model belongs to the exponential family (see [27]). As it can be seen from (17), approximating the function $\tau \mapsto h(x, \tau)$ by its first order Taylor expansion at θ_i leads to a quadratic approximation of g_θ . This approach is commonly used in the SLAM framework to use the properties of linear Gaussian models (see e.g. [3]).

As the landmarks are not observed all the time, we choose a slowly increasing sequence $\{\tau_n \propto n^{1.1}\}_{n \geq 1}$ so that the number of updates is not too small (in this experiment, we have 60 updates for a total number of observations of 2000). As the total number of observations is not so large (the largest block is of length 60), the number of particles is chosen to be constant on each block: for all $n \geq 1$, $N_n = 50$. For the SMC step, we apply Algorithm 2 with the *bootstrap filter*.

For each run the estimated path (equal to the weighted mean of the particles) and the estimated map at the end of the loop ($T = 2000$) are stored. Figure 4 represents the mean estimated path and the mean map over 50 independent Monte Carlo runs. It highlights the good performance of the P-BOEM algorithm in a more complex framework.

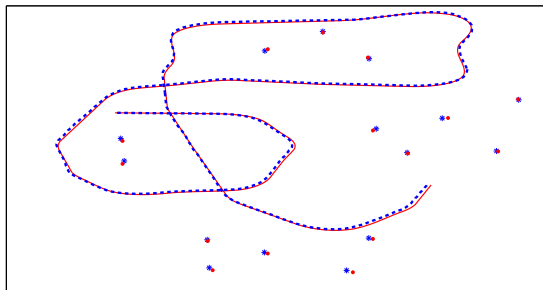


Figure 4: True trajectory (bold line) and true landmark positions (balls) with the estimated path (dotted line) and the landmarks estimated positions (stars) at the end of the run ($T = 2000$).

We also compare our algorithm to the *marginal SLAM* proposed by [30]. In this algorithm, the map is also modeled as a parameter to learn in a HMM model; SMC methods are used to estimate the map in the maximum likelihood sense. Marginal SLAM is a gradient-based approach for solving the recursive maximum likelihood procedure. Figure 5 illustrates the estimation of the position of each landmark. P-BOEM is applied using the same parameters as above and the *marginal SLAM* uses a sequence of step-size $\{\gamma_n \propto n^{-0.6}\}_{n \geq 1}$. We use the averaged version of P-BOEM and a Polyak-Ruppert based averaging procedure for the *marginal SLAM* (see [35]). For each landmark the last estimation (at the end of the loop) of the position is stored for each of the 50 independent Monte Carlo runs. Figure 5 displays the distance between the estimated position and the true position for each landmark. In this experiment, the P-BOEM based SLAM algorithm outperforms the *marginal SLAM*.

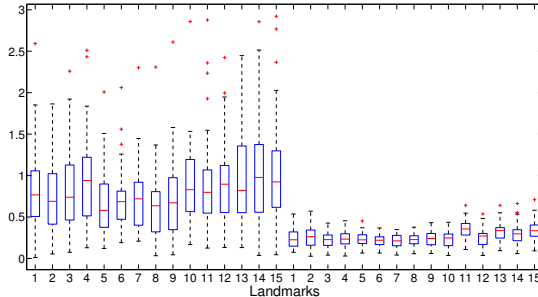


Figure 5: Distance between the final estimation and the true position for each of the 15 landmarks with the averaged *marginal SLAM* (left) and the averaged P-BOEM algorithm (right).

4 Convergence of the Particle Block Online EM algorithms

In this section, we analyze the limiting points of the P-BOEM algorithm. We prove in Theorem 4.2 that P-BOEM has the same limit points as a so-called *limiting EM* algorithm, which would consist of defining a sequence $\{\theta_n\}_{n \geq 0}$ by $\theta_{n+1} = \bar{\theta} \circ \bar{S}(\theta_n)$ where $\bar{S}(\theta)$ is the a.s. limit $\lim_{\tau \rightarrow +\infty} \bar{S}_\tau^T(\theta, \mathbf{Y})$ (defined by (6)). As discussed in [25, Section 4.3.], the set of the limit points of the limiting EM is the set of the stationary points of the contrast function $c_*(\theta)$, defined as the a.s. limit of the normalized log-likelihood of the observations. This convergence result on P-BOEM requires two sets of assumptions: conditions A2 to A5 and part of A7 are the same as in [25] and imply the convergence of BOEM; assumptions A6 and A7 are introduced to control the difference between P-BOEM and BOEM.

4.1 Assumptions

Consider the following assumptions

A2 There exist σ_- and σ_+ s.t. for any $(x, x') \in \mathbb{X}^2$ and any $\theta \in \Theta$, $0 < \sigma_- \leq m_\theta(x, x') \leq \sigma_+$. Set $\rho \stackrel{\text{def}}{=} 1 - (\sigma_-/\sigma_+)$.

Define the shift operator ϑ onto $\mathbb{Y}^{\mathbb{Z}}$ by $(\vartheta \circ \mathbf{y})_k = \mathbf{y}_{k+1}$ for any $k \in \mathbb{Z}$; and by induction, define the s -iterated shift operator

$$\vartheta^{s+1} \circ \mathbf{y} = \vartheta \circ (\vartheta^s \circ \mathbf{y}), \quad \forall s \geq 0, \quad (18)$$

with the convention that ϑ^0 is the identity operator. The shift operator is said to be ergodic for \mathbb{P}_* if for each set A in $\{A \in \mathcal{B}(\mathbb{Y})^{\otimes \mathbb{Z}}; A = \vartheta^{-1}(A)\}$, $\mathbb{P}_*(A) \in \{0, 1\}$ (see [2, p.314]). Define, for all $y \in \mathbb{Y}$,

$$b_-(y) \stackrel{\text{def}}{=} \inf_{\theta \in \Theta} \int g_\theta(x, y) \lambda(dx) \quad \text{and} \quad b_+(y) \stackrel{\text{def}}{=} \sup_{\theta \in \Theta} \int g_\theta(x, y) \lambda(dx). \quad (19)$$

A3-(γ) $\mathbb{E}_\star [\sup_{x,x' \in \mathbb{X}^2} |S(x, x', \mathbf{Y}_0)|^\gamma] < +\infty$.

A4 (a) Under \mathbb{P}_\star , \mathbf{Y} is a stationary sequence.

(b) The shift operator is ergodic with respect to \mathbb{P}_\star .

(c) $\mathbb{E}_\star [|\log b_-(\mathbf{Y}_0)| + |\log b_+(\mathbf{Y}_0)|] < +\infty$.

For any sequence of r.v. $Z \stackrel{\text{def}}{=} \{Z_t\}_{t \in \mathbb{Z}}$ on $(\Omega, \tilde{\mathbb{P}}, \mathcal{F})$, let

$$\mathcal{F}_k^Z \stackrel{\text{def}}{=} \sigma(\{Z_u\}_{u \leq k}) \quad \text{and} \quad \mathcal{G}_k^Z \stackrel{\text{def}}{=} \sigma(\{Z_u\}_{u \geq k}) \quad (20)$$

be σ -fields associated to Z . We also define the mixing coefficients by, see [9],

$$\beta^Z(n) \stackrel{\text{def}}{=} \sup_{u \in \mathbb{Z}} \beta(\mathcal{G}_{u+n}^Z, \mathcal{F}_u^Z), \forall n \geq 0, \quad (21)$$

where for any σ -algebras \mathcal{F} and \mathcal{G} ,

$$\beta(\mathcal{G}, \mathcal{F}) \stackrel{\text{def}}{=} \sup_{B \in \mathcal{G}} |\tilde{\mathbb{P}}(B|\mathcal{F}) - \tilde{\mathbb{P}}(B)|. \quad (22)$$

A5 There exist $C \in [0, 1)$ and $\beta \in (0, 1)$ s.t. for any $n \geq 0$, $\beta^{\mathbf{Y}}(n) \leq C\beta^n$, where $\beta^{\mathbf{Y}}$ is defined in (21).

Assumptions A2 to A5 are the same as in [25]. A2, referred to as the *strong mixing condition*, is used to prove the uniform forgetting property of the initial condition of the filter, see e.g. [13, 14]. This assumption is easy to check in finite state-space HMM or when the state-space is compact when the Markov kernel m_θ is sufficiently regular. As noted in [25], it can fail to hold in quite general situations. Nevertheless, the exponential forgetting property needed to ensure the convergence results could be checked under weaker assumptions (see [17] for a Doeblin assumption). However, it would imply quite technical supplementary results out of the scope of this paper. Examples of observation sequences satisfying A4(a-b) and A5 can be found in [26]; it includes, for example, stationary ψ -irreducible and positive recurrent Markov chains which are geometrically ergodic (see e.g. [32] for Markov chains theory).

We conclude this set of assumptions with conditions on the SMC approximation, on the length of the blocks and on the number of particles per block.

A6-(γ)

(a) There exists a filtration $\{\tilde{\mathcal{F}}_{T_n}\}_{n \geq 0}$ on (Ω, \mathcal{F}) such that for any $n \geq 0$,

(i) $\mathcal{F}_{T_n}^{\mathbf{Y}} \subseteq \tilde{\mathcal{F}}_{T_n}$,

(ii) $\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})$ is $\tilde{\mathcal{F}}_{T_{n+1}}$ -measurable, where $\{\theta_n\}_{n \geq 0}$ is the P-BOEM sequence,

(iii) for any $m \geq 1$ and any $B \in \mathcal{G}_{T_n+m}^{\mathbf{Y}}$, $\mathbb{P}_\star(B|\tilde{\mathcal{F}}_{T_n}) = \mathbb{P}_\star(B|\mathcal{F}_{T_n}^{\mathbf{Y}})$.

(b) For any compact $\mathcal{K} \subseteq \Theta$, there exists C such that for any $N, n > 0$,

$$\mathbb{E}_* \left[\left| \tilde{S}_{\tau_{n+1}}^{N, T_n}(\theta_n, \mathbf{Y}) - \bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}) \right|^\gamma \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \leq C \left(\frac{1}{N^\gamma} + \frac{1}{(\tau_{n+1}N)^{\gamma/2}} \right).$$

We will discuss in Section 4.3 below how to check A6 when the SMC approximation is computed as described in Section 2.4.

A7-(γ) The block size sequence $\{\tau_n\}_{n \geq 1}$ and the number of particles $\{N_n\}_{n \geq 1}$ satisfy

$$\sum_{k \geq 1} \frac{1}{\tau_k^{\gamma/2}} + \sum_{k \geq 1} \frac{1}{N_k^\gamma} < \infty.$$

If, for all block k , we choose $N_k = \sqrt{\tau_k}$, A7-(γ) is reduced to $\sum_{k \geq 1} \tau_k^{-\gamma/2} < \infty$, which is the same assumption as A6-(γ) in [25].

4.2 Asymptotic behavior of the Particle Block Online EM algorithms

Following [25], we address the convergence of the P-BOEM algorithm as the convergence of a perturbed version of the *limiting EM* recursion. The following result, which is proved in [25, Theorem 4.1.], shows that when τ is large, the BOEM statistic $\tilde{S}_\tau^T(\theta, \mathbf{Y})$ is an approximation of a deterministic quantity $\bar{S}(\theta)$; the *limiting EM* is the iterative algorithm defined by $\theta_{n+1} = \mathbf{R}(\theta_n)$ where

$$\mathbf{R}(\theta) \stackrel{\text{def}}{=} \bar{\theta}(\bar{S}(\theta)), \forall \theta \in \Theta; \quad (23)$$

the mapping $\bar{\theta}$ is given by A1.

Theorem 4.1. *Let $S : \mathbb{X}^2 \times \mathbb{Y} \rightarrow \mathbb{R}^d$ be a measurable function s.t. A3-(1) holds. Assume A2 and A4(a-b). For any $\theta \in \Theta$, there exists a \mathbb{P}_* -integrable r.v. $\mathbb{E}_\theta [S(X_{-1}, X_0, \mathbf{Y}_0) | \mathbf{Y}]$ s.t. for any $T > 0$,*

$$\tilde{S}_\tau^T(\theta, \mathbf{Y}) \xrightarrow{\tau \rightarrow +\infty} \bar{S}(\theta) \stackrel{\text{def}}{=} \mathbb{E}_* [\mathbb{E}_\theta [S(X_{-1}, X_0, \mathbf{Y}_0) | \mathbf{Y}]] \quad \mathbb{P}_* - a.s. \quad (24)$$

Moreover, $\theta \mapsto \bar{S}(\theta)$ is continuous on Θ .

The asymptotic behavior of the limiting EM algorithm is addressed in [25, Section 4.3.]: the main ingredient is that the map \mathbf{R} admits a Lyapunov function W w.r.t. the set $\mathcal{L} \stackrel{\text{def}}{=} \{\theta \in \Theta; \mathbf{R}(\theta) = \theta\}$. It is proved in [25, Proposition 4.2] that there exists a positive and continuous Lyapunov function W i.e. (i) $W \circ \mathbf{R}(\theta) \geq W(\theta)$ for any $\theta \in \Theta$ and, (ii) for any compact subset \mathcal{K} of $\Theta \setminus \mathcal{L}$, $\inf_{\theta \in \mathcal{K}} W \circ \mathbf{R}(\theta) - W(\theta) > 0$. This Lyapunov function is equal to $\exp(c_*(\theta))$, where the contrast function $c_*(\theta)$ is the (deterministic) limit of the normalized log-likelihood of the observations (see [26, Theorem 4.6]).

Theorem 4.2 establishes the convergence of P-BOEM to the set \mathcal{L} . The proof of Theorem 4.2 is on the same lines as the proof of [25, Theorem 4.4.] and details

are postponed to Appendix B. The main ingredient is the L_p -mean control of the error when, starting from θ_n , the limiting EM update \bar{S} is replaced with the SMC approximation $\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\cdot, \mathbf{Y})$:

$$\left| \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) - \bar{S}(\theta_n) \right| \mathbf{1}_{\theta_n \in \mathcal{K}},$$

where \mathcal{K} is a compact subset of Θ . Such a control is derived in Proposition 6.1 in Section 6.

Let $\text{Cl}(A)$ be the closure of the set A .

Theorem 4.2. *Assume A1-2, A3-(\bar{p}), A4-5, A6-(p) and A7-(p) for some $2 < p < \bar{p}$. Assume in addition that $W(\mathcal{L})$ is compact and, for any $M > 0$, the level set $\{\theta \in \Theta; W(\theta) \geq M\}$ is compact. Then,*

- (a) $\limsup_n p_n < +\infty$ \mathbb{P}_\star - a.s where p_n is defined in (8).
- (b) $\{W(\theta_n)\}_{n \geq 0}$ converges to a connected component of $W(\mathcal{L})$.
- (c) If $W(\mathcal{L} \cap \text{Cl}(\{\theta_n\}_{n \geq 0}))$ has an empty interior, there exists w_\star s.t. $\{W(\theta_n)\}_{n \geq 0}$ converges almost surely to w_\star and $\{\theta_n\}_{n \geq 0}$ converges to $\{\theta \in \mathcal{L}; W(\theta) = w_\star\}$.

The assumptions made in Theorem 4.2 are of common use to prove the convergence of EM based procedures or stochastic approximation algorithms. They are used in [38] to find the limit points of the classical EM algorithm. See also [15] and [22] for the stability of the Monte Carlo EM algorithm and of a stochastic approximation of the EM algorithm. The compactness of the level holds if $\lim W(\theta) = 0$, as $\theta \rightarrow \partial\Theta$ (where $\partial\Theta$ is the boundary of Θ), and thus highly depends on the model. Moreover, if W is sufficiently regular, Sard's theorem states that $W(\mathcal{L})$ has Lebesgue measure 0 and hence has an empty interior.

Under the assumptions of Theorem 4.2, it can be proved that along any converging P-BOEM sequence $\{\theta_n\}_{n \geq 0}$ to θ_\star in \mathcal{L} , the averaged P-BOEM statistics $\{\tilde{S}_n\}_n$ defined by (9) (see also (11)) converge to $\bar{S}(\theta_\star)$ - see Proposition 6.2 in Section 6. Since $\bar{\theta}$ is continuous, the averaged P-BOEM sequence $\{\tilde{\theta}_n\}_{n \geq 0}$ converges to $\bar{\theta}(\bar{S}(\theta_\star)) = R(\theta_\star)$. Since $\theta_\star \in \mathcal{L}$, $R(\theta_\star) = \theta_\star$ showing that the averaged P-BOEM algorithm has the same limit points as the P-BOEM algorithm.

4.3 Comments on Assumption A6

Assumption A6 is needed to control the L_p -error on each block between $\tilde{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y})$ and its SMC approximation. In this section, we give sufficient conditions to check A6 in the case of the forward only SMC presented in Section 2.4.

For each block n , denote by $\{v_{t,n}\}_{t \leq \tau_{n+1}}$ and $\{q_{t,n}\}_{t \leq \tau_{n+1}}$ resp. the adjustment multipliers and the instrumental kernels in the SMC propagation step (see (15)). For all $y \in \mathbb{Y}$ and all compact subset $\mathcal{K} \subseteq \Theta$, define

$$\omega_+^{\mathcal{K}}(y) = \sup_{\theta \in \mathcal{K}} \sup_{\substack{(x, x') \in \mathbb{X} \times \mathbb{X} \\ t \geq 0, n \geq 0}} \frac{m_\theta(x, x') g_\theta(x', y)}{v_{t,n}(x) q_{t,n}(x, x')}.$$

It is assumed:

$$\mathbf{A8} \quad |v|_\infty \stackrel{\text{def}}{=} \sup_{t,n} |v_{t,n}|_\infty < \infty.$$

$$\mathbf{A9}-(\gamma) \quad \mathbb{E}_* \left[\left| \frac{\omega_+^{\mathcal{K}}(\mathbf{Y}_0)}{b_-(\mathbf{Y}_0)} \right|^\gamma \right] < +\infty.$$

In the case of the *Bootstrap filter*, A8 holds (since $v_{t,n} = 1$); furthermore, note that $\omega_+^{\mathcal{K}}(y) = \sup_{\theta \in \mathcal{K}} \sup_{x \in \mathbb{X}} \rho_\theta(x, y)$.

Proposition 4.3. *Let $S : \mathbb{X}^2 \times \mathbb{Y} \rightarrow \mathbb{R}^d$ be a measurable function s.t. A3-(\bar{p}) holds for some $\bar{p} > 2$. Assume A2, A4(a), A8. Define $\Delta p \stackrel{\text{def}}{=} 2\bar{p}p/(\bar{p} - p)$ and assume A9-(Δp) holds for some $p \in (2, \bar{p})$. Then, A6-(p) holds when the SMC approximation is computed by the algorithm described in Section 2.4.*

5 Rate of convergence of the Particle Block On-line EM algorithm

In this section, we consider a converging P-BOEM sequence $\{\theta_n\}_{n \geq 0}$ with limiting point $\theta_* \in \mathcal{L}$. It can be shown, as in [25, Proposition 5.1], that the convergence of the sequence $\{\theta_n\}_{n \geq 0}$ is equivalent to the convergence of the sufficient statistics $\{\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})\}_{n \geq 0}$: along any P-BOEM sequence converging to θ_* , this sequence of sufficient statistics converges to $s_* = \bar{S}(\theta_*)$. We will thus discuss the rate of convergence of the sufficient statistics.

Let $G : \mathcal{S} \rightarrow \mathcal{S}$ be the limiting EM map defined on the space of sufficient statistics by

$$G(s) \stackrel{\text{def}}{=} \bar{S}(\bar{\theta}(s)), \quad \forall s \in \mathcal{S}. \quad (25)$$

The rate of convergence of P-BOEM and of its averaged version are established as in [25, Theorem 5.2]. These results rely on two sets of assumptions: A10 and A11 are the same as in [25] to provide convergence rates for BOEM and A12 is required to control the SMC approximation introduced in P-BOEM.

A10 (a) \bar{S} and $\bar{\theta}$ are twice continuously differentiable on Θ and \mathcal{S} .
 (b) There exists $0 < \gamma < 1$ s.t. $\text{sp}(\nabla_s G(s_*)) \leq \gamma$ where sp denotes the spectral norm.

A11 (a) $\{\tau_{n+1}/\tau_n\}_{n \geq 0}$ converges to q and $\gamma q < 1$.
 (b) $\limsup_n \sum_{k=1}^n \left\{ \left| \frac{\tau_{k+1}}{\tau_k} - q \right| \tau_k^{1/2} + \log \tau_k \right\} / \sqrt{T_n} < \infty$.

As discussed in [25], A11(a-b) are satisfied with polynomial rates $\tau_n \sim cn^b$ with $b \geq 1$ (in this case $q = 1$) and geometric rates $\tau_n \sim a\tau^n$ with $\tau \in (1, \gamma^{-1})$ (in this case $q > 1$).

A12 (a) $\{N_{n+1}/N_n\}_{n \geq 0}$ converges to α and $\gamma\alpha^2 < 1$.

$$(b) \limsup_n \sum_{k=1}^n \left\{ \sqrt{\frac{\tau_{k+1}}{N_{k+1}}} + \frac{\tau_{k+1}}{N_{k+1}} \right\} / \sqrt{T_n} < +\infty.$$

A12 can be checked for a number of particles of the form $N_n = \tau_n^a$, $a > 0$. In this case and under A11(a), we have $\alpha = q^a$. Hence, if $a \leq 1/2$, A11(a) implies A12(a). If $a > 1/2$, A11(a) should be strenghtened into $\gamma q^{2a} < 1$. In the case of polynomial rates $\tau_n \sim cn^b$, $b > 1$, A12(b) is satisfied when $a \geq \frac{1+b}{2b}$. In the case of geometric rates, $\tau_n \sim a\tau^n$, A12(b) is satisfied when $a \geq 1/2$.

We will use the following notations: for any sequence of random variables $\{Z_n\}_{n \geq 0}$, write $Z_n = \mathcal{O}_{L_p}(1)$ if $\limsup_n \mathbb{E}_* [|Z_n|^p] < \infty$; and $Z_n = \mathcal{O}_{a.s.}(1)$ if $\sup_n |Z_n| < +\infty$ $\mathbb{P}_* - a.s.$

Theorem 5.1. *Assume A1-2, A3-(\bar{p}), A4-5, A6-(p), A7-(p), A10, A11(a) and A12(a) for some $2 < p < \bar{p}$. Then,*

$$\begin{aligned} & \left[\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) - s_* \right] \mathbf{1}_{\lim_k \tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) = s_*} \\ &= \mathcal{O}_{L_p} \left(\frac{1}{\sqrt{\tau_n}} + \frac{1}{N_n} \right) + \mathcal{O}_{L_{p/2}} \left(\frac{1}{N_n^2} + \frac{1}{\tau_n} \right) \mathcal{O}_{a.s.}(1). \end{aligned} \quad (26)$$

If in addition A11(b) and A12(b) hold then,

$$\left[\tilde{\Sigma}_n - s_* \right] \mathbf{1}_{\lim_k \tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) = s_*} = \mathcal{O}_{L_p} \left(\frac{1}{\sqrt{T_n}} \right) + \mathcal{O}_{L_{p/2}} \left(\frac{n}{T_n} \right) \mathcal{O}_{a.s.}(1). \quad (27)$$

Theorem 5.1 gives the rate of convergence of particle BOEM for the sufficient statistics. However, as noted in [25, Theorem 5.1], this result is enough to prove the same rate of convergence for the sequence of parameter estimates $\{\theta_n\}_{n \geq 0}$ (and for the averaged estimates).

The proof of Theorem 5.1 is along the same lines as the proof of [25, Theorem 5.2.]. The main difficulty for the proof of Eq.(27) is to obtain a L_p -moment of the error

$$\frac{1}{T_{n+1}} \sum_{k=1}^n \tau_{k+1} \left[\tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) - \bar{S}(\theta_k) \right] \mathbf{1}_{\left| \tilde{S}_{\tau_k}^{N_k, T_{k-1}}(\theta_{k-1}, \mathbf{Y}) - s_* \right| \leq \eta},$$

for any $\eta > 0$. Such a control is provided in Section 6, Lemma 6.3. The proof of Theorem 5.1 is omitted and can be found in the extended version of this paper (see [24]).

Eq. (26) shows that the leading term in the error $\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) - s_*$ has a L_p -norm decreasing as $N_n^{-1} \wedge \tau_n^{-1/2}$; the L_p -error decreasing at the rate N_n^{-1} (resp. $\tau_n^{-1/2}$) is due to the SMC approximation of the BOEM algorithm (resp. the BOEM approximation of the limiting EM). Hence, the rate of the P-BOEM algorithm depends both on the choice of the number of observations and the number of particles per block. By choosing $N_n \sim \sqrt{\tau_n}$, the SMC error and the BOEM error are balanced; and the rate of convergence of P-BOEM decreases as $1/\sqrt{\tau_n}$. Unfortunately, such a rate is obtained after a total number of observations T_n ; therefore, as discussed in [25], is quite sub-optimal. Eq (26) shows

that the rate of convergence equal to the square root of the total number of observations up to block n , can be reached by using the averaged P-BOEM: the error $\tilde{\Sigma}_n - s_\star$ has a rate of convergence proportional to $T_n^{-1/2}$ when $\lim_n n/\sqrt{T_n} = 0$. This is the case with polynomial rates $\tau_n \sim cn^b$ when $b > 1$ and with geometric rates, $\tau_n \sim a\tau^n$, $\tau > 1$. In addition, the rate of convergence of the averaged P-BOEM does not depend on the number of particles per block. This can be observed in the experiments presented in Section 3, Figure 2.

6 Proofs

For $p > 0$ and Z a random variable measurable w.r.t. the σ -algebra \mathcal{F} , set

$$\|Z\|_{\star,p} \stackrel{\text{def}}{=} (\mathbb{E}_\star [|Z|^p])^{1/p} .$$

6.1 Proof of Proposition 4.3

(a) For all $n \geq 0$, define the σ -algebra $\tilde{\mathcal{F}}_{T_n}$ by

$$\tilde{\mathcal{F}}_{T_n} \stackrel{\text{def}}{=} \sigma \left\{ \mathcal{F}_{T_n}^{\mathbf{Y}}, \left\{ \left(\xi_{t_k}^{\ell_k,k}, \omega_{t_k}^{\ell_k,k} \right) \right\}_{\ell_k=1}^N, k \in \{0, \dots, n-1\}, t_k \in \{0, \dots, \tau_{k+1}\} \right\} ,$$

where $\mathcal{F}_{T_n}^{\mathbf{Y}}$ is defined by (20) and where $\left\{ \left(\xi_{t_k}^{\ell_k,k}, \omega_{t_k}^{\ell_k,k} \right) \right\}_{\ell_k=1}^{N_{k+1}}$ is the SMC approximation of the filtering distribution at time t_k in the block k . It is clear that $\mathcal{F}_{T_n}^{\mathbf{Y}} \subseteq \tilde{\mathcal{F}}_{T_n}$, showing that $\tilde{\mathcal{F}}_{T_n}$ satisfies A6a(i). Note that the random variable θ_n is measurable with respect to $\tilde{\mathcal{F}}_{T_n}$ and that $\tilde{S}_{\tau_{n+1}}^{N,T_n}(\theta_n, \mathbf{Y})$ is measurable with respect to $\tilde{\mathcal{F}}_{T_{n+1}}$, ensuring A6a(ii). According to the selection and mutation procedure given in Section 2.4, the σ -algebra $\tilde{\mathcal{F}}_{T_n}$ can be written as,

$$\tilde{\mathcal{F}}_{T_n} \stackrel{\text{def}}{=} \sigma \{ \mathcal{F}_{T_n}^{\mathbf{Y}}, \mathcal{H}_{T_n} \} ,$$

where \mathcal{H}_{T_n} is independent from \mathbf{Y} (the σ -algebra \mathcal{H}_{T_n} is generated by the random variables independent from the observations \mathbf{Y} used to produce the weighted samples at each time step). Hence, for any positive integer m and any $B \in \mathcal{G}_{T_n+m}^{\mathbf{Y}}$, since \mathcal{H}_{T_n} is independent from B and from $\mathcal{F}_{T_n}^{\mathbf{Y}}$, A6a(iii) holds.

(b) For any $t \in \{0, \dots, \tau_{n+1}\}$, define the σ -algebra $\mathcal{F}_{n,t}^N$ by

$$\mathcal{F}_{n,t}^N \stackrel{\text{def}}{=} \sigma \{ \theta_n, \mathbf{Y}_{T_n+1:T_n+t+1}, (\xi_s^\ell, \omega_s^\ell); \ell \in \{1, \dots, N\}; 0 \leq s \leq t \} . \quad (28)$$

Under A2 and A8, Propositions B.5, B.8 and B.9 in Appendix B can be applied so that

$$\mathbb{E}_\star \left[\left| \tilde{S}_{\tau_{n+1}}^{N,T_n}(\theta_n, \mathbf{Y}) - \tilde{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \leq C (I_{1,n} + I_{2,n}) , \quad (29)$$

where (since $p/2 \geq 1$, $\alpha \stackrel{\text{def}}{=} \bar{p}/p \geq 1$ and $\beta^{-1} \stackrel{\text{def}}{=} 1 - \alpha^{-1} \leq 1$),

$$I_{1,n} \stackrel{\text{def}}{=} \frac{1}{\tau_{n+1}^{\frac{p}{2}+1} N^{\frac{p}{2}}} \times \sum_{t=0}^{\tau_{n+1}} \mathbb{E}_\star \left[\left| \frac{\omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})}{b_-(\mathbf{Y}_{t+T_n})} \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\} \right|^p \right],$$

$$I_{2,n} \stackrel{\text{def}}{=} \frac{1}{\tau_{n+1} N^p} \times \sum_{t=0}^{\tau_{n+1}} \mathbb{E}_\star \left[\left| \frac{\omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})}{b_-(\mathbf{Y}_{t+T_n})} \right|^{2p} \mathbb{E}_\star \left[\left| \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\} \right|^{\bar{p}} \middle| \mathcal{F}_{t-1,n}^N \right]^{p/\bar{p}} \right].$$

We use $S_s(x, x')$ as a shorthand notation for $S(x, x', \mathbf{Y}_s)$. By the Hölder inequality applied with $\alpha = \bar{p}/p$ and $\beta^{-1} = 1 - \alpha^{-1}$,

$$I_{1,n} \leq \frac{1}{\tau_{n+1}^{\frac{p}{2}+1} N^{\frac{p}{2}}} \sum_{t=1}^{\tau_{n+1}} \left\| \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S_{s+T_n}\} \right\|_{\star, \bar{p}}^p \times \left\| \frac{\omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})}{b_-(\mathbf{Y}_{t+T_n})} \right\|_{\star, \bar{p}p/(\bar{p}-p)}^p.$$

By A3-(\bar{p}), A4(a) and A9-(Δp), we have

$$I_{1,n} \leq \frac{C}{\tau_{n+1}^{\frac{p}{2}} N^{\frac{p}{2}}}.$$

Using similar arguments for $I_{2,n}$ yields $I_{2,n} \leq C N^{-p}$.

6.2 L_p -controls

Proposition 6.1. *Let $S : \mathbb{X}^2 \times \mathbb{Y} \rightarrow \mathbb{R}^d$ be a measurable function s.t. A3-(\bar{p}) holds for some $\bar{p} > 2$. Assume A2, A4(a), A5 and A6(p) for some $p \in (2, \bar{p})$. For any compact set $\mathcal{K} \subseteq \Theta$, there exists a constant C s.t. for any $n \geq 1$,*

$$\left\| \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) - \bar{S}(\theta_n) \right\|_{\star, p} \leq C \left(\frac{1}{\sqrt{\tau_{n+1}}} + \frac{1}{N_{n+1}} \right).$$

Proof. The proof is written with $N \leftarrow N_{n+1}$. Under A6(p), θ_n is $\tilde{\mathcal{F}}_{T_n}$ -measurable and, for any positive integer m and any $B \in \mathcal{G}_{T_n+m}^{\mathbf{Y}}$,

$$\sup_{B \in \mathcal{G}_{T_n+m}^{\mathbf{Y}}} |\mathbb{P}_\star(B | \tilde{\mathcal{F}}_{T_n}) - \mathbb{P}_\star(B)| = \sup_{B \in \mathcal{G}_{T_n+m}^{\mathbf{Y}}} |\mathbb{P}_\star(B | \mathcal{F}_{T_n}^{\mathbf{Y}}) - \mathbb{P}_\star(B)|.$$

Hence, the mixing coefficients defined in (22) are such that

$$\beta(\mathcal{G}_{T_n+m}^{\mathbf{Y}}, \tilde{\mathcal{F}}_{T_n}) = \beta(\mathcal{G}_{T_n+m}^{\mathbf{Y}}, \mathcal{F}_{T_n}^{\mathbf{Y}}).$$

Therefore, under A2, A3-(\bar{p}), A4(a) and A5, following the same steps as in [25, Proposition 6.5] it can be proved that there exists a constant C s.t.

$$\left\| \bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}) - \bar{S}(\theta_n) \right\|_{\star, p} \leq \frac{C}{\sqrt{\tau_{n+1}}}.$$

Moreover, under A6(p),

$$\mathbb{E} \left[\left| \tilde{S}_{\tau_{n+1}}^{N, T_n}(\theta_n, \mathbf{Y}) - \bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \leq C \left(\frac{1}{N^p} + \frac{1}{\tau_{n+1}^{p/2} N^{p/2}} \right),$$

which concludes the proof. \square

It is proved in Proposition 6.2 that there is no need to apply the stabilization step to the averaged sequence $\{\tilde{\theta}_n\}_{n \geq 0}$.

Proposition 6.2. *Let $S : \mathbb{X}^2 \times \mathbb{Y} \rightarrow \mathbb{R}^d$ be a measurable function s.t. A3-(\bar{p}) holds for some $\bar{p} > 2$. Assume A2, A4-5, A6(p) and A7-(p) for some $2 < p < \bar{p}$. Let $\{\theta_n\}_n$ be the P-BOEM sequence. For any $\theta_\star \in \Theta$, on the set $\{\lim_n \theta_n = \theta_\star\}$,*

$$\tilde{\Sigma}_n \rightarrow \bar{S}(\theta_\star), \mathbb{P}_\star - a.s.,$$

where \bar{S} is defined in (24) and $\tilde{\Sigma}_n$ in (9).

Proof. By (9), $\tilde{\Sigma}_n$ can be written as

$$\tilde{\Sigma}_n = \frac{1}{T_n} \sum_{j=1}^n \tau_j \left[\tilde{S}_{\tau_j}^{N_j, T_{j-1}}(\theta_{j-1}, \mathbf{Y}) - \bar{S}(\theta_{j-1}) \right] + \frac{1}{T_n} \sum_{j=1}^n \tau_j \bar{S}(\theta_{j-1}). \quad (30)$$

By Theorem 4.1, \bar{S} is continuous so, by the Cesaro Lemma, the second term in the rhs of (30) converges to $\bar{S}(\theta_\star)$ \mathbb{P}_\star -a.s., on the set $\{\lim_n \theta_n = \theta_\star\}$.

On the set $\{\lim_n \theta_n = \theta_\star\}$, the number of projections is finite (w.p.1). Therefore, in order to prove that the first term in (30) converges to zero a.s., it is sufficient to prove that the convergence holds on the set $\{\forall q, \theta_q \in \mathcal{K}\}$ for any compact set \mathcal{K} . By Proposition 6.1, there exists a constant C such that for any n ,

$$\left\| \left| \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) - \bar{S}(\theta_n) \right| \mathbf{1}_{\theta_n \in \mathcal{K}} \right\|_{\star, p} \leq C \left(\frac{1}{\sqrt{\tau_{n+1}}} + \frac{1}{N_{n+1}} \right).$$

Hence, by A7-(p) and the Borel-Cantelli Lemma,

$$\left| \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) - \bar{S}(\theta_n) \right| \mathbf{1}_{\{\forall q, \theta_q \in \mathcal{K}\}} \leq \left| \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) - \bar{S}(\theta_n) \right| \mathbf{1}_{\theta_n \in \mathcal{K}} \rightarrow 0, \mathbb{P}_\star - a.s.$$

The proof is concluded by applying the Cesaro Lemma. \square

Lemma 6.3. *Assume A1-2, A3-(\bar{p}), A4-5, A6(p), A10, A11(b) and A12(b) for some $p \in (2, \bar{p})$. Then, for any $\eta > 0$,*

$$\limsup_{n \rightarrow +\infty} \frac{1}{\sqrt{T_{n+1}}} \left\| \sum_{k=1}^n \tau_{k+1} \left[\tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) - \bar{S}(\theta_k) \right] \mathbf{1}_{\left| \tilde{S}_{\tau_k}^{N_k, T_{k-1}}(\theta_{k-1}, \mathbf{Y}) - s_\star \right| \leq \eta} \right\|_{\star, p} < \infty,$$

where \bar{S} is defined by (24).

Proof. Let $\eta > 0$. We have

$$\tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) - \bar{S}(\theta_k) = \bar{S}_{\tau_{k+1}}^{T_k}(\theta_k, \mathbf{Y}) - \bar{S}(\theta_k) + \left[\tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) - \bar{S}_{\tau_{k+1}}^{T_k}(\theta_k, \mathbf{Y}) \right].$$

Under A2, A3-(\bar{p}), A4-5, A6(p), A10 and A11(b), using similar arguments than in [25, Lemma 6.7] (by replacing $\mathcal{F}_{T_n}^{\mathbf{Y}}$ by $\tilde{\mathcal{F}}_{T_n}$ as in the proof of Proposition 4.3 and noting that by A6(p), $\mathbf{1}_{|\tilde{S}_{\tau_k}^{N_k, T_{k-1}}(\theta_{k-1}, \mathbf{Y}) - s_*| \leq \eta}$ is $\tilde{\mathcal{F}}_{T_n}$ -measurable) yields

$$\limsup_{n \rightarrow +\infty} \frac{1}{\sqrt{T_{n+1}}} \left\| \sum_{k=1}^n \tau_{k+1} \left(\bar{S}_{\tau_{k+1}}^{T_k}(\theta_k, \mathbf{Y}) - \bar{S}(\theta_k) \right) \mathbf{1}_{|\tilde{S}_{\tau_k}^{N_k, T_{k-1}}(\theta_{k-1}, \mathbf{Y}) - s_*| \leq \eta} \right\|_{\star, p} < \infty.$$

Since $\bar{\theta}$ is a continuous function (see A1), there exists a compact set \mathcal{K} of Θ such that on the set $\{|\tilde{S}_{\tau_k}^{N_k, T_{k-1}}(\theta_{k-1}, \mathbf{Y}) - s_*| \leq \eta\}$, $\theta_k = \bar{\theta}(\tilde{S}_{\tau_k}^{N_k, T_{k-1}}(\theta_{k-1}, \mathbf{Y})) \in \mathcal{K}$. Therefore, by A6(p), there exists a constant C s.t. for any $n \geq 1$,

$$\begin{aligned} \frac{1}{\sqrt{T_{n+1}}} \sum_{k=1}^n \tau_{k+1} \left\| \left| \tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) - \bar{S}_{\tau_{k+1}}^{T_k}(\theta_k, \mathbf{Y}) \right| \mathbf{1}_{|\tilde{S}_{\tau_k}^{N_k, T_{k-1}}(\theta_{k-1}, \mathbf{Y}) - s_*| \leq \eta} \right\|_{\star, p} \\ \leq \frac{1}{\sqrt{T_{n+1}}} \sum_{k=1}^n \tau_{k+1} \left\| \left| \tilde{S}_{\tau_{k+1}}^{N_{k+1}, T_k}(\theta_k, \mathbf{Y}) - \bar{S}_{\tau_{k+1}}^{T_k}(\theta_k, \mathbf{Y}) \right| \mathbf{1}_{\theta_k \in \mathcal{K}} \right\|_{\star, p} \\ \leq \frac{C}{\sqrt{T_{n+1}}} \sum_{k=1}^n \tau_{k+1} \left(\frac{1}{\sqrt{\tau_{k+1} N_{k+1}}} + \frac{1}{N_{k+1}} \right), \\ \leq \frac{C}{\sqrt{T_{n+1}}} \sum_{k=1}^n \left(\sqrt{\frac{\tau_{k+1}}{N_{k+1}}} + \frac{\tau_{k+1}}{N_{k+1}} \right). \end{aligned}$$

The proof is concluded upon noting that the last term is bounded under A12(b). \square

A Detailed SMC algorithm

In this section, we give a detailed description of the SMC algorithm used to compute sequentially the quantities $\tilde{S}_{\tau_{n+1}}^{N, T_n}(\theta_n, \mathbf{Y})$, $n \geq 0$. This is the algorithm proposed by [5] and [12].

At each time step, the weighted samples are produced using sequential importance sampling and sampling importance resampling steps. In Algorithm 2, the instrumental proposition kernel used to select and propagate the particles is π_t (see (15) and [18, 19, 29] for further details on this SMC step).

It is readily seen from the description below that the observations \mathbf{Y}_t are processed sequentially.

B Additional proofs

For $p > 0$ and Z a random variable measurable w.r.t. the σ -algebra \mathcal{F} , set $\|Z\|_{\star, p} \stackrel{\text{def}}{=} (\mathbb{E}_{\star} [|Z|^p])^{1/p}$.

Algorithm 2 Forward SMC step

Require: $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$.

Ensure: $\tilde{S}_{\tau_{n+1}}^{N, T_n}(\theta_n, \mathbf{Y})$.

Sample $\{\xi_0^\ell\}_{\ell=1}^N$ i.i.d. with distribution χ .

Set $\omega_0^\ell = 1/N$ for all $\ell \in \{1, \dots, N\}$.

Set $R_{0, \theta_n}^\ell = 0$ for all $\ell \in \{1, \dots, N\}$.

for $t = 1$ to τ_{n+1} **do**

for $\ell = 1$ to N **do**

 Conditionally to $(\theta_n, Y_{T_n+1:T_n+t}, \{J_{t-1}^\ell, \xi_{t-1}^\ell\}_{\ell=1}^N)$, sample independently $(J_t^\ell, \xi_t^\ell) \sim \pi_t(i, dx)$, where $\pi_t(i, dx) \propto \omega_{t-1}^i v_t(\xi_{t-1}^i) q_t(\xi_{t-1}^i, x) \lambda(dx)$.

 Set

$$\omega_t^\ell = \frac{m_{\theta_n}(\xi_{t-1}^{J_t^\ell}, \xi_t^\ell) g_{\theta_n}(\xi_t^\ell, \mathbf{Y}_{T_n+t})}{v_t(\xi_{t-1}^{J_t^\ell}) q_t(\xi_{t-1}^{J_t^\ell}, \xi_t^\ell)} .$$

 Set

$$R_{t, \theta_n}^\ell = \frac{1}{t} \sum_{j=1}^N \omega_{t-1}^j m_{\theta_n}(\xi_{t-1}^j, \xi_{t+1}^\ell) \frac{S(\xi_{t-1}^j, \xi_t^\ell, \mathbf{Y}_{T_n+t}) + (t-1) R_{t-1, \theta_n}^j}{\sum_{k=1}^N \omega_{t-1}^k m_{\theta_n}(\xi_{t-1}^k, \xi_t^\ell)} .$$

end for

end for

Set

$$\tilde{S}_{\tau_{n+1}}^{N, T_n}(\theta_n, \mathbf{Y}) = \sum_{\ell=1}^N \omega_{\tau_{n+1}}^\ell R_{\tau_{n+1}, \theta_n}^\ell .$$

B.1 Proof of Theorem 4.2

We check the assumptions of [22, Proposition 9] and [22, Proposition 11] with $T(\theta) \stackrel{\text{def}}{=} \mathbf{R}(\theta)$ (see (23)), $F_n(\theta) \stackrel{\text{def}}{=} \bar{\theta} \left(\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}) \right)$ and $\mathcal{L} \stackrel{\text{def}}{=} \{\theta \in \Theta; \mathbf{R}(\theta) = \theta\}$.

We start by checking the conditions of [22, Proposition 11]. Under the stated assumptions, (a) holds. For (c) we prove that for any compact subset $\mathcal{K} \subseteq \Theta$,

$$\left| \mathbf{W} \circ \mathbf{R}(\theta_n) - \mathbf{W} \circ \bar{\theta}(\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})) \right| \mathbf{1}_{\theta_n \in \mathcal{K}} \xrightarrow{n \rightarrow +\infty} 0 \quad \mathbb{P}_* - \text{a.s.} \quad (31)$$

By Theorem 4.1, the function \bar{S} is continuous on Θ and then $\bar{S}(\mathcal{K}) \stackrel{\text{def}}{=} \{s \in \mathcal{S}; \exists \theta \in \mathcal{K}, s = \bar{S}(\theta)\}$ is compact. For any $\delta > 0$ (small enough), we can define the compact subset $\bar{S}(\mathcal{K}, \delta) \stackrel{\text{def}}{=} \{s \in \mathbb{R}^d; d(s, \bar{S}(\mathcal{K})) \leq \delta\}$ of \mathcal{S} , where $d(s, \bar{S}(\mathcal{K})) \stackrel{\text{def}}{=} \inf_{s' \in \bar{S}(\mathcal{K})} |s - s'|$. Let $\delta > 0$ (small enough) and $\varepsilon > 0$. Since $\mathbf{W} \circ \bar{\theta}$ is continuous (see A1(c) and [25, Proposition 4.2]) and $\bar{S}(\mathcal{K}, \delta)$ is compact, $\mathbf{W} \circ \bar{\theta}$ is uniformly continuous on $\bar{S}(\mathcal{K}, \delta)$ and there exists $\eta > 0$ s.t.,

$$\forall x, y \in \bar{S}(\mathcal{K}, \delta), \quad |x - y| \leq \eta \Rightarrow |\mathbf{W} \circ \bar{\theta}(x) - \mathbf{W} \circ \bar{\theta}(y)| \leq \varepsilon . \quad (32)$$

Set $\alpha \stackrel{\text{def}}{=} \delta \wedge \eta$ and $\Delta S_n \stackrel{\text{def}}{=} |\bar{S}(\theta_n) - \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})| \mathbf{1}_{\theta_n \in \mathcal{K}}$. We write,

$$\begin{aligned} & \mathbb{P}_\star \left\{ \left| W \circ \bar{\theta}(\bar{S}(\theta_n)) - W \circ \bar{\theta}(\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})) \right| \mathbf{1}_{\theta_n \in \mathcal{K}} \geq \varepsilon \right\} \\ &= \mathbb{P}_\star \left\{ \left| W \circ \bar{\theta}(\bar{S}(\theta_n)) - W \circ \bar{\theta}(\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})) \right| \mathbf{1}_{\theta_n \in \mathcal{K}} \geq \varepsilon; \Delta S_n > \delta \right\} \\ &\quad + \mathbb{P}_\star \left\{ \left| W \circ \bar{\theta}(\bar{S}(\theta_n)) - W \circ \bar{\theta}(\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})) \right| \mathbf{1}_{\theta_n \in \mathcal{K}} \geq \varepsilon; \Delta S_n \leq \delta \right\} \\ &\leq \mathbb{P}_\star \{ \Delta S_n > \delta \} + \mathbb{P}_\star \{ \Delta S_n > \eta \} \leq 2\mathbb{P}_\star \{ \Delta S_n > \alpha \} . \end{aligned}$$

By the Markov inequality and Proposition 6.1, there exists a constant C s.t.

$$\begin{aligned} & \mathbb{P}_\star \left\{ \left| W \circ \bar{\theta}(\bar{S}(\theta_n)) - W \circ \bar{\theta}(\tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})) \right| \mathbf{1}_{\theta_n \in \mathcal{K}} \geq \varepsilon \right\} \\ &\leq \frac{2}{\alpha^p} \mathbb{E}_\star \left[|\bar{S}(\theta_n) - \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y})|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \\ &\leq C \left(\left[\frac{1}{\tau_{n+1}} \right]^{p/2} + \left[\frac{1}{N_{n+1}} \right]^p \right) . \end{aligned}$$

(31) follows from A7-(p) and the Borel-Cantelli lemma. The proof of the condition (b) follows the same lines. By [22, Proposition 11], this implies that $\limsup_n p_n < +\infty$ \mathbb{P}_\star -a.s and that $\{\theta_n\}_{n \geq 0}$ is a compact sequence \mathbb{P}_\star -a.s. For the other statements, we apply [22, Proposition 9]. $\mathcal{L} \cap \mathcal{K}$ is compact since \mathcal{L} is closed and \mathcal{K} is compact. We now prove that for any compact subset $\mathcal{K} \subset \Theta$,

$$|W(\theta_{n+1}) - W \circ R(\theta_n)| \mathbf{1}_{\theta_n \in \mathcal{K}} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \mathbb{P}_\star \text{-a.s.} \quad (33)$$

Since $\limsup_n p_n < +\infty$ \mathbb{P}_\star -a.s, it is sufficient to prove this convergence on the set $\{\omega \in \Omega; \limsup_n p_n(\omega) < +\infty\}$. For any ω s.t. $\limsup_n p_n(\omega) < +\infty$, there exists (a random) n_0 s.t., for any $n \geq n_0$, $p_n(\omega) = p_{n+1}(\omega)$ and then $\theta_{n+1}(\omega) = \theta_{n+1/2}(\omega)$, see (8). Therefore (33) follows from (31).

B.2 Proof of Theorem 5.1

We introduce the shorthand notation

$$\bar{S}_0 \stackrel{\text{def}}{=} \bar{S}_{\tau_1}^0(\theta_0, \mathbf{Y}) \quad \text{and} \quad \bar{S}_n \stackrel{\text{def}}{=} \bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}), \quad \forall n \geq 0, \quad (34)$$

and

$$\tilde{S}_0 \stackrel{\text{def}}{=} \tilde{S}_{\tau_1}^{N_1, 0}(\theta_0, \mathbf{Y}) \quad \text{and} \quad \tilde{S}_n \stackrel{\text{def}}{=} \tilde{S}_{\tau_{n+1}}^{N_{n+1}, T_n}(\theta_n, \mathbf{Y}), \quad \forall n \geq 0. \quad (35)$$

We will also use the following notations: for any sequence of random variables $\{Z_n\}_{n \geq 0}$, write $Z_n = \mathcal{O}_{L_p}(1)$ if $\limsup_n \mathbb{E}_\star [|Z_n|^p] < \infty$; $Z_n = o_{L_p}(1)$ if $\limsup_n \mathbb{E}_\star [|Z_n|^p] = 0$; $Z_n = \mathcal{O}_{\text{a.s.}}(1)$ if $\sup_n |Z_n| < +\infty$ \mathbb{P}_\star -a.s; and $Z_n = o_{\text{a.s.}}(1)$ if $\limsup_n |Z_n| = 0$ \mathbb{P}_\star -a.s

B.2.1 Proof of Eq. (26)

The proof relies on the following decomposition of $\tilde{S}_n - s_*$: since $G(s_*) = s_*$, we write

$$\tilde{S}_n - s_* = \Gamma(\tilde{S}_{n-1} - s_*) + \tilde{S}_n - G(\tilde{S}_{n-1}) + G(\tilde{S}_{n-1}) - G(s_*) - \Gamma(\tilde{S}_{n-1} - s_*) .$$

Let \mathcal{K} be a compact subset of Θ and $\eta > 0$. Define $\{\tilde{\mu}_n\}_{n \geq 0}$ and $\{\tilde{\rho}_n\}_{n \geq 0}$ s.t. $\tilde{\mu}_0 = 0$, $\tilde{\rho}_0 = \tilde{S}_0 - s_*$ and

$$\tilde{\mu}_n \stackrel{\text{def}}{=} \Gamma \tilde{\mu}_{n-1} + \tilde{e}_n , \quad \tilde{\rho}_n \stackrel{\text{def}}{=} \tilde{S}_n - s_* - \tilde{\mu}_n , \quad n \geq 1 , \quad (36)$$

where,

$$\tilde{e}_n \stackrel{\text{def}}{=} \left[\tilde{S}_n - \bar{S}(\theta_n) \right] \mathbf{1}_{|\tilde{S}_{n-1} - s_*| \leq \eta} , \quad n \geq 1 . \quad (37)$$

Proposition B.1. *Assume A1-2, A3-(\bar{p}), A4-5, A6-(p), A7-(p), A10, A11(a) and A12(a) for some $2 < p < \bar{p}$. Then,*

$$\begin{aligned} \tilde{\mu}_n &= \mathcal{O}_{L_p} \left(\frac{1}{\sqrt{\tau_n}} \right) + \mathcal{O}_{L_p} \left(\frac{1}{N_n} \right) , \\ \tilde{\rho}_n \mathbf{1}_{\lim_k \tilde{S}_k = s_*} &= \mathcal{O}_{\text{a.s.}}(1) \mathcal{O}_{L_{p/2}} \left(\frac{1}{\tau_n} + \frac{1}{N_n^2} \right) . \end{aligned}$$

Proof. By (36), for all $n \geq 1$, $\tilde{\mu}_n = \sum_{k=0}^{n-1} \Gamma^k \tilde{e}_{n-k}$. By A10 and the Minkowski inequality, for all $n \geq 1$, $\|\tilde{\mu}_n\|_{*,p} \leq \sum_{k=0}^{n-1} \gamma^k \|\tilde{e}_{n-k}\|_{*,p}$. By Proposition 6.1, there exists a constant C s.t. for any $n \geq 1$,

$$\|\tilde{\mu}_n\|_{*,p} \leq C \sum_{k=0}^{n-1} \gamma^k \sqrt{\frac{1}{\tau_{n+1-k}}} + C \sum_{k=0}^{n-1} \gamma^k \frac{1}{N_{n+1-k}} .$$

By [34, Result 178, p. 39], A11(a) and A12(a) (upon noting that $q \geq 1$ and $\alpha \geq 1$, so that $\sqrt{q}\gamma < 1$ and $\alpha\gamma < 1$), this yields

$$\tilde{\mu}_n = \mathcal{O}_{L_p} \left(\frac{1}{\sqrt{\tau_n}} \right) + \mathcal{O}_{L_p} \left(\frac{1}{N_n} \right) . \quad (38)$$

By A10, using a Taylor expansion with integral form of the remainder term,

$$\begin{aligned} &G(\tilde{S}_{n-1}) - G(s_*) - \Gamma(\tilde{S}_{n-1} - s_*) \\ &= \sum_{i,j=1}^d \left(\tilde{S}_{n-1,i} - s_{*,i} \right) \left(\tilde{S}_{n-1,j} - s_{*,j} \right) \tilde{R}_{n-1}(i,j) \\ &= \sum_{i,j=1}^d (\tilde{\mu}_{n-1,i} + \tilde{\rho}_{n-1,i}) (\tilde{\mu}_{n-1,j} + \tilde{\rho}_{n-1,j}) \tilde{R}_{n-1}(i,j) , \end{aligned}$$

where $x_{n,i}$ denotes the i -th component of $x_n \in \mathbb{R}^d$ and

$$\tilde{R}_n(i, j) \stackrel{\text{def}}{=} \int_0^1 (1-t) \frac{\partial^2 \mathbf{G}}{\partial s_i \partial s_j} \left(s_* + t(\tilde{S}_n - s_*) \right) dt, \quad n \in \mathbb{N}, 1 \leq i, j \leq d.$$

Observe that under A10, $|\tilde{R}_n| \mathbf{1}_{\lim_n \theta_n = \theta_*} = \mathcal{O}_{\text{a.s.}}(1)$. Define for $n \geq 1$ and $k \leq n$,

$$\tilde{H}_n \stackrel{\text{def}}{=} \sum_{i=1}^d (2\tilde{\mu}_{n,i} + \tilde{\rho}_{n,i}) \tilde{R}_n(i, \cdot), \quad \tilde{r}_n \stackrel{\text{def}}{=} \sum_{i,j=1}^d \tilde{R}_n(i, j) \tilde{\mu}_{n,i} \tilde{\mu}_{n,j}, \quad (39)$$

$$\tilde{\delta}_n \stackrel{\text{def}}{=} \tilde{S}(\theta_n) - \mathbf{G}(\tilde{S}_{n-1}), \quad \tilde{\psi}(n, k) \stackrel{\text{def}}{=} (\Gamma + \tilde{H}_n) \cdots (\Gamma + \tilde{H}_k), \quad (40)$$

$$\tilde{\epsilon}_n \stackrel{\text{def}}{=} \left[\tilde{S}_n - \tilde{S}(\theta_n) \right] \mathbf{1}_{|\tilde{S}_{n-1} - s_*| > \eta}, \quad (41)$$

with the convention $\tilde{\psi}(n, n+1) \stackrel{\text{def}}{=} \text{Id}$. By (36),

$$\tilde{\rho}_n = \tilde{\psi}(n-1, 0) \tilde{\rho}_0 + \sum_{k=0}^{n-1} \tilde{\psi}(n-1, k+1) \tilde{r}_k + \sum_{k=1}^n \tilde{\psi}(n-1, k) \left\{ \tilde{\delta}_k + \tilde{\epsilon}_k \right\}. \quad (42)$$

By (38), A7-(p) implies that $\tilde{\mu}_n = o_{\text{a.s.}}(1)$. This yields, by (36), $\tilde{\rho}_n \mathbf{1}_{\lim_n \tilde{S}_n = s_*} = o_{\text{a.s.}}(1)$ and, by (39), it follows $\lim_{n \rightarrow +\infty} \left| \tilde{H}_n \right| \mathbf{1}_{\lim_n \tilde{S}_n = s_*} = o_{\text{a.s.}}(1)$.

Let $\epsilon \stackrel{\text{def}}{=} q^{-1} \wedge \alpha^{-2}$ and $\tilde{\gamma} \in (\gamma, \epsilon)$, where γ is given by A10 and α by A12(a) (this is possible since by A11(a) and A12(a), $\gamma q < 1$ and $\gamma \alpha \leq \gamma \alpha^2 < 1$).

Since $\lim_{n \rightarrow +\infty} \left| \tilde{H}_n \right| \mathbf{1}_{\lim_n \tilde{S}_n = s_*} = 0$, there exists a \mathbb{P}_* -a.s finite random variable Z_1 s.t., for all $0 \leq k \leq n-1$,

$$\left| \tilde{\psi}(n-1, k) \right| \mathbf{1}_{\lim_n \tilde{S}_n = s_*} \leq \tilde{\gamma}^{n-k} Z_1 \mathbf{1}_{\lim_n \tilde{S}_n = s_*}. \quad (43)$$

Therefore, $\left| \tilde{\psi}(n-1, 0) \tilde{\rho}_0 \right| \mathbf{1}_{\lim_n \tilde{S}_n = s_*} = \tilde{\gamma}_n \mathcal{O}_{\text{a.s.}}(1)$, and, by A3-(\bar{p}), (2), (6) and Proposition 6.1 $\mathbb{E}_* [|\tilde{\rho}_0|^p] < +\infty$ which implies that $\tilde{\rho}_0 < +\infty$ \mathbb{P}_* -a.s. Since $\alpha^2 \tilde{\gamma} < 1$, the first term in the RHS of (42) is $N_n^{-2} o_{L^p}(1) \mathcal{O}_{\text{a.s.}}(1)$.

We now consider the second term in the RHS of (42). From equation (43),

$$\left| \sum_{k=0}^{n-1} \tilde{\psi}(n-1, k+1) \tilde{r}_k \right| \mathbf{1}_{\lim_n \tilde{S}_n = s_*} \leq Z_1 \sum_{k=0}^{n-1} \tilde{\gamma}^{n-k-1} |\tilde{r}_k| \mathbf{1}_{\lim_n \tilde{S}_n = s_*} \quad \mathbb{P}_* \text{-a.s.}$$

By (39) and A10, there exists a \mathbb{P}_* -a.s finite random variable Z_2 s.t.

$$|\tilde{r}_k| \mathbf{1}_{\lim_n \tilde{S}_n = s_*} \leq Z_2 \sum_{i,j=1}^d \tilde{\mu}_{k,i} \tilde{\mu}_{k,j}, \quad \mathbb{P}_* \text{-a.s.}$$

In addition, by (38), there exists a constant C s.t.

$$\left\| \sum_{k=0}^{n-1} \tilde{\gamma}^{n-k-1} \sum_{i,j=1}^d \tilde{\mu}_{k,i} \tilde{\mu}_{k,j} \right\|_{\star, p/2} \leq C \sum_{k=0}^{n-1} \tilde{\gamma}^{n-k-1} \left(\frac{1}{\tau_k} + \frac{1}{N_k^2} \right).$$

Under A11(a) and A12(a), applying again [34, Result 178, p. 39] yields that the second term in the RHS of (42) is $\left(\frac{1}{\tau_n} + \frac{1}{N_n^2}\right) \mathcal{O}_{\text{a.s.}}(1) \mathcal{O}_{L_{p/2}}(1)$.

We finally consider the third term in the RHS of (42). By Theorem 4.2, the number of projections is finite w.p.1. so that $\theta_n = \bar{\theta}(\tilde{S}_{n-1})$ for any n large enough (depending upon the path). In addition, on the set $\{\lim_n \tilde{S}_n = s_\star\}$, $\tilde{\epsilon}_n = 0$ for any n large enough (depending upon the path). Therefore, $\tilde{\delta}_n + \tilde{\epsilon}_n = 0$ for any n large enough. Hence,

$$\left| \sum_{k=1}^n \tilde{\psi}(n-1, k) \left\{ \tilde{\delta}_k + \tilde{\epsilon}_k \right\} \mathbf{1}_{\lim_n \tilde{S}_n = s_\star} \right| = \tilde{\gamma}^n \mathcal{O}_{\text{a.s.}}(1).$$

Since $\alpha^2 \tilde{\gamma} < 1$, the third term in the RHS of (42) is $N_n^{-2} \mathcal{O}_{\text{a.s.}}(1)$. \square

B.2.2 Proof of Eq. (27)

We write $\tilde{\Sigma}_n - s_\star = \bar{\mu}_n + \bar{\rho}_n$ with

$$\bar{\mu}_n \stackrel{\text{def}}{=} \frac{1}{T_n} \sum_{k=1}^n \tau_k \tilde{\mu}_{k-1} \quad \text{and} \quad \bar{\rho}_n \stackrel{\text{def}}{=} \frac{1}{T_n} \sum_{k=1}^n \tau_k \tilde{\rho}_{k-1}. \quad (44)$$

Proposition B.2. *Assume A1-2, A3-(\bar{p}), A4-5, A6-(p), A7-(p) and A10-12 for some $2 < p < \bar{p}$. Then,*

$$\sqrt{T_n} \bar{\mu}_n = \mathcal{O}_{L_p}(1), \quad \frac{T_n}{n} \bar{\rho}_n \mathbf{1}_{\lim_k \tilde{S}_k = s_\star} = \mathcal{O}_{L_{p/2}}(1) \mathcal{O}_{\text{a.s.}}(1).$$

Proof. Set $A \stackrel{\text{def}}{=} (I - q\Gamma)$. Under A10, A^{-1} exists. By (36) and (44),

$$A \sqrt{T_n} \bar{\mu}_n = -\frac{\tau_{n+1} \tilde{\mu}_n}{\sqrt{T_n}} + \frac{1}{\sqrt{T_n}} \sum_{k=1}^n \tau_{k+1} \tilde{\epsilon}_k + \frac{1}{\sqrt{T_n}} \sum_{k=1}^n \tau_k \left(\frac{\tau_{k+1}}{\tau_k} - q \right) \Gamma \tilde{\mu}_{k-1}. \quad (45)$$

By Proposition B.1, Lemma 6.3 and A11 the first two terms in the RHS of (45) are bounded in L_p . By (38) the third term in the RHS of (45) satisfies

$$\begin{aligned} & \left\| \frac{1}{\sqrt{T_n}} \sum_{k=1}^n \tau_k \left(\frac{\tau_{k+1}}{\tau_k} - q \right) \Gamma \tilde{\mu}_{k-1} \right\|_{\star, p} \\ & \leq \frac{C}{\sqrt{T_n}} \sum_{k=1}^n \frac{\tau_k}{N_k} \left| \frac{\tau_{k+1}}{\tau_k} - q \right| + \frac{C}{\sqrt{T_n}} \sum_{k=1}^n \sqrt{\tau_k} \left| \frac{\tau_{k+1}}{\tau_k} - q \right|. \quad (46) \end{aligned}$$

By A11(a), $\left\{ \frac{\tau_{k+1}}{\tau_k} \right\}_{k \geq 1}$ converges to q and then, by A12(b), the first term of the RHS of (46) is bounded. The second term is bounded by A11(b).

The proof of the second assertion follows from (44) and Proposition B.1. \square

B.3 L_p -controls of SMC approximations

In this section, we give further details on the L_p control on each block (see (29)):

$$\mathbb{E}_\star \left[\left| \tilde{S}_{\tau_{n+1}}^{N, T_n}(\theta_n, \mathbf{Y}) - \bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right],$$

\bar{S}_τ^T is defined by (6) (we recall that, χ being fixed, it is dropped from the notations) and $\tilde{S}_\tau^{N, T}$ is the SMC approximation of \bar{S}_τ^T based on N particles computed as described in Section 2.4.

The following results are technical lemmas taken from [18] (stated here for a better clarity) or extensions of the L_p controls derived in [21].

Hereafter, “time t ” corresponds to time t in the block n . Therefore, even if it is not explicit in the notations (in order to make them simpler), the following quantities depend upon the observations $\mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$.

Denote by ϕ_s^θ the filtering distribution at time s , and let

$$B_{\phi_t^\theta}^\theta(x, dx') \stackrel{\text{def}}{=} \frac{m_\theta(x', x)}{\int m_\theta(u, x) \phi_t^\theta(du)} \phi_t^\theta(dx')$$

be the backward kernel smoothing kernel at time $t+1$. For all $0 \leq s \leq \tau-1$ and for all bounded measurable function h on $\mathbb{X}^{\tau-s+1}$, define recursively $\phi_{s:\tau|s}^\theta[h]$ backward in time, according to

$$\phi_{s:\tau|s}^\theta[h] = \int \cdots \int B_{\phi_s^\theta}^\theta(x_{s+1}, dx_s) \phi_{s+1:\tau|s}^\theta(dx_{s+1:\tau}) h(x_{s:\tau}), \quad (47)$$

starting from $\phi_{\tau:\tau|s}^\theta = \phi_\tau^\theta$. By convention, $\phi_0^\theta = \chi$.

For $t \geq 1$, let $\{(\xi_t^\ell, \omega_t^\ell)\}_{\ell=1}^N$ be the weighted samples obtained as described in Section 2.4 (see also Algorithm 2 in Appendix A); it approximates the filtering distribution ϕ_t^θ . Denote by $\phi_t^{N, \theta}$ this approximation. For $0 \leq s \leq \tau-1$, an approximation of the backward kernel can be obtained

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

and inserting this expression into (47) gives the following particle approximation of the fixed-interval smoothing distribution $\phi_{0:\tau|s}^\theta[h]$

$$\phi_{0:\tau|s}^{N, \theta}[h] = \sum_{i_0=1}^N \cdots \sum_{i_\tau=1}^N \left(\prod_{u=1}^{\tau} \frac{\omega_{u-1}^{i_u} m_\theta(\xi_{u-1}^{i_{u-1}}, \xi_u^{i_u})}{\sum_{\ell=1}^N \omega_{u-1}^\ell m_\theta(\xi_{u-1}^\ell, \xi_u^{i_u})} \right) \times \frac{\omega_\tau^{i_\tau}}{\Omega_\tau^N} h(\xi_0^{i_0}, \dots, \xi_\tau^{i_\tau}), \quad (48)$$

with $\Omega_\tau^N \stackrel{\text{def}}{=} \sum_{\ell=1}^N \omega_\tau^\ell$.

Lemma B.3. Let $\{(\xi_t^\ell, \omega_t^\ell), 1 \leq \ell \leq N, 0 \leq t \leq \tau_{n+1}\}$ be the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$. Then,

$$\left[\tilde{S}_{\tau_{n+1}}^{N, T_n}(\theta_n, \mathbf{Y}) - \bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}) \right] = \frac{1}{\tau_{n+1}} \left(\phi_{0:\tau_{n+1}|\tau_{n+1}}^{N, \theta_n} [\mathbf{S}_{\tau_{n+1}}] - \phi_{0:\tau_{n+1}|\tau_{n+1}}^{\theta_n} [\mathbf{S}_{\tau_{n+1}}] \right), \quad (49)$$

where

$$\mathbf{S}_\tau(x_{0:\tau}) \stackrel{\text{def}}{=} \sum_{s=1}^{\tau} S(x_{s-1}, x_s, \mathbf{Y}_{s:T}). \quad (50)$$

For all $t \in \{0, \dots, \tau\}$ and all bounded measurable function h on $\mathbb{X}^{\tau+1}$, define the kernel $L_{t,\tau} : \mathbb{X}^{t+1} \times \mathcal{X}^{\otimes \tau+1} \rightarrow [0, 1]$ by

$$L_{t,\tau}^\theta h(x_{0:t}) \stackrel{\text{def}}{=} \int \prod_{u=t+1}^{\tau} m_\theta(x_{u-1}, x_u) g_\theta(x_u, \mathbf{Y}_{u:T}) h(x_{0:\tau}) \lambda(dx_{t+1:\tau}); \quad (51)$$

by convention, $L_{\tau,\tau}^\theta h = h$. Let $\mathcal{L}_{t,\tau}^{N,\theta}$ and $\mathcal{L}_{t,\tau}^\theta$ be two kernels on $\mathbb{X} \times \mathcal{X}^{\otimes(\tau+1)}$ defined for all $x_t \in \mathbb{X}$ by

$$\mathcal{L}_{t,\tau}^\theta h(x_t) \stackrel{\text{def}}{=} \int B_{\phi_{t-1}^\theta}^\theta(x_t, dx_{t-1}) \cdots B_{\phi_0^\theta}^\theta(x_1, dx_0) L_{t,\tau}^\theta h(x_{0:t}) \quad (52)$$

$$\mathcal{L}_{t,\tau}^{N,\theta} h(x_t) \stackrel{\text{def}}{=} \int B_{\phi_{t-1}^{N,\theta}}^\theta(x_t, dx_{t-1}) \cdots B_{\phi_0^{N,\theta}}^\theta(x_1, dx_0) L_{t,\tau}^\theta h(x_{0:t}). \quad (53)$$

Note that

$$\mathcal{L}_{t,\tau}^\theta \mathbf{1}(x_t) = \int m_\theta(x_t, x') g_\theta(x', \mathbf{Y}_{T+t+1}) \mathcal{L}_{t+1,\tau}^\theta \mathbf{1}(x') \lambda(dx'). \quad (54)$$

Lemma B.4, Proposition B.5, Lemma B.6 and B.7 can be found in [18].

Lemma B.4. Let $\{(\xi_t^\ell, \omega_t^\ell), 1 \leq \ell \leq N, 0 \leq t \leq \tau_{n+1}\}$ be the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$. Then,

$$\phi_{0:\tau_{n+1}|\tau_{n+1}}^{N, \theta_n} [h] - \phi_{0:\tau_{n+1}|\tau_{n+1}}^{\theta_n} [h] = \sum_{t=0}^{\tau_{n+1}} \frac{\sum_{\ell=1}^N \omega_t^\ell G_{t,\tau_{n+1}}^{N, \theta_n} h(\xi_t^\ell)}{\sum_{\ell=1}^N \omega_t^\ell \mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}(\xi_t^\ell)}, \quad (55)$$

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

$$G_{t,\tau}^{N,\theta} h(x) \stackrel{\text{def}}{=} \mathcal{L}_{t,\tau}^{N,\theta} h(x) - \frac{\phi_{t-1}^{N,\theta} [\mathcal{L}_{t-1,\tau}^{N,\theta} h]}{\phi_{t-1}^{N,\theta} [\mathcal{L}_{t-1,\tau}^{N,\theta} \mathbf{1}]} \mathcal{L}_{t,\tau}^{N,\theta} \mathbf{1}(x). \quad (56)$$

Proof. By definition of $\mathbf{L}_{t,\tau}^\theta$,

$$\phi_{0:\tau|\tau}^\theta[h] = \frac{\phi_{0:t|t}^\theta[\mathbf{L}_{t,\tau}^\theta h]}{\phi_{0:t|t}^\theta[\mathbf{L}_{t,\tau}^\theta \mathbf{1}]}.$$

We write

$$\phi_{0:\tau|\tau}^{N,\theta}[h] - \phi_{0:\tau|\tau}^\theta[h] = \sum_{t=0}^{\tau} \left\{ \frac{\phi_{0:t|t}^{N,\theta}[\mathbf{L}_{t,\tau}^\theta h]}{\phi_{0:t|t}^{N,\theta}[\mathbf{L}_{t,\tau}^\theta \mathbf{1}]} - \frac{\phi_{0:t-1|t-1}^\theta[\mathbf{L}_{t-1,\tau}^\theta h]}{\phi_{0:t-1|t-1}^\theta[\mathbf{L}_{t-1,\tau}^\theta \mathbf{1}]} \right\},$$

where we used the convention

$$\frac{\phi_{0:-1|t-1}^{N,\theta}[\mathbf{L}_{t-1,\tau}^\theta h]}{\phi_{0:-1|t-1}^{N,\theta}[\mathbf{L}_{t-1,\tau}^\theta \mathbf{1}]} = \frac{\chi[\mathbf{L}_{0,\tau}^\theta h]}{\chi[\mathbf{L}_{0,\tau}^\theta \mathbf{1}]} = \phi_{0:\tau|\tau}^\theta[h].$$

We have for all $0 \leq t \leq \tau$,

$$\phi_{0:t|t}^{N,\theta}[\mathbf{L}_{t,\tau}^\theta h] = \int \phi_t^{N,\theta}(\mathrm{d}x_t) \prod_{j=0}^{t-1} \mathbf{B}_{\phi_j^{N,\theta}}(x_{j+1}, \mathrm{d}x_j) \mathbf{L}_{t,\tau}^\theta h(x_{0:t}) = \phi_t^{N,\theta}[\mathcal{L}_{t,\tau}^{N,\theta} h].$$

Therefore, for all $1 \leq t \leq \tau$,

$$\frac{\phi_{0:t|t}^{N,\theta}[\mathbf{L}_{t,\tau}^\theta h]}{\phi_{0:t|t}^{N,\theta}[\mathbf{L}_{t,\tau}^\theta \mathbf{1}]} - \frac{\phi_{0:t-1|t-1}^{N,\theta}[\mathbf{L}_{t-1,\tau}^\theta h]}{\phi_{0:t-1|t-1}^{N,\theta}[\mathbf{L}_{t-1,\tau}^\theta \mathbf{1}]} = \frac{\phi_t^{N,\theta}[\mathcal{L}_{t,\tau}^{N,\theta} h]}{\phi_t^{N,\theta}[\mathcal{L}_{t,\tau}^{N,\theta} \mathbf{1}]} - \frac{\phi_{t-1}^{N,\theta}[\mathcal{L}_{t-1,\tau}^{N,\theta} h]}{\phi_{t-1}^{N,\theta}[\mathcal{L}_{t-1,\tau}^{N,\theta} \mathbf{1}]} = \frac{\phi_t^{N,\theta}[G_{t,\tau}^{N,\theta} h]}{\phi_t^{N,\theta}[\mathcal{L}_{t,\tau}^{N,\theta} \mathbf{1}]}.$$

□

Proposition B.5. *Let $\{(\xi_t^\ell, \omega_t^\ell), 1 \leq \ell \leq N, 0 \leq t \leq \tau_{n+1}\}$ be the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$. Then,*

$$\left[\tilde{S}_{\tau_{n+1}}^{N,T_n}(\theta_n, \mathbf{Y}) - \bar{S}_{\tau_{n+1}}^{T_n}(\theta_n, \mathbf{Y}) \right] = \frac{1}{\tau_{n+1}} \sum_{t=0}^{\tau_{n+1}} D_{t,\tau_{n+1}}^{N,\theta_n}(\mathbf{S}_{\tau_{n+1}}) + \frac{1}{\tau_{n+1}} \sum_{t=0}^{\tau_{n+1}} C_{t,\tau_{n+1}}^{N,\theta_n}(\mathbf{S}_{\tau_{n+1}})$$

where \mathbf{S}_τ is given by (50) and

$$D_{t,\tau_{n+1}}^{N,\theta_n}(h) \stackrel{\text{def}}{=} \frac{\phi_{t-1}^{N,\theta_n}[v_t]}{\phi_{t-1}^{N,\theta_n} \left[\frac{\mathcal{L}_{t-1,\tau_{n+1}}^{\theta_n} \mathbf{1}}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right]} N^{-1} \sum_{\ell=1}^N \omega_t^\ell \frac{G_{t,\tau_{n+1}}^{N,\theta_n} h(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty}; \quad (57)$$

$$C_{t,\tau_{n+1}}^N(h) \stackrel{\text{def}}{=} \left[\frac{1}{N^{-1} \sum_{i=1}^N \omega_t^i \frac{\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}(\xi_t^i)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty}} - \frac{\phi_{t-1}^{N,\theta_n}[v_t]}{\phi_{t-1}^{N,\theta_n} \left[\frac{\mathcal{L}_{t-1,\tau_{n+1}}^{\theta_n} \mathbf{1}}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right]} \right] N^{-1} \sum_{\ell=1}^N \omega_t^\ell \frac{G_{t,\tau_{n+1}}^{N,\theta_n} h(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty}. \quad (58)$$

Proof. (55) can be rewritten as follows:

$$\phi_{0:\tau_{n+1}|\tau_{n+1}}^{N,\theta_n}[h] - \phi_{0:\tau_{n+1}|\tau_{n+1}}^{\theta_n}[h] = \sum_{t=0}^{\tau_{n+1}} D_{t,\tau_{n+1}}^{N,\theta_n}(h) + \sum_{t=0}^{\tau_{n+1}} C_{t,\tau_{n+1}}^{N,\theta_n}(h). \quad (59)$$

The proof is concluded by Lemma B.3. \square

For any $t \in \{0, \dots, \tau_{n+1}\}$, we recall the definition of $\mathcal{F}_{n,t}^N$ given by (28)

$$\mathcal{F}_{n,t}^N = \sigma \left\{ \theta_n, \mathbf{Y}_{T_n+1:T_n+t+1}, (\xi_s^\ell, \omega_s^\ell); \ell \in \{1, \dots, N\}; 0 \leq s \leq t \right\},$$

where $\{(\xi_t^\ell, \omega_t^\ell)\}_{\ell=1}^N$ are the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$.

Lemma B.6. *Let $\{(\xi_t^\ell, \omega_t^\ell), 1 \leq \ell \leq N, 0 \leq t \leq \tau_{n+1}\}$ be the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$. Then, for any $1 \leq t \leq \tau_{n+1}$ and any $1 \leq \ell \leq N$,*

$$\mathbb{E}_\star [\omega_t^\ell h(\xi_t^\ell) | \mathcal{F}_{n,t-1}^N] = \frac{\phi_{t-1}^{N,\theta_n} [\int m_{\theta_n}(\cdot, x) g_{\theta_n}(x, \mathbf{Y}_{T_n+t}) h(x) \lambda(dx)]}{\phi_{t-1}^{N,\theta_n}[v_t]}. \quad (60)$$

Proof. By definition of the weighted particles,

$$\begin{aligned} & \mathbb{E}_\star [\omega_t^\ell h(\xi_t^\ell) | \mathcal{F}_{n,t-1}^N] \\ &= \mathbb{E}_\star \left[\frac{m_{\theta_n}(\xi_{t-1}^1, \xi_t^1) g_{\theta_n}(\xi_t^1, \mathbf{Y}_{t+T_n})}{v_t(\xi_{t-1}^1) q_t(\xi_{t-1}^1, \xi_t^1)} h(\xi_t^1) \middle| \mathcal{F}_{n,t-1}^N \right] \\ &= \left(\sum_{i=1}^N \omega_{t-1}^i v_t(\xi_{t-1}^i) \right)^{-1} \sum_{i=1}^N \int \omega_{t-1}^i v_t(\xi_{t-1}^i) q_t(\xi_{t-1}^i, x) \\ & \quad \times \frac{m_{\theta_n}(\xi_{t-1}^i, x) g_{\theta_n}(x, \mathbf{Y}_{t+T_n})}{v_t(\xi_{t-1}^i) q_t(\xi_{t-1}^i, x)} h(x) \lambda(dx) \\ &= \left(\sum_{i=1}^N \omega_{t-1}^i v_t(\xi_{t-1}^i) \right)^{-1} \sum_{i=1}^N \int \omega_{t-1}^i m_{\theta_n}(\xi_{t-1}^i, x) g_{\theta_n}(x, \mathbf{Y}_{t+T_n}) h(x) \lambda(dx). \end{aligned}$$

\square

Lemma B.7. *Assume A2 and A8. Let $\{(\xi_t^\ell, \omega_t^\ell), 1 \leq \ell \leq N, 0 \leq t \leq \tau_{n+1}\}$ be the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$.*

(i) *For any $t \in \{0, \dots, \tau_{n+1}\}$ and any measurable function h on $\mathbb{X}^{\tau_{n+1}+1}$, the random variables $\left\{ \omega_t^\ell G_{t,\tau_{n+1}}^{N,\theta_n} h(\xi_t^\ell) | \mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}_{|\infty}^{-1} \right\}_{\ell=1}^N$ are:*

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

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

(ii) For any $t \in \{0, \dots, \tau_{n+1}\}$:

$$\left| \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathcal{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right| \leq \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\}, \quad (61)$$

where \mathcal{S}_τ is defined by (50).

(iii) For all $x \in \mathbb{X}$ and any $t \in \{0, \dots, \tau_{n+1}\}$,

$$\frac{\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}(x)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \geq \frac{\sigma_-}{\sigma_+}, \quad \frac{\mathcal{L}_{t-1,\tau_{n+1}}^{\theta_n} \mathbf{1}(x)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \geq \frac{\sigma_-^2}{\sigma_+} b_-(\mathbf{Y}_{t+T_n}).$$

Proof. The proof of (i) is given by [18, Lemma 3].

Proof of (ii). Let $\Pi_{s-1:s,\tau}$ be the operator which associates to any bounded and measurable function h on $\mathbb{X} \times \mathbb{X}$ the function $\Pi_{s-1:s,\tau} h$ given, for any $(x_0, \dots, x_\tau) \in \mathbb{X}^{\tau+1}$, by

$$\Pi_{s-1:s,\tau} h(x_{0:\tau}) \stackrel{\text{def}}{=} h(x_{s-1:s}).$$

Using this notation, we may write $\mathcal{S}_\tau = \sum_{s=1}^\tau \Pi_{s-1:s,\tau} S(\cdot, \cdot, \mathbf{Y}_{s+T})$ and $G_{t,\tau}^{N,\theta} \mathcal{S}_\tau = \sum_{s=1}^\tau G_{t,\tau}^{N,\theta} \Pi_{s-1:s,\tau} S(\cdot, \cdot, \mathbf{Y}_{s+T})$. Following the same lines as in [18, Lemma 10],

$$\begin{aligned} |G_{t,\tau}^{N,\theta} \Pi_{s-1:s,\tau} S(\cdot, \cdot, \mathbf{Y}_{s+T})|_\infty &\leq \rho^{s-1-t} \text{osc}(S(\cdot, \cdot, \mathbf{Y}_{s+T})) |\mathcal{L}_{t,\tau}^\theta \mathbf{1}|_\infty \quad \text{if } t \leq s-1, \\ |G_{t,\tau}^{N,\theta} \Pi_{s-1:s,\tau} S(\cdot, \cdot, \mathbf{Y}_{s+T})|_\infty &\leq \rho^{t-s} \text{osc}(S(\cdot, \cdot, \mathbf{Y}_{s+T})) |\mathcal{L}_{t,\tau}^\theta \mathbf{1}|_\infty \quad \text{if } t \geq s. \end{aligned}$$

Consequently,

$$\left| G_{t,\tau}^{N,\theta} \mathcal{S}_\tau \right|_\infty \leq \sum_{s=1}^\tau |G_{t,\tau}^{N,\theta} \Pi_{s-1:s,\tau} S(\cdot, \cdot, \mathbf{Y}_{s+T})|_\infty \leq \left(\sum_{s=1}^\tau \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T})\} \right) |\mathcal{L}_{t,\tau}^\theta \mathbf{1}|_\infty,$$

which shows (ii).

Proof of (iii). From the definition (52), for all $x \in \mathbb{X}$ and all $t \in \{1, \dots, \tau\}$,

$$\mathcal{L}_{t,\tau}^\theta \mathbf{1}(x) = \int m_\theta(x, x_{t+1}) g_\theta(x_{t+1}, \mathbf{Y}_{t+T+1}) \prod_{u=t+2}^\tau m_\theta(x_{u-1}, dx_u) g_\theta(x_u, \mathbf{Y}_{u+T}) \lambda(dx_{t+1:\tau}).$$

Hence, by A2,

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

which concludes the proof of the first statement. By (54), A2 and (19),

$$\frac{\mathcal{L}_{t-1,\tau}^\theta \mathbf{1}(x)}{|\mathcal{L}_{t,\tau}^\theta \mathbf{1}|_\infty} = \int m_\theta(x, x') g_\theta(x', \mathbf{Y}_{t+T}) \frac{\mathcal{L}_{t,\tau}^\theta \mathbf{1}(x')}{|\mathcal{L}_{t,\tau}^\theta \mathbf{1}|_\infty} \lambda(dx') \geq \frac{\sigma_-^2}{\sigma_+} b_-(\mathbf{Y}_{t+T}).$$

□

The proofs of Propositions B.8 and B.9 follow the same lines as [21, Propositions 1-2]. The upper bounds given here provide an explicit dependence on the observations.

Proposition B.8. *Assume A2 and A8. Let $\{(\xi_t^\ell, \omega_t^\ell), 1 \leq \ell \leq N, 0 \leq t \leq \tau_{n+1}\}$ be the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_n+1:T_n+\tau_{n+1}}$. For all $p > 1$, there exists a constant C such that for any compact set $\mathcal{K} \subseteq \Theta$,*

$$\begin{aligned} & \mathbb{E}_\star \left[\left| \sum_{t=0}^{\tau_{n+1}} D_{t, \tau_{n+1}}^{N, \theta_n}(\mathbf{S}_{\tau_{n+1}}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \\ & \leq C \frac{\tau_{n+1}^{\left(\frac{p}{2}-1\right)\vee 0}}{N^{p-\left(\frac{p}{2}\vee 1\right)}} \sum_{t=0}^{\tau_{n+1}} \mathbb{E}_\star \left[\left| \frac{\omega_t^{\mathcal{K}}(\mathbf{Y}_{t+T_n})}{b_-(\mathbf{Y}_{t+T_n})} \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\} \right|^p \right]. \end{aligned} \quad (62)$$

where $D_{t, \tau}^{N, \theta}$ is defined in (57).

Proof. By Lemma B.7(iii),

$$\frac{\phi_{t-1}^{N, \theta_n}[v_t]}{\phi_{t-1}^{N, \theta_n} \left[\frac{\mathcal{L}_{t-1, \tau_{n+1}}^{\theta_n} \mathbf{1}}{|\mathcal{L}_{t, \tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right]} \leq \frac{\sigma_+ |v|_\infty}{\sigma_-^2 b_-(\mathbf{Y}_{t+T_n})}.$$

By Lemma B.7(i) and since θ_n is $\mathcal{F}_{n,t}^N$ -measurable for all $t \in \{0, \dots, \tau_{n+1}\}$, $\left\{ D_{t, \tau_{n+1}}^{N, \theta_n}(\mathbf{S}_{\tau_{n+1}}) \mathbf{1}_{\theta_n \in \mathcal{K}}, \mathcal{F}_{n,t}^N \right\}_{0 \leq t \leq \tau_{n+1}}$ is a martingale difference. Since $p > 1$, Burkholder's inequality (see [23, Theorem 2.10, page 23]) states the existence of a constant C depending only on p such that:

$$\mathbb{E}_\star \left[\left| \sum_{t=0}^{\tau_{n+1}} D_{t, \tau_{n+1}}^{N, \theta_n}(\mathbf{S}_{\tau_{n+1}}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \leq C \mathbb{E} \left[\left| \sum_{t=0}^{\tau_{n+1}} \left| D_{t, \tau_{n+1}}^{N, \theta_n}(\mathbf{S}_{\tau_{n+1}}) \right|^2 \right|^{p/2} \mathbf{1}_{\theta_n \in \mathcal{K}} \right].$$

Hence,

$$\begin{aligned} & \mathbb{E}_\star \left[\left| \sum_{t=0}^{\tau_{n+1}} D_{t, \tau_{n+1}}^{N, \theta_n}(\mathbf{S}_{\tau_{n+1}}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \leq C \left(\frac{\sigma_+ |v|_\infty}{\sigma_-^2} \right)^p \\ & \quad \times \mathbb{E}_\star \left[\left| \sum_{t=0}^{\tau_{n+1}} \left| N^{-1} \sum_{\ell=1}^N \frac{\omega_t^\ell}{b_-(\mathbf{Y}_{t+T_n})} \frac{G_{t, \tau_{n+1}}^{N, \theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t, \tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^2 \right|^{p/2} \mathbf{1}_{\theta_n \in \mathcal{K}} \right], \end{aligned}$$

which implies, using the convexity inequality $(\sum_{k=1}^{\tau} a_k)^{p/2} \leq \tau^{(p/2-1)\vee 0} \sum_{k=1}^{\tau} a_k^{p/2}$,

$$\begin{aligned} & \mathbb{E}_\star \left[\left| \sum_{t=0}^{\tau_{n+1}} D_{t, \tau_{n+1}}^{N, \theta_n}(\mathbf{S}_{\tau_{n+1}}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \leq C \left(\frac{\sigma_+ |v|_\infty}{\sigma_-^2} \right)^p \\ & \quad \times \frac{(\tau_{n+1} + 1)^{\left(\frac{p}{2}-1\right)\vee 0}}{N^p} \sum_{t=0}^{\tau_{n+1}} \mathbb{E}_\star \left[\left| \frac{1}{b_-(\mathbf{Y}_{t+T_n})} \sum_{\ell=1}^N \omega_t^\ell \frac{G_{t, \tau_{n+1}}^{N, \theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t, \tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right]. \end{aligned}$$

Since \mathbf{Y}_{t+T_n} and θ_n are $\mathcal{F}_{n,t-1}^N$ -measurable,

$$\begin{aligned} & \mathbb{E}_\star \left[\left| \frac{1}{b_-(\mathbf{Y}_{t+T_n})} \sum_{\ell=1}^N \omega_t^\ell \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \\ &= \mathbb{E}_\star \left[\mathbb{E}_\star \left[\left| \sum_{\ell=1}^N \omega_t^\ell \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^p \middle| \mathcal{F}_{n,t-1}^N \right] \frac{1}{b_-(\mathbf{Y}_{t+T_n})^p} \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \end{aligned}$$

By Lemma B.7(i), using again the Burkholder and convexity inequalities, there exists C s.t.

$$\begin{aligned} \mathbb{E}_\star \left[\left| \sum_{\ell=1}^N \omega_t^\ell \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^p \middle| \mathcal{F}_{t-1,n}^N \right] &\leq CN^{(\frac{p}{2}-1)\vee 0} \mathbb{E}_\star \left[\sum_{\ell=1}^N \left| \omega_t^\ell \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^p \middle| \mathcal{F}_{n,t-1}^N \right] \\ &\leq CN^{\frac{p}{2}\vee 1} \mathbb{E}_\star \left[\left| \omega_t^1 \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^1)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^p \middle| \mathcal{F}_{n,t-1}^N \right]. \end{aligned}$$

The proof is concluded by (61). \square

Proposition B.9. *Assume A2 and A8. Let $\{(\xi_t^\ell, \omega_t^\ell), 1 \leq \ell \leq N, 0 \leq t \leq \tau_{n+1}\}$ be the weighted samples obtained by Algorithm 2 in Appendix A, with input variables $\theta_n, \tau_{n+1}, N, \mathbf{Y}_{T_{n+1}:T_n+\tau_{n+1}}$. For all $\bar{p} > 1$ and all $p \in (1, \bar{p})$, there exists a constant C s.t. for any compact set $\mathcal{K} \subseteq \Theta$ and any $t \in \{0, \dots, \tau_{n+1}\}$,*

$$\begin{aligned} & \mathbb{E}_\star \left[\left| C_{t,\tau_{n+1}}^{N,\theta_n} (\mathbf{S}_{\tau_{n+1}}) \right|^p \mathbf{1}_{\theta_n \in \mathcal{K}} \right] \leq CN^{(\frac{p}{2}\vee \frac{1}{\alpha}) + (\frac{p}{2}\vee \frac{1}{\beta}) - 2p} \\ & \times \mathbb{E}_\star \left[\left| \frac{\omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})}{b_-(\mathbf{Y}_{t+T_n})} \right|^{2p} \mathbb{E}_\star \left[\left| \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\} \right|^{\bar{p}} \middle| \mathcal{F}_{n,t-1}^N \right]^{p/\bar{p}} \right], \end{aligned} \quad (63)$$

where $C_{t,\tau}^{N,\theta}$ is defined in (58) and $\alpha \stackrel{\text{def}}{=} \bar{p}/p$ and $\beta^{-1} = 1 - \alpha^{-1}$.

Proof. Lemma B.6 applied with the function $h = \mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}$ and (54) yield for any $1 \leq \ell \leq N$

$$\mathbb{E}_\star \left[\omega_t^\ell \mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}(\xi_t^\ell) \middle| \mathcal{F}_{n,t-1}^N \right] = \frac{\phi_{t-1}^{N,\theta_n} \left[\mathcal{L}_{t-1,\tau_{n+1}}^{\theta_n} \mathbf{1} \right]}{\phi_{t-1}^{N,\theta_n} [v_t]}.$$

Therefore, by definition of $C_{t,\tau}^{N,\theta}$ (see (58)), $C_{t,\tau_{n+1}}^{N,\theta}(\mathbf{S}_{\tau_{n+1}})$ is equal to

$$\begin{aligned} & \frac{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} B_{n,t}^N = (\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N) \cdots \\ & \times \frac{\Omega_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} \left(\frac{B_{n,t}^N}{\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N]} + \frac{B_{n,t}^N}{\Omega_{n,t}^N \mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N]} (\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N] - \Omega_{n,t}^N) \right) \\ & = B_{n,t}^N (\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N) \frac{\Omega_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} \frac{1}{\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N]} \\ & + (\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N) (\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N] - \Omega_{n,t}^N) \frac{B_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} \frac{1}{\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N]} \end{aligned}$$

with

$$\begin{aligned} A_{n,t}^N & \stackrel{\text{def}}{=} N^{-1} \sum_{\ell=1}^N \omega_t^\ell \frac{\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty}, \\ B_{n,t}^N & \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\ell=1}^N \omega_t^\ell \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathbf{S}_{\tau_{n+1}}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty}, \\ \Omega_{n,t}^N & \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\ell=1}^N \omega_t^\ell. \end{aligned}$$

This can be rewritten,

$$C_{t,\tau}^{N,\theta} = C_1 + C_2,$$

with

$$C_1 = B_{n,t}^N (\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N) \frac{\Omega_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} \frac{1}{\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N]}$$

and

$$\begin{aligned} C_2 & = (\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N) (\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N] - \Omega_{n,t}^N) \\ & \quad \times \frac{B_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} \frac{1}{\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N]}. \end{aligned}$$

By Lemmas B.6 and Lemmas B.7(iii), and A8,

$$\frac{1}{\mathbb{E}_\star [\Omega_{n,t}^N | \mathcal{F}_{n,t-1}^N]} \leq \frac{\sigma_- |v|_\infty}{b_-(\mathbf{Y}_{t+T_n})}; \quad \frac{\Omega_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} \leq \left(\frac{\sigma_+}{\sigma_-} \right)^2 \frac{|v|_\infty}{\sigma_- b_-(\mathbf{Y}_{t+T_n})};$$

and by Lemma B.7(ii)

$$\frac{B_{n,t}^N}{\mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] A_{n,t}^N} \leq \left(\frac{\sigma_+}{\sigma_-} \right)^2 \frac{|v|_\infty}{\sigma_- b_-(\mathbf{Y}_{t+T_n})} \left(\sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}(S(\cdot, \cdot, \mathbf{Y}_{s+T_n})) \right).$$

Therefore, there exists a constant C s.t.

$$\mathbb{E}_\star [|C_1|^p | \mathcal{F}_{n,t-1}^N] \mathbf{1}_{\theta_n \in \mathcal{K}} \leq C \left| \frac{1}{b_-(\mathbf{Y}_{t+T_n})} \right|^{2p} \mathbb{E}_\star \left[|B_{n,t}^N|^p | \mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N|^p | \mathcal{F}_{n,t-1}^N \right] \mathbf{1}_{\theta_n \in \mathcal{K}}$$

Applying the Holder inequality with $\alpha \stackrel{\text{def}}{=} \bar{p}/p \geq 1$ and $\beta^{-1} \stackrel{\text{def}}{=} 1 - \alpha^{-1}$ yields

$$\begin{aligned} & \mathbb{E}_\star \left[|B_{n,t}^N|^p | \mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N|^p | \mathcal{F}_{n,t-1}^N \right] \\ & \leq \mathbb{E}_\star \left[|B_{n,t}^N|^{\alpha p} | \mathcal{F}_{n,t-1}^N \right]^{1/\alpha} \mathbb{E}_\star \left[| \mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N |^{\beta p} | \mathcal{F}_{n,t-1}^N \right]^{1/\beta}. \end{aligned}$$

By Proposition B.8,

$$\begin{aligned} & \mathbb{E}_\star \left[|B_{n,t}^N|^{\alpha p} | \mathcal{F}_{n,t-1}^N \right]^{1/\alpha} \mathbf{1}_{\theta_n \in \mathcal{K}} \\ & \leq CN^{(\frac{p}{2} \vee \frac{1}{\alpha}) - p} \mathbb{E}_\star \left[\left| \omega_t^1 \frac{G_{t,\tau_{n+1}}^{N,\theta_n} \mathcal{S}_{\tau_{n+1}}(\xi_t^1)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right|^{\alpha p} | \mathcal{F}_{n,t-1}^N \right]^{1/\alpha} \mathbf{1}_{\theta_n \in \mathcal{K}} \\ & \leq CN^{(\frac{p}{2} \vee \frac{1}{\alpha}) - p} \omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})^p \mathbb{E}_\star \left[\left| \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\} \right|^{\alpha p} | \mathcal{F}_{n,t-1}^N \right]^{1/\alpha}, \end{aligned}$$

Given $\mathcal{F}_{n,t-1}^N$, the random variables $\left\{ \mathbb{E}_\star \left[\omega_t^\ell \frac{\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}(\xi_t^1)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} | \mathcal{F}_{n,t-1}^N \right] - \omega_t^\ell \frac{\mathcal{L}_{t,\tau_{n+1}} \mathbf{1}(\xi_t^\ell)}{|\mathcal{L}_{t,\tau_{n+1}}^{\theta_n} \mathbf{1}|_\infty} \right\}_{\ell=1}^N$ are conditionally independent, centered and bounded by Lemma B.7. Following the same steps as in the proof of Proposition B.8, there exists a constant C such that

$$\mathbb{E}_\star \left[| \mathbb{E}_\star [A_{n,t}^N | \mathcal{F}_{n,t-1}^N] - A_{n,t}^N |^{\beta p} | \mathcal{F}_{n,t-1}^N \right]^{1/\beta} \mathbf{1}_{\theta_n \in \mathcal{K}} \leq CN^{(\frac{p}{2} \vee \frac{1}{\beta}) - p} \omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})^p.$$

Hence,

$$\begin{aligned} & \mathbb{E}_\star [|C_1|^p \mathbf{1}_{\theta_n \in \mathcal{K}}] \leq CN^{(\frac{p}{2} \vee \frac{1}{\alpha}) + (\frac{p}{2} \vee \frac{1}{\beta}) - 2p} \\ & \times \mathbb{E}_\star \left[\left| \frac{\omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})}{b_-(\mathbf{Y}_{t+T_n})} \right|^{2p} \mathbb{E}_\star \left[\left| \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\} \right|^{\bar{p}} | \mathcal{F}_{n,t-1}^N \right]^{p/\bar{p}} \right]. \end{aligned}$$

Similarly, using

$$\mathbb{E}_\star \left[| \mathbb{E}_\star [\Omega_{t,n}^N | \mathcal{F}_{n,t-1}^N] - \Omega_{t,n}^N |^{\alpha p} | \mathcal{F}_{n,t-1}^N \right]^{1/\alpha} \mathbf{1}_{\theta_n \in \mathcal{K}} \leq CN^{(\frac{p}{2} \vee \frac{1}{\alpha}) - p} \omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})^p,$$

yields

$$\begin{aligned} & \mathbb{E}_\star [|C_2|^p \mathbf{1}_{\theta_n \in \mathcal{K}}] \leq CN^{-p} \\ & \times \mathbb{E}_\star \left[\left| \frac{\omega_+^{\mathcal{K}}(\mathbf{Y}_{t+T_n})}{b_-(\mathbf{Y}_{t+T_n})} \right|^{2p} \mathbb{E}_\star \left[\left| \sum_{s=1}^{\tau_{n+1}} \rho^{|t-s|} \text{osc}\{S(\cdot, \cdot, \mathbf{Y}_{s+T_n})\} \right|^{\bar{p}} | \mathcal{F}_{n,t-1}^N \right]^{p/\bar{p}} \right]. \end{aligned}$$

□

References

- [1] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot. Consistency of the EKF-SLAM algorithm. In *IEEE International Conference on Intelligent Robots and Systems*, pages 3562–3568, 2006.
- [2] P. Billingsley. *Probability and Measure*. Wiley, New York, 3rd edition, 1995.
- [3] W. Burgard, D. Fox, and S. Thrun. *Probabilistic robotics*. Cambridge, MA:MIT Press, 2005.
- [4] O. Cappé. Online sequential Monte Carlo EM algorithm. In *IEEE Workshop on Statistical Signal Processing (SSP)*, pages 37–40, 2009.
- [5] O. Cappé. Online EM algorithm for Hidden Markov Models. *To appear in J. Comput. Graph. Statist.*, 2011.
- [6] O. Cappé and E. Moulines. Online Expectation Maximization algorithm for latent data models. *J. Roy. Statist. Soc. B*, 71(3):593–613, 2009.
- [7] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer. New York, 2005.
- [8] H. Chen, A. Gao, and L. Guo. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stoch. Proc. Appl.*, 27:217–231, 1988.
- [9] J. Davidson. *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press, 1994.
- [10] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer. New York, 2004.
- [11] P. Del Moral, A. Doucet, and S.S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5):947–975, 2010.
- [12] P. Del Moral, A. Doucet, and S.S. Singh. Forward smoothing using sequential Monte Carlo. Technical report, arXiv:1012.5390v1, 2010.
- [13] P. Del Moral and A. Guionnet. Large deviations for interacting particle systems: applications to non-linear filtering. *Stoch. Proc. Appl.*, 78:69–95, 1998.
- [14] P. Del Moral, M. Ledoux, and L. Miclo. On contraction properties of Markov kernels. *Probab. Theory Related Fields*, 126(3):395–420, 2003.
- [15] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27(1), 1999.
- [16] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39(1):1–38 (with discussion), 1977.

- [17] R. Douc, G. Fort, E. Moulines, and P. Priouret. Forgetting the initial distribution for hidden Markov models. *Stochastic Processes and their Applications*, 119(4):1235–1256, 2009.
- [18] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *To appear in Ann. Appl. Probab.*, 4 2010.
- [19] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.
- [20] A. Doucet, G. Poyiadjis, and S.S. Singh. Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, 98(1):65–80, 2010.
- [21] C. Dubarry and S. Le Corff. Non-asymptotic deviation inequalities for smoothed additive functionals in non-linear state-space models. Technical report, arXiv:1012.4183v1, 2011.
- [22] G. Fort and E. Moulines. Convergence of the Monte Carlo Expectation Maximization for curved exponential families. *Ann. Statist.*, 31(4):1220–1259, 2003.
- [23] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, New York, London, 1980.
- [24] S. Le Corff and G. Fort. Convergence of a particle-based approximation of the Block Online Expectation Maximization algorithm. Technical report, arXiv, 2011.
- [25] S. Le Corff and G. Fort. Online Expectation Maximization based algorithms for inference in Hidden Markov Models. Technical report, arXiv:1108.3968v1, 2011.
- [26] S. Le Corff and G. Fort. Supplement paper to "Online Expectation Maximization based algorithms for inference in Hidden Markov Models". Technical report, arXiv:1108.4130v1, 2011.
- [27] S. Le Corff, G. Fort, and E. Moulines. Online EM algorithm to solve the SLAM problem. In *IEEE Workshop on Statistical Signal Processing (SSP)*, 2011.
- [28] F. Le Gland and L. Mevel. Recursive estimation in HMMs. In *Proc. IEEE Conf. Decis. Control*, pages 3468–3473, 1997.
- [29] J.S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.
- [30] Ruben Martinez-Cantin. *Active Map Learning for Robots: Insights into Statistical Consistency*. PhD thesis, University of Zaragoza, 2008.

- [31] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, New York, 1997.
- [32] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, London, 1993.
- [33] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In *Proceedings of the Sixteenth IJCAI*, Mexico, 2003.
- [34] G. Pólya and G. Szegő. *Problems and Theorems in Analysis. Vol. II*. Springer, 1976.
- [35] B. T. Polyak. A new method of stochastic approximation type. *Autom. Remote Control*, 51:98–107, 1990.
- [36] Vladislav B. Tadić. Analyticity, convergence, and convergence rate of recursive maximum-likelihood estimation in hidden Markov models. *IEEE Trans. Inf. Theor.*, 56:6406–6432, December 2010.
- [37] D. M. Titterton. Recursive parameter estimation using incomplete data. *J. Roy. Statist. Soc. B*, 46(2):257–267, 1984.
- [38] C. F. J. Wu. On the convergence properties of the EM algorithm. *Ann. Statist.*, 11:95–103, 1983.